Ivana Marković, BSc

Moment-Based SNR Estimation

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieurin

Master's degree programme: Electrical Engineering

Submitted to

Graz University of Technology

Supervisor

Assoc. Prof. Dipl.-Ing. Dr. techn. Wilfried Gappmair

Institute of Communication Networks and Satellite Communications

Graz University of Technology, Austria

Graz, December 2015

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

_____ _____

Date Signature

Acknowledgments

First of all, I would like to express my very great appreciation to my supervisor Assoc. Prof. Dipl.-Ing. Dr. techn. Wilfried Gappmair for his continuous and helpful mentoring and valuable advices which substantially contributed to the quality of this master thesis.

Furthermore, I owe more than thanks to my family members: my mother Emilija, my sister Kristina, my brother-in-law Milos and my nephews Mihailo and Luka for their unconditional love, concern and encouragement throughout my study and life.

Moreover, I would like to express my gratitude to all my friends, who have supported me in any way during my work on this thesis.

Special thanks should be given to Nemanja, who has always supported me and stood by me through the good times and bad.

Finally, I dedicate this thesis to the memory of my beloved father, Zvonko Marković (1948-2014), who is a constant source of my inspiration.

Abstract

In wireless communication systems, the channel introduces distortions and noise to the communication link between transmitter and receiver. In this context, the signal-to-noise ratio (SNR) is one of the major figures of merit, which characterizes the channel quality. Typically, modern communication systems require knowledge of the SNR for proper operation, such as power control for adaptive coding and modulation (ACM) or iterative soft-decoding procedures as most prominent examples. For this reason, the development of powerful SNR estimation techniques is a hot and timely topic.

Intending to optimize the overall system performance, next-generation satellite systems, as they are specified in the DVB-S2 and DVB-S2x standards, include higher-order modulation schemes such as 16-APSK or 32-APSK constellations, which enable the improvement in terms of spectral efficiency. However, these modulation formats are particularly challenging, when envelope-based SNR estimators have to be envisaged.

After introduction and some basic notes on parameter estimation, the current thesis provides in the first part an assessment of different procedures for SNR estimation available from the open literature. Of particular interest is the modified version of an algorithm based on second- and fourth-order moments, which produces reliable SNR estimates for modulation schemes with non-constant envelope. Furthermore, a new concept for SNR estimation, based on a dual-filter framework, will be presented in the second part of this thesis.

Zusammenfassung

In funkgestützten Systemen zur Nachrichtenübertragung verursacht der Kanal Störungen verschiedenster Art zwischen Sender und Empfänger. In diesem Zusammenhang ist das Signal-Rauschleistungsverhältnis (SNR) eine wichtige Kenngröße, welche die Kanalqualität charakterisiert. Typischerweise verlangen moderne Nachrichtensysteme die Kenntnis des jeweiligen SNR-Wertes, um die richtige Funktionsweise zu garantieren, etwa als Kontrollmechanismus bei adaptiver Codierung und Modulation (ACM) oder bei iterativen Kanaldecodern. Aus diesem Grund ist die Entwicklung von leistungsfähigen SNR-Schätzern ein wichtiges Forschungsfeld.

In der Absicht die Systemleistung zu optimieren, sind für Satellitensysteme der nächsten Generation, wie sie etwa im DVB-S2 bzw. DVB-S2x Standard spezifiziert sind, auch Modulationsverfahren höherer Ordnung vorgesehen, beispielsweise 16-APSK oder 32-APSK Konstellationen. Allerdings sind diese Modulationsschemata besonders anspruchsvoll, wenn es um die SNR-Schätzung ohne Kenntnis der Trägerphase sowie der darunterliegenden Daten geht.

Nach der Einleitung und einigen grundlegenden Betrachtungen über Parameterschätzung geht es im ersten Teil der vorliegenden Masterarbeit um die Bewertung verschiedener Verfahren zur SNR-Schätzung, wie sie in der offenen Literatur zu finden sind. Von besonderem Interesse ist dabei die modifizierte Form eines Algorithmus, der auf den Momenten zweiter bzw. vierter Ordnung beruht und der zuverlässige Schätzwerte für SNR liefert, ohne dass Trägerphase und Daten bekannt sein müssen. Im zweiten Teil der Arbeit wird ein neuartiges Konzept zur SNR-Schätzung diskutiert, das auf dem sogenannten dualen Filterkonzept beruht.

Contents

Table of Figures

References

[1] R. Lopez-Valcarce, J. Villares, J. Riba, W. Gappmair, and C. Mosquera, “Cramer-Rao bounds for SNR estimation of oversampled linearly modulated signals,” *IEEE Trans. Signal Processing*, vol. 63, pp. 1675–1683, April 2015.

[2] W. Gappmair, R. Lopez-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, pp. 1622-1627, June 2009.

[3] W. Gappmair, *Software Defined Radio*, Lecture Script, Feb. 2014.

[4] W. Gappmair, R. Lopez-Valcarce, and C. Mosquera, “New Results on envelope-based SNR estimation for nonconstant modulus signals”, *Internal Document*, Jan. 2014

[5] J. G. Proakis and D. G. Manokalis, *Digital Signal Processing: Principles, Algorithms and Applications,* Prentice Hall, Upper Saddle River, NJ, 1996

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, 1993.

[7] J. Riba, J. Villares, and G. Vazquez, “A nondata-aided SNR estimation technique for multilevel modulations exploiting signal cyclostationarity,” *IEEE Trans. Signal Process*., vol. 58, pp. 5767-5778, Nov. 2010.

[8] U. Mengali and A. N. D’Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, New York, 1997.

[9] D. R. Pauluzzi and N. C. Beaulieu, “A comparison of SNR estimation techniques for the AWGN channel,” *IEEE Trans. Commun*., vol. 48, pp. 1681-1691, Oct. 2000.

[10] 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. 301-304.

[11] M. Alvarez-Diaz, R. Lopez-Valcarce, and C. Mosquera, “SNR estimation for multilevel constellations using higher-order moments”, *IEEE Trans. Signal Process*., vol. 58, pp. 1515-1526, March 2010.

[12] R. Lopez-Valcarce and C. Mosquera, “Sixth-order statistics-based non-data-aided SNR estimation”, *IEEE Commun. Lett*., vol. 11, pp. 351-353, April 2007.

[13] G. Ferrari, G. Colavolpe and R. Raheli, *Detection Algorithms for Wireless Communications*, John Wiley & Sons, Chichester, UK, 2004.

[14] C. E. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal, 1948

[15] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: turbo codes”, in *Proc. IEEE Commun. Conf.*, Geneva, Switzerland, May 1993, pp.1064-1070.

[16] D. J. C. MacKay and R. M. Neal, “Near Shannon Limit Performance of Low Density Parity Check Codes”, *Electron. Lett*., vol. 33, pp. 457–458, March 1997.

[17] HEL: Radio Laboratory Guidelines, Helsinki University of Technology, http://www.comlab.hut.fi/opetus/333/2004_2005_slides/modulation_methods.pdf

[18] EN 302 307, Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2), *ETSI Standard*, Aug. 2009.

[19] M. Bergmann, Parameter Estimation on Interference-Imposed Satellite Channels, *PhD Thesis*, Graz Univ. Technol., Nov. 2012.

[20] K. Gentile, Digital Pulse-Shaping Filter Basic, http://www.analog.com/media/en/technical-documentation/application-notes/AN-922.pdf

[21] Advantech Wireless, DVB-S2 Technical Presentation, http://www.advantechwireless.com/wp-content/uploads/DVB-S2-theory.pdf

Acronyms

ACF	Autocorrelation Function
ACM	Adaptive Coding and Modulation
APSK	Amplitude and Phase Shift Keying
AWGN	Additive White Gaussian Noise
CM	Constant Modulus
CPM	Constant Phase Modulation
CRLB	Cramer-Rao Lower Bound
DA	Data-Aided
DFT	Discrete Fourier Transform
DMF	Derivate Matched Filter
DVB	Digital Video Broadcasting
EVB	Envelope-Based
ISI	Inter-symbol Interference
LPF	Low-Pass Filter
M2M4	Second- and Fourth-Order Moments
MF	Matched Filter
ML	Maximum Likelihood
MSE	Mean Square Error
MVUE	Minimum Variance Unbiased Estimator
NCRLB	Normalized Cramer-Rao Lower Bound
NDA	Non-Data-Aided
NMSE	Normalized Mean Square Error
NSE	Noise Subspace Estimator
PDF	Probability Density Function
PSD	Power Spectral Density
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RCos	Raised Cosine
RRCos	Root-Raised Cosine
SNR	Signal-to-Noise Ratio

1. Introduction

Due to long transmission distances, satellite communications is characterized by a long round trip times (RTT) and significant signal attenuation. Power amplifiers are often driven close to the saturation point resulting in non-linear distortions. A simplified signal model, considering modulation and coding only, is depicted in the Figure 1-1 [13] [19]. Analog front ends performing radio frequency carrier modulation and demodulation are considered to be a part of the channel.

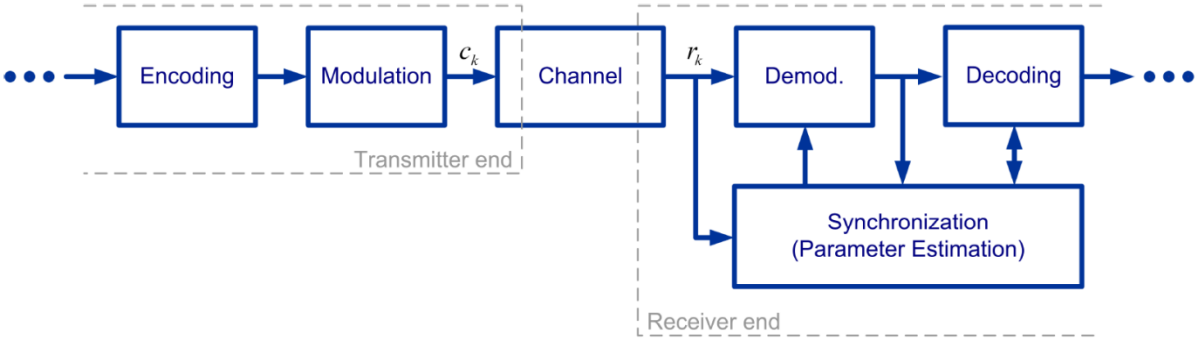


Figure 1-1: Baseband model of a communication system

The information rate at which error free transmission for a given SNR value turns out to be possible is theoretically bounded by the so-called channel capacity. This figure of merit was introduced in 1948 by Claude E. Shannon in his *theorem on channel coding*, but he did not present any algorithms that are able to achieve this bound [14]. In the 1990s iterative channel coding schemes were published, which approach the Shannon limit. Most prominent examples in this respect are the turbo and LDPC codes [15] [16]. Figure 1-2 shows the comparison of the performance in the terms of the bit error rate (BER) of different forward error correction (FEC) schemes. Binary phase shift keying (BPSK) and additive white Gaussian noise (AWGN) channel are assumed.

The approach of the Shannon bound by iterative decoding techniques increases the synchronization and parameter estimation requirements. As it can be seen in the Figure 1-2, whenever powerful FEC schemes are used, system operation at very low SNR values (0 dB or less) must be guaranteed, which means that the demodulator has to be reliably operated under these conditions. This means also that, in order to

facilitate the system operation near the Shannon bound, one must estimate and synchronize the most important transmission parameters although the energy of the signal might not be larger than the noise energy [3]. The most important parameters needed for reliable demodulation are carrier frequency/phase and symbol timing [19].

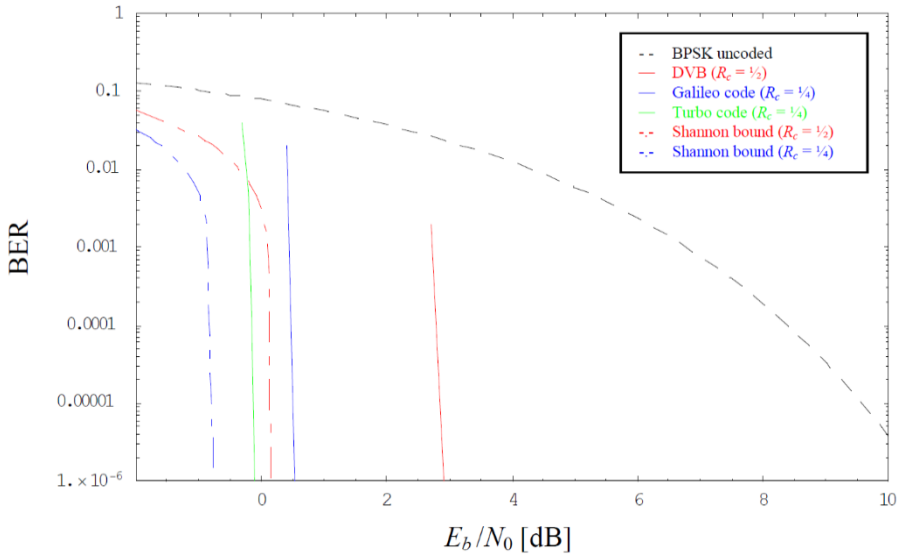


Figure 1-2: Performance of different FEC schemes in BPSK/AWGN channels

2. Baseband Signal Decomposition

2.1. Modulation

Modulation techniques can be linear or nonlinear. If the input-output relation of a modulator satisfies the superposition property, i.e. the output of a modulator produced by a number of inputs applied simultaneously is equal to the sum of the outputs that result when the inputs are applied one at a time, then we are talking about linear modulation. If the input-output relation of modulator does not (partially or fully) satisfy the superposition property, the modulation process is nonlinear [17]. Linear modulations, like PSK, APSK and QAM are typically used in satellite communications, although some nonlinear modulations, e.g. Gaussian Minimum Shift Keying, have been adopted some mobile standards.

Digital transmission via satellite is affected by power and bandwidth limitations. Due to the power limitations, the high power amplifier (HPA) is often operated close to saturation which introduces strong nonlinear effects. This is the reason why M-ary PSK schemes, which are rather robust to nonlinear distortions, are still preferred in satellite communications. However, new generation satellite systems use M-ary APSK and QAM constellations as well in order to increase spectral efficiency in case of good propagation conditions. Figure 2-1 shows constellation diagrams of QPSK and 32-APSK schemes [18].

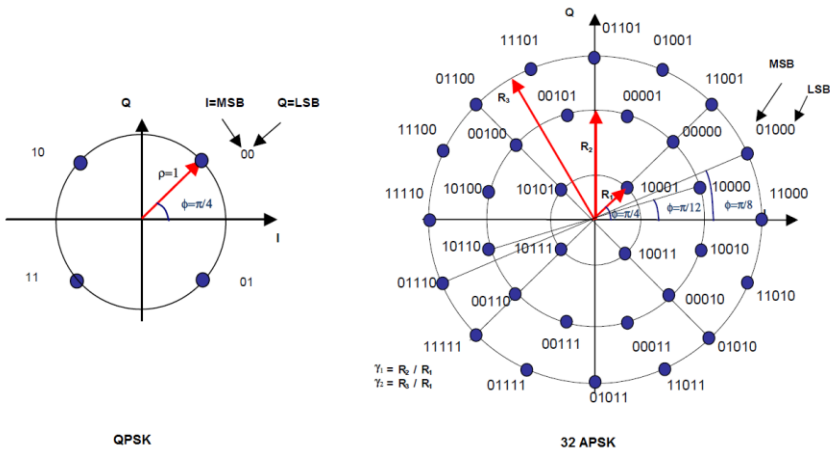


Figure 2-1: Constaltion diagramms for QPSK and 32-APSK

Typically, one has to make a trade-off between spectral and power efficiency when choosing a modulation scheme. Constant envelope constellations like M-PSK have high power efficiency because of the fact that the amplitude is constant at sampling time, which results in robustness to nonlinear distortions. On the other hand, M-APSK schemes, for example, are more efficient regarding bandwidth but much more sensitive to nonlinearities.

Nonlinear modulation techniques like continuous phase modulation (CPM) offer high bandwidth efficiency due to the absence of abrupt phase transitions while exhibiting excellent power efficiency as consequence of their constant envelope. However, such modulation techniques are still of minor importance in satellite communications because of the complex signal processing [3].

State-of-the-art satellite systems, e.g. specified in standard for the digital video broadcasting via satellite (DVB-S2) [18], typically facilitate bi-directional data services using time division multiple access (TDMA). Transmission occurs in frames or bursts which do not only contain useful data but also training sequences for carrier and timing recovery as well as periodically inserted pilot sequences for tracking purposes [19].

2.2. Pulse Shaping Filter

In today's communications systems there is a constant need for more channels, i.e. more users, so that the crowding of the frequency spectrum might become an issue. At the same time, there is a constant demand for high data rates. Therefore, in order to deal with these two conflicting requirements, appropriate pulse shaping techniques are mandatory.

2.2.1. Inter-Symbol Interference

In band-limited systems, increasing the channel bandwidth to achieve higher data rates, inter-symbol interference (ISI) is a major issue, where the energy from one symbol leaks into other symbols (Figure 2-2). This is clearly an unwanted phenomenon, since the received signal becomes distorted and we observe a higher error probability at the output.

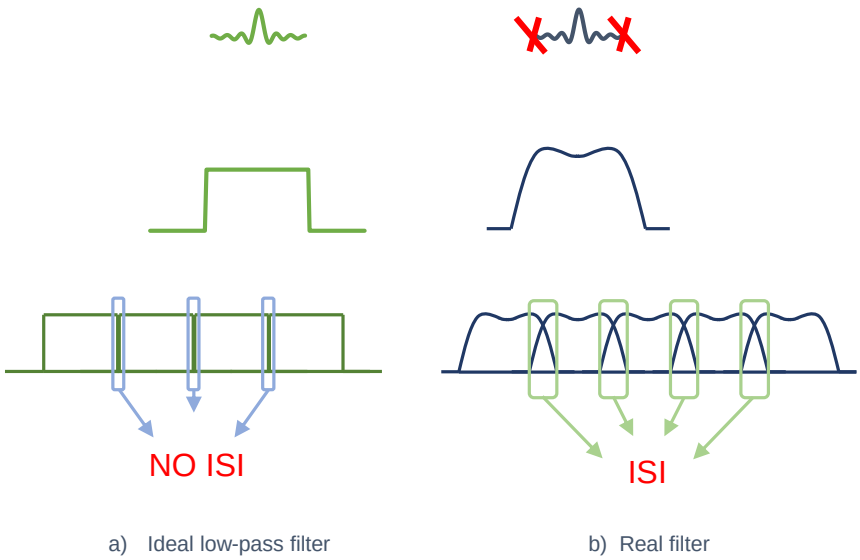


Figure 2-2: Background of inter-symbol interference

As already pointed out, the spectral content of the data to be transmitted has to be limited to a certain bandwidth. This is achieved by forcing the data to pass through a so called pulse shaping filter. In order to avoid ISI, the related filter must be designed according to the following properties:

- Excluding the origin, the impulse response has zeros at integer multiples of the symbol period (see Figure 2-3).
- In the frequency domain, the filter must satisfy the first Nyquist criterion.

The simplest solution is an ideal low-pass filter (LPF). As it can be seen from Figure 2-2 a), the frequency response of the ideal LPF, which has a rectangular shape, is narrow enough, whereas its related impulse response $\mathrm{sinc}(t/T)$, has side lobes that are zero at integer multiples of *T* so that no ISI occurs, which is exemplified in Figure 2-3. However, due to the sharp edges of the transfer function, such a filter is physically not realizable.

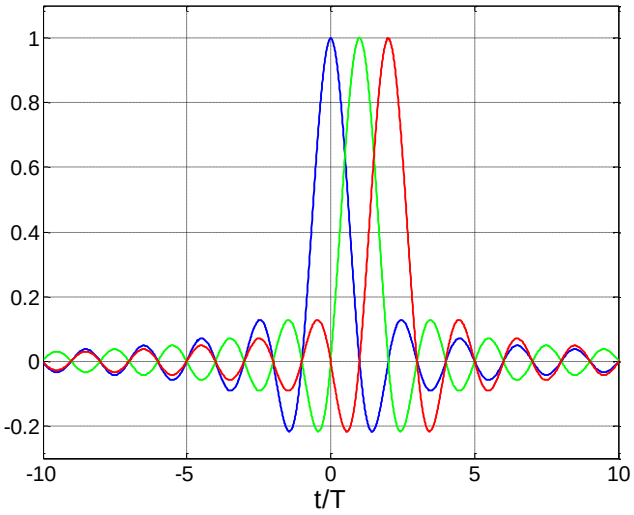


Figure 2-3: Impulse responses illustrating the zero-crossing property

2.2.2. Raised Cosine Filter

Clearly, in order to design a realizable filter that satisfies the requirements mentioned above, a tradeoff has to be made between the ripple amplitudes in time domain and the bandwidth constraints in frequency domain.

A typical approach in this context is to use a smoother transition function instead of problematic sharp edges. The most prominent example is raised cosine (RCos) filter. Its impulse response $g(t)$ is given as:

$$g(t) = \mathrm{sinc}(t/T)\frac{\cos(\pi\alpha t/T)}{1-(2\alpha\, t/T)^2}, \qquad 0 \le \alpha \le 1\,. \tag{2.1}$$

In this context, α is so called roll-off factor or excess bandwidth, which characterizes the filter. It can take values between 0 and 1, i.e. $0 \le \alpha \le 1$. Furthermore, T denotes the symbol period. Applying the Fourier transform on (2.1), we obtain the transfer function of a RCos filter furnished by:

$$G(f) = \begin{cases} T, & |f| \leq \dfrac{1-\alpha}{2T} \\ T\cos^2\left[\dfrac{\pi}{2\alpha}(|fT|) - \dfrac{1-\alpha}{2}\right], & \dfrac{1-\alpha}{2T} < |f| < \dfrac{1+\alpha}{2T} \\ 0, & |f| > \dfrac{1+\alpha}{2T} \end{cases} \tag{2.2}$$

Figures 2-4 and 2-5 depict the evolution of a RCos filter for different values of α both in time and frequency domain, respectively. Clearly, for $\alpha = 0$, the RCos filter embodies the ideal LPF. In this particular case (in Figure 2.5 depicted in blue color), the ringing artifacts due to the Gibbs phenomenon, which will be described later, are evident. Furthermore, it is to be noticed that larger values for α result in larger bandwidths in the frequency domain and side lobes that decay faster in time domain. By looking at the impulse response of the RCos filter (see Figure 2-4), one can observe that α does not impact the position of zeros, but only the amplitude of the side lobes are affected.

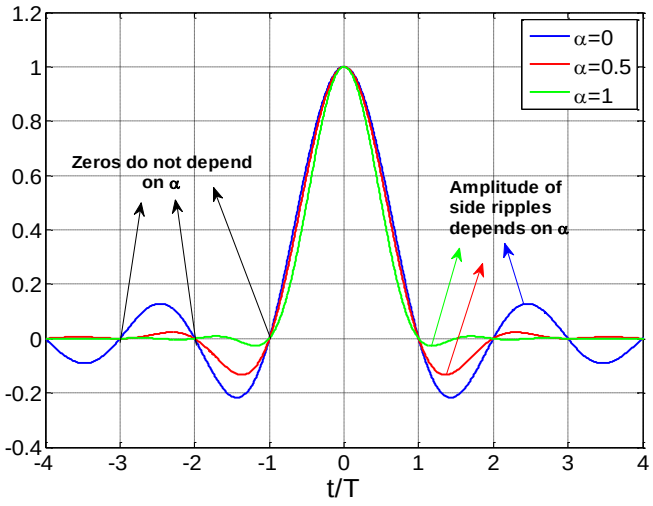


Figure 2-4: Impulse response of RCos shapes for different roll-off factors

Considering only the transfer function of a RCos pulse (see Figure 2-5), we can identify three major spectral positions:

- Nyquist frequency $f = \pm 1/2T$, i.e. half of the symbol rate, which is the minimum bandwidth where perfect recovery is guaranteed. For any value of α, the magnitude is reduced by 3 dB.

- Stop band frequency at $f = \pm(1+\alpha)/2T$, where the magnitude of the frequency response reaches zero.
- Pass band frequency at $f = \pm(1-\alpha)/2T$, where frequency response starts to get smaller than the maximum magnitude.

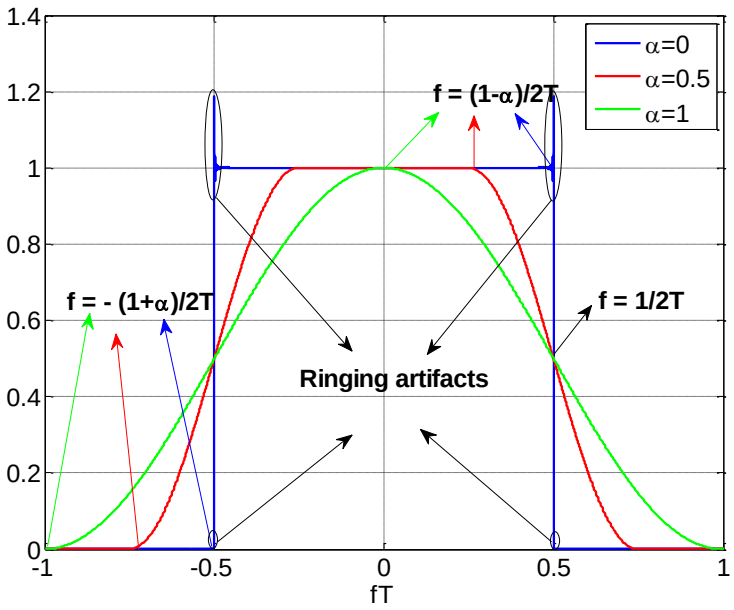


Figure 2-5: Frequency response of RCos pulse shape for different roll-off factors

It is important to notice that, RCos filters occupy a part of the bandwidth beyond the Nyquist bandwidth of $\pm 1/2T$. This part of the bandwidth, so-called excess bandwidth, is strongly related to the value of α, i.e. $B_{excess} = \alpha/T$. Therefore, the total spectral support of RCos pulses is given as $B = B_{excess} + B_{Nyquist} = (1+\alpha)/T$ (double-sided spectrum representation).

2.2.3. Root-Raised Cosine Filters

In order to improve noise cancelation, pulse shaping filters are usually split into two parts, with one at the transmitter and other at the receiver end [20]. Thus, the transfer function of each filter becomes the square root of the RCos filter function:

$$G(f) = |H(f)|^2 \quad \leftrightarrow \quad g(t) = h(t) \otimes h^*(-t). \tag{2.3}$$

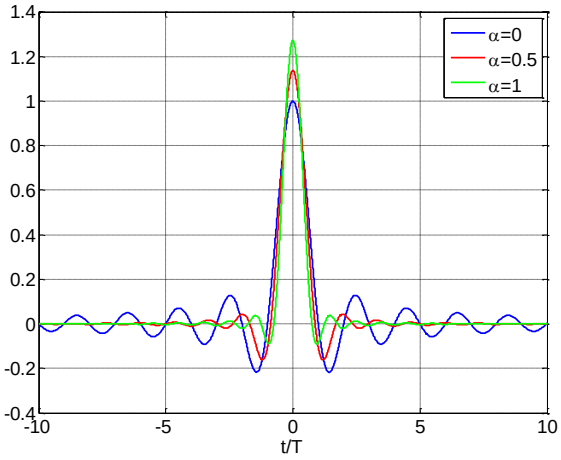
This means that the impulse response of root-raised cosine (RRCos) filter $h(t)$ is simply computed in that the inverse Fourier transform is applied to $H(f)$:

$$
\begin{aligned}
h(t) &= h^*(-t) \\
&= \int_{-\infty}^{\infty} H(f) e^{j2\pi f t}\, df \\
&= \frac{2}{\sqrt{T}} \int_{0}^{(1-\alpha)/2} \cos(2\pi v\, t/T)\, dv \\
&\quad + \frac{2}{\sqrt{T}} \int_{(1-\alpha)/2}^{(1+\alpha)/2} \cos\left[\frac{\pi}{2\alpha}\left(v - \frac{1-\alpha}{2}\right)\right] \cos(2\pi v t/T)\, dv \\
&= \frac{1}{\sqrt{T}} \frac{4\alpha \cos[\pi\,(1+\alpha)t/T] + \pi(1-\alpha)sinc[(1-\alpha)t/T]}{\pi[1-(4\alpha\, t/T)^2]}.
\end{aligned}
\tag{2.4}
$$

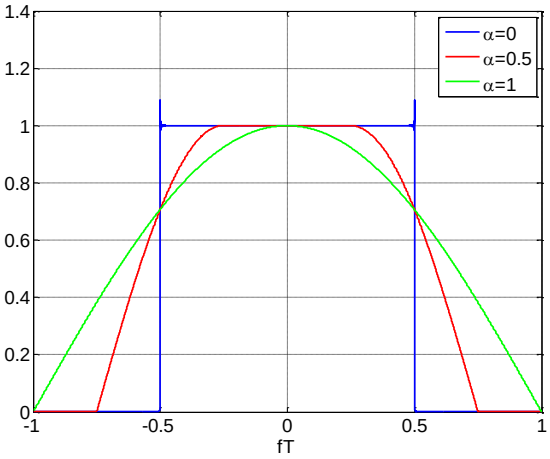
By sampling $g(t)$ at $t = 0$ and bearing in mind that the relationship between $|H(f)|^2$ and $|h(t)|^2$ is determined through the Parseval theorem, we can get the normalized value for the energy per symbol [3]:

$$
g(0) = \int_{-\infty}^{\infty} G(f) df = \int_{-\infty}^{\infty} |H(f)|^2 df = \int_{-\infty}^{\infty} |h(t)|^2 dt = 1\,. \tag{2.5}
$$

Figure 2-6 depicts the RRCos filter in time and frequency domain for different values of α. It can be seen that, in contrast to RCos, the impulse response of the RRCos filter function does not satisfy the zero-crossing property required for ISI-free conditions.



a) Evolution of RRCos in time domain



b) Evolution of RRCos in frequency domain

Figure 2-6: RRCos shapes for different roll-off factors

2.2.4. Design of RRCos Filters

In the sequel, the implementation of a digital finite impulse response (FIR) RRCos filter will be briefly described. The following discussion is also relevant for the filter that will be introduced in the Chapter 5, since it is derived from a RRCos shape.

The beauty of the RRCos pulse shape lies in the fact that the output of related matched filter is ISI-free. The bad thing is that it has theoretically infinite duration in the time domain, so that it cannot be implemented in practical systems. Hence, its impulse response has to be truncated to a finite length, which leads to oscillations and ringing effect in the frequency domain. This effect is called the Gibbs phenomenon.

As mentioned before, for $\alpha \neq 0$, the RRcos filter exceeds Nyquist bandwidth $1/T$ by an amount of α/T, so that the largest frequency points $\pm(1+\alpha)/2T$ are located at

more than half the data rate. Therefore, in order avoid any alias effect, signals must be oversampled by a factor of $1 + \alpha$, which gives in the worst case of $\alpha = 1$ a factor of $N_{OS} = 2$ (typically 2 to 4 samples per symbol are used in practice).

An appropriate truncation of RRCos filters for $\alpha = 0$, due to rapid transition between pass-band and stop-band, is not possible. Hence, in this case, severe ringing artifacts will exist, no matter how large the number of the filter taps is used for the final design.

On the other hand, for $\alpha > 0$, where the edge effect is smoother, this undesirable effect is reduced. However, as a consequence of the impulse response truncation, the frequency response will not be zero in the stop-band, i.e. the truncation introduces non-zero side lobes.

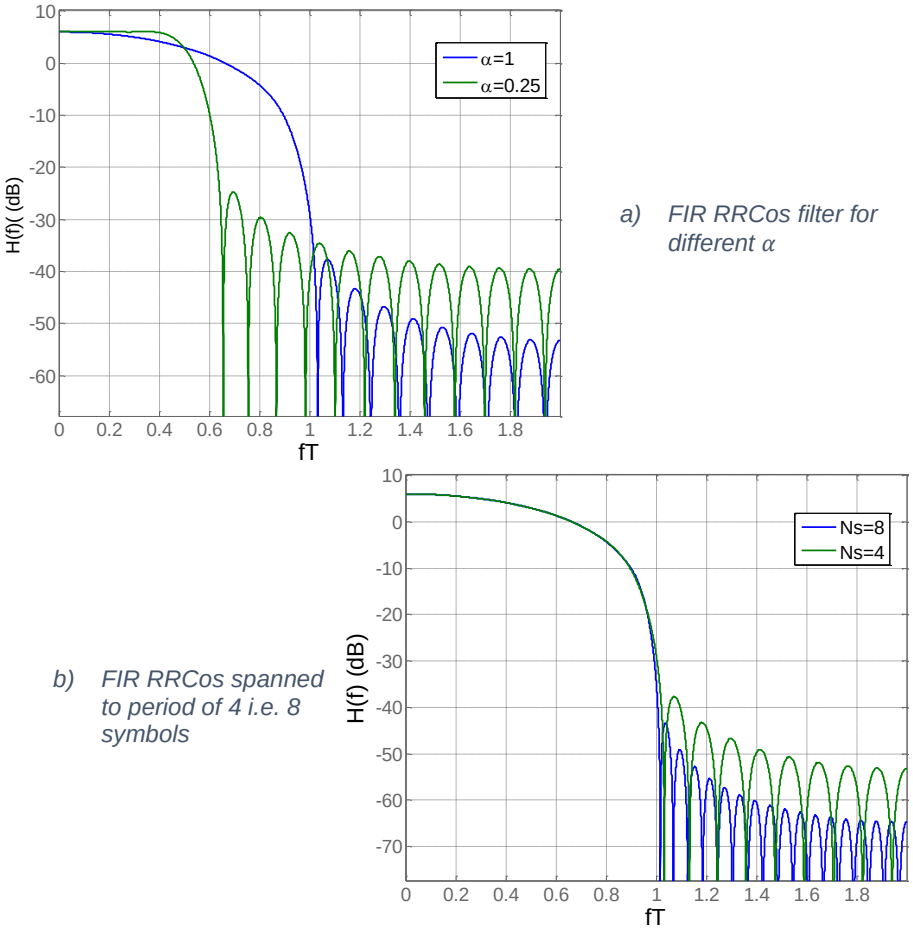
The degree to which the truncated filter response matches the ideal one, depends on oversampling factor N_{OS} and the number of filter taps N. Since N is generally chosen to be an integer multiple of N_{OS}, we get

$$N = N_{OS} \cdot N_S \,, \tag{2.6}$$

where N_S represents the number of symbols spanned by the impulse response of the filter.

Figure 2-7 depicts the magnitude response of a RRCos filter as result of truncation in the time domain. It can be seen in a) that the amplitude of the out-of-band side lobes increases as α decreases. Hence, in order to maintain a given level of stop-band for smaller values of α, the truncated time span N_S of the filter has to be increased. As shown in b), for larger N_S we will have better approximation of the ideal filter response. However, since computational complexity of the filter is directly proportional to N_S, it make sense to use the smallest possible value of N_S that meets the desired filtering requirements (minimum stop-band attenuation, maximum acceptable pass-band ripple, desired range of α values).

At last, it has to be mentioned that the oversampling factor N_{OS} impacts filter performance by reducing the variation of stop-band attenuation for small changes in N_S and α.



a) FIR RRCos filter for different α

b) FIR RRCos spanned to period of 4 i.e. 8 symbols

Figure 2-7: Magnitude response of RRCos shapes implemented as FIR filters

2.3. Noise Model

On its way to the receiver, signals are exposed to a variety of impairments such as thermal noise, nonlinear distortions, fading effects etc. Thermal noise is usually modeled by *additive white Gaussian noise* (AWGN) and it will be in this thesis considered as the only source of impairments:

- Additive, because it adds to the useful signal
- White, because it has approximately flat PSD
- Gaussian, because it exhibits a normal distribution in the time domain, with an average value of zero

2.4. Signal Model

In satellite communications, due to the presence of line-of-sight conditions between transmitter and receiver, the only impairment to communication is an addition of thermal noise. Therefore, satellite links can be well modeled by AWGN channels. Under such conditions, the received baseband signal is affected only by AWGN so that, apart from the useful component, it will only include noise. Typical baseband signal model for satellite communications is illustrated by Figure 2-8.

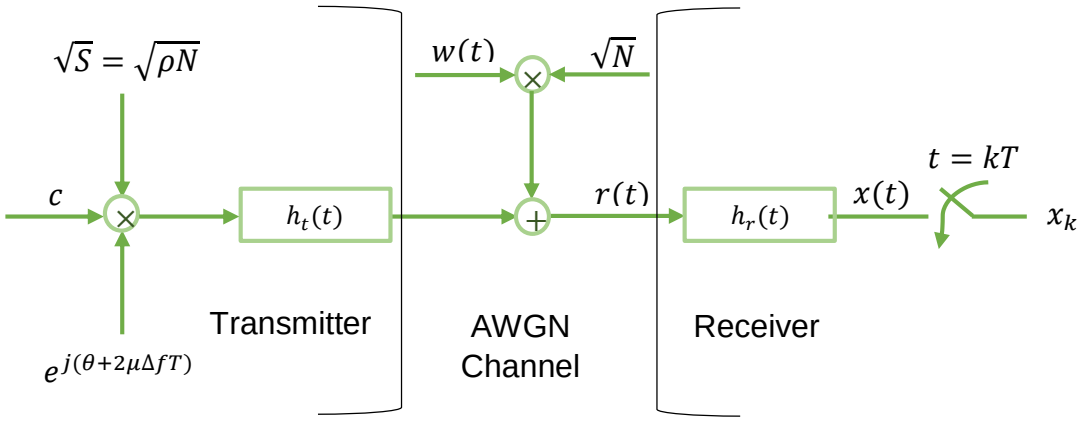


Figure 2-8: Baseband signal model in satellite communications

We consider the transmission of symbols $c_k = a_k + jb_k \in \mathbb{C}$, where $\mathbb{C}$ represents selected modulation scheme. Pulse shaping is achieved by a filter $h(t)$, which is a root-raised cosine functions in the frequency domain. Furthermore, the signal is transmitted with symbol rate $1/T$ over a frequency-flat channel and is impaired by carrier phase, carrier frequency and timing offsets denoted as θ, Δf and ε, respectively. Hence, the complex envelope of the received signal $r(t)$ can be modeled as:

$$r(t) = \sqrt{S} e^{j(2\pi\Delta f t+\theta)} \sum_i c_i h(t - iT - \varepsilon T) + \sqrt{N}\, w(t)\,, \tag{2.7}$$

where $w(t)$ stands for a circular-complex, zero-mean Gaussian process with unit variance. As can be seen from the Figure 2-8, the received signal $r(t)$ passes the matched filter (MF) $h_r(t) = {h_t}^*(-t)$. Since RRCos filters are real and even symmetric, we have that ${h_r}^*(-t) = h_r(t)$, i.e. $h_t(t) = h_r(t) = h(t)$. Hence, the MF output (t) is given as:

$$\begin{aligned} x_0(t) &\doteq r(t) \otimes h(t) \\ &= \left(\sqrt{S} e^{j(2\pi \Delta f t + \theta)} \sum_i c_i h(t - iT - \varepsilon T) + \sqrt{N}\, w(t) \right) \otimes h(t)\,, \end{aligned} \tag{2.8}$$

where $\otimes$ denotes convolution. Due to the properties of matched filtering, with $g(t) = h(t) \otimes h^*(-t)$, the signal part at the MF output sampled at $t = kT$, with $k \in \mathbb{Z}$, simply develops as

$$\begin{aligned} s_{0,k} &= \sqrt{S} e^{j(\theta + 2\pi \Delta f t)} \sum_i c_i h(t - iT - \varepsilon T) \otimes h(t)|_{t=kT} \\ &= \sqrt{S} e^{j(2\pi \Delta f t + \theta)} \sum_i c_i g(t - iT - \varepsilon T)|_{t=kT}\,. \end{aligned} \tag{2.9}$$

The corresponding noise samples, denoted as $n_{0,k}$, are given as

$$n_{0,k} = w(t) \otimes h(t)|_{t=kT}\,. \tag{2.10}$$

Finally, we have that

$$x_{0,k} = s_{0,k} + \sqrt{N}\, n_{0,k}\ . \tag{2.11}$$

3. Parameter Estimation

Say we have a discrete-time set of *N* observations $\{x[0], x[1], \dots, x[N-1]\}$ which depends on some deterministic, but unknown parameter θ. Our goal is to identify θ based only on the knowledge of the *N* observables, or to find some function f that delivers an estimate $\hat{\theta}$ of the related parameter. In this context, the function f is called an estimator [6]:

$$\hat{\theta} = f(x[0], x[1], \dots, x[N-1]) \,. \tag{3.1}$$

Assume that the true value θ of the parameter to be estimated, can take any value in the range (a, b). The estimator $\hat{\theta}$, on average converging to the true value θ, constitutes an unbiased estimate, i.e.

$$E\{\hat{\theta}\} = \theta, \quad \theta \in (a, b), \tag{3.2}$$

or equivalently, with $\boldsymbol{x} = (x[0], x[1], \dots, x[N-1])^T$ and $p(\boldsymbol{x}; \theta)$ denoting the PDF of the observations parameterized by the unknown parameter θ, we have [6]:

$$E\{\hat{\theta}\} = \int f(\boldsymbol{x}) p(\boldsymbol{x}; \theta) \, d\boldsymbol{x} = \theta \,. \tag{3.3}$$

On the other hand, an estimator that can never attain the true value of the unknown parameter is called biased:

$$E\{\hat{\theta}\} = \theta + b(\theta). \tag{3.4}$$

From Figure 3-1 one can see that the biased estimator deviates from the true value of the unknown parameter by $b(\theta)$, which is called bias of the estimator, i.e. $b(\theta) = \theta - E\{\hat{\theta}\}$, whereas under unbiased conditions the PDF is concentrated around the true value.

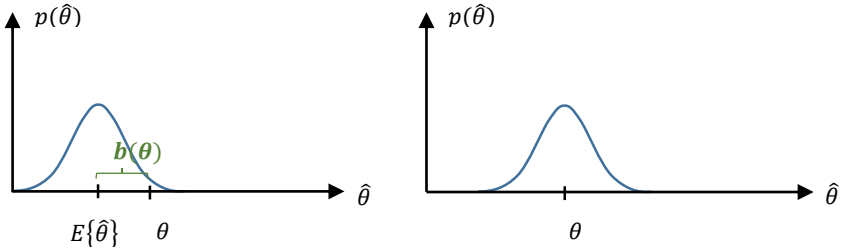


Figure 3-1: Biased vs. unbiased estimator

The unbiased property of an estimator is of the great importance, whenever several estimators are combined so that we have a set of estimates $\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n\}$. In order to get hopefully a better estimate of θ, it seems reasonable to average them. If all estimators are unbiased, uncorrelated with each other and of the same variance, we have that [6]

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_i \,, \tag{3.5}$$

$$\mathrm{var}(\hat{\theta}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(\hat{\theta}_i) = \frac{\mathrm{var}(\hat{\theta}_1)}{n} \,. \tag{3.6}$$

However, in case of biased estimators, when $E\{\hat{\theta}_i\} = \theta + b(\theta)$, combining them together will result also in a biased estimator:

$$E\{\hat{\theta}\} = \frac{1}{n} E\{\hat{\theta}_i\} = \theta + b(\theta) \,. \tag{3.7}$$

As it can be seen from Figure 3-2 that the averaging of estimators reduces the variance by factor $1/n$. This means that, if estimators are unbiased, we will end up definitely with a better estimate of θ, so that $\hat{\theta} \to \theta$ for $n \to \infty$. Unfortunately, this cannot be said for biased estimators, no matter how many independent estimates are available.

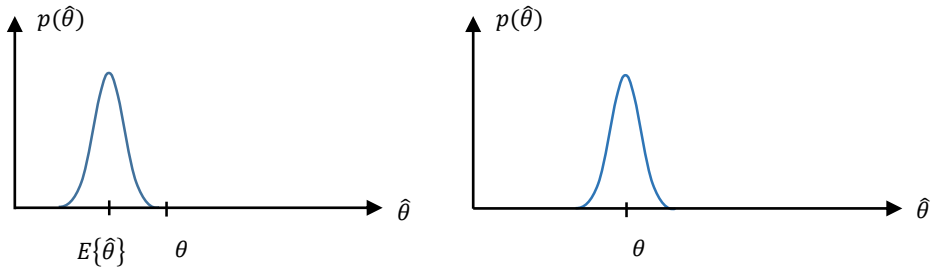


Figure 3-2: Combining biased vs. unbiased estimators

It has to be mentioned as well that an unbiased estimator guarantees only that it provides the true value of the desired parameter, but this does not necessary imply that related estimator is optimal. On the other hand, biased estimators are always poor estimators, since they introduce a systematic error.

Generally, there are two estimation approaches to extract parameters of interest from a set of observations [6], which is also illustrated in Figure 3-3:

- Classical approach
  - Maximum Likelihood Estimator (MLE)
  - Last Squares Estimators (LSE)
  - Moment-based Estimator
- Bayesian approach
  - Minimum Mean Square Error Estimator (MMSE)
  - Maximum a Posteriori Estimator (MAP)

In the classical approach the desired parameters are assumed to be deterministic but unknown, whereas Bayesian estimators come into play when *a priori* information can be exploited, so that unknown parameters are assumed to be no longer deterministic but randomly distributed.

Optimality of an estimation algorithm is predominantly measured in terms of its variance. In this context, an unbiased estimator is defined as optimal, if it has the smallest variance among all other estimators for all θ, so that the estimation error $\hat{\theta} - \theta$ is concentrated about zero. Such an estimator is denoted as minimum variance unbiased estimator (MVUE) [6]. Unfortunately, the MVUE does not always exists or it cannot be

found. Hence, in practice, we have deal with approximately optimal or suboptimal estimation algorithms.

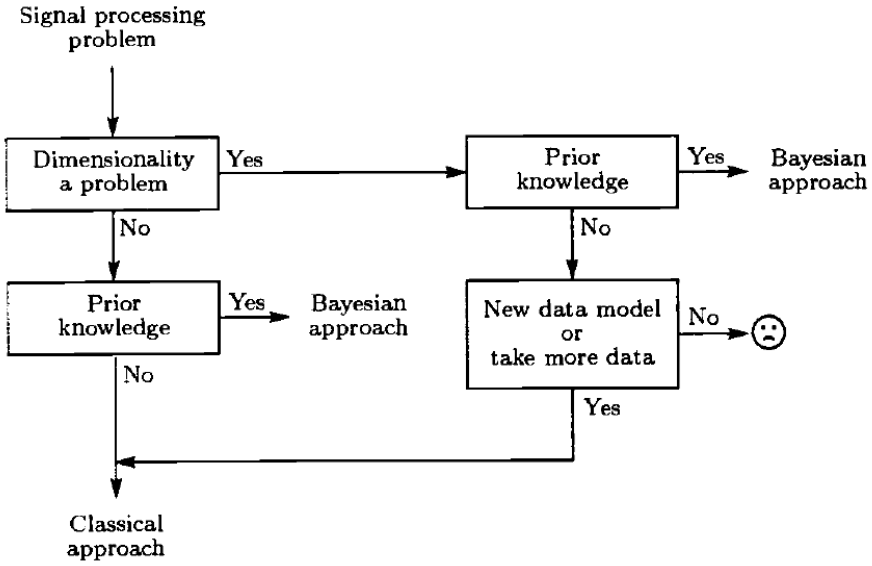


Figure 3-3: Decision-making process in estimator selection [6]

On the other hand, even if we can find an optimal estimator, it may not be efficient. In order to evaluate the efficiency of any bias-free estimator, it is considered to be optimal, approximately optimal or suboptimal. For this purpose, the so-called *Cramer-Rao lower bound* (CRLB) is introduced and widely used in this respect. The CRLB represents a theoretical limit for the achievable estimation accuracy embodied by the variance of any unbiased estimator. Whenever the error performance of the developed estimation algorithm coincides with the CRLB, then it is considered to be efficient:

$$\mathrm{var}\left(\hat{\theta}\right) \geq \mathrm{CRLB}(\theta)\,. \tag{3.8}$$

The CRLB is given by the inverse of the *Fisher information matrix* $FIM(\theta)$ according to

$$\mathrm{CRLB}(\theta) = \frac{1}{FIM(\theta)}\,. \tag{3.9}$$

With $\Lambda(\boldsymbol{x};\theta)$ denotes the log-likelihood function for the parameter to be estimated, i.e. $\Lambda(\boldsymbol{x};\theta) = \log p(\boldsymbol{x};\theta)$, *Fisher information matrix* is determined as

$$FIM(\theta) = E\left\{\left(\frac{\partial \Lambda(\boldsymbol{x};\theta)}{\partial \theta}\right)^2\right\} = -E\left\{\frac{\partial^2 \Lambda(\boldsymbol{x};\theta)}{\partial \theta^2}\right\} . \tag{3.10}$$

Figure 3-4 provides a graphical plot of the previous discussion: For the situation depicted as in a) $\hat{\theta}_1$ is an MVUE, but not efficient estimator for the parameter θ; if $\mathrm{var}(\hat{\theta}) = \mathrm{CRLB}$, as for $\hat{\theta}_1$ in b), such an estimator must be an MVUE and efficient; in c) none of the estimation algorithms is either MVUE or efficient, since $\hat{\theta}_1$ has better performance for $\theta > \theta_0$ and $\hat{\theta}_2$ for $\theta < \theta_0$.

Remark: due to simplification, all derivations in previous section were obtained under assumption that only one parameter has to be extracted from the observation set, but they can be extended when more than one parameter has to be estimated.

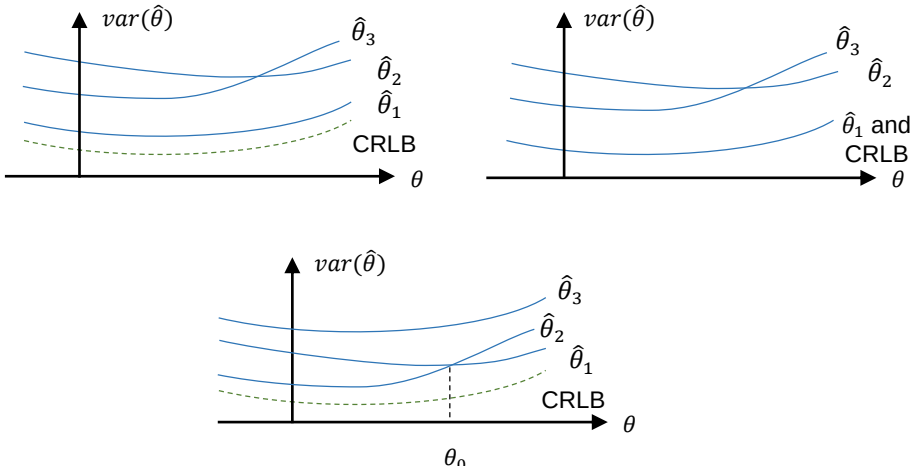


Figure 3-4: Evolution of the variance for different estimates

Choosing an appropriate estimation algorithm for a particular application depends on many factors. Generally, even if an optimal estimator can be found, it tends to be computationally complex, so that alternative, suboptimal but less complex procedures are perhaps preferable in practice. Hence, a tradeoff between computational complexity and performance has to be encountered for many practical situations.

In practice, the following properties are very desirable, when it comes to the development of an estimation algorithm [3]:

- Low computational complexity: Reducing the number of additive and multiplicative operations to a minimum is of great importance, since this decreases hardware requirements-
- Bias-free behavior: The algorithm itself should not introduce a systematic error, even if this might be corrected in case it is known in advance.
- Consistency: By increasing the length of observation, the variance of the developed estimation algorithm should decrease.
- Jitter variance as close as possible to the CRLB: The algorithm that is closer to the CRLB is the more efficient one.

3.1. Synchronization

As already mentioned, synchronization is a vital part of communication systems. Synchronization failures typically result in an inferior system performance. This section offers a brief overview of parameter estimation techniques used for data transmission without discussing specific algorithms. Assuming the signal model introduced previously, we have that

$$r(t) = \sqrt{S}e^{j(2\pi\Delta f t+\theta)} \sum_i c_i h(t - iT - \varepsilon T) + \sqrt{N}\, w(t)\,. \tag{3.11}$$

It becomes clear that the estimation of the parameters Δf, θ and ε is necessary for reliable detection of data, representing the carrier frequency/phase offset and the timing error, respectively.

Due to the presence of noise and nonlinear components in communication systems as well as the fact that multiple parameters have to be recovered, synchronization of transmission might be regarded as a nonlinear-stochastic and multi-dimensional optimization problem [3]. The Euclidian distance between the received signal and a test function is typically used as cost function. As a consequence of nonlinearities, the cost function will have several local optima and one global optimum [3], which is also illustrated in Figure 3-5.

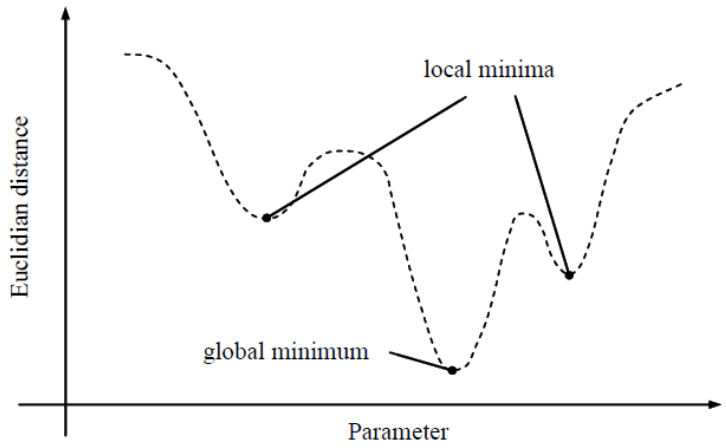


Figure 3-5: One- dimensional schematic of the synchronization process

In order to end up in the global optimum, the synchronization process is split in two phases:

- Initial acquisition
- Tracking

Acquisition represents the initial synchronization. In this phase, we are searching for a combination of parameters which results in a Euclidian distance close to the global minimum of the cost function. On the other hand, tracking corresponds to a loop controlling small deviations from the stable equilibrium point.

Furthermore, synchronization algorithms can be categorized as

- Feedforward algorithms
- Feedback algorithms

Feedback algorithms are almost exclusively used for tracking purposes, whereas initial acquisition is achieved by feedforward schemes because of convergence and stability constraints.

3.1.1. Frequency and Phase Errors

The presence of a frequency error has multiple effects on the transmitted signal. Unfortunately, the most severe effects are the distortion of the transmitted symbols and

generation of ISI, which becomes significant as the frequency error increases. A constant carrier frequency error corresponds to the rotation of the signal constellation, while a constant carrier phase error corresponds to the rotation of the constellation points by a fixed angle. If the parameters are estimated with sufficient accuracy, these effects can easily be compensated by a complex multiplier in baseband processing prior to detection, which results in an appropriate de-rotation of the received symbols [8]. Figure 3-6 demonstrates the effect of phase and frequency errors on the constellation of the received signal.

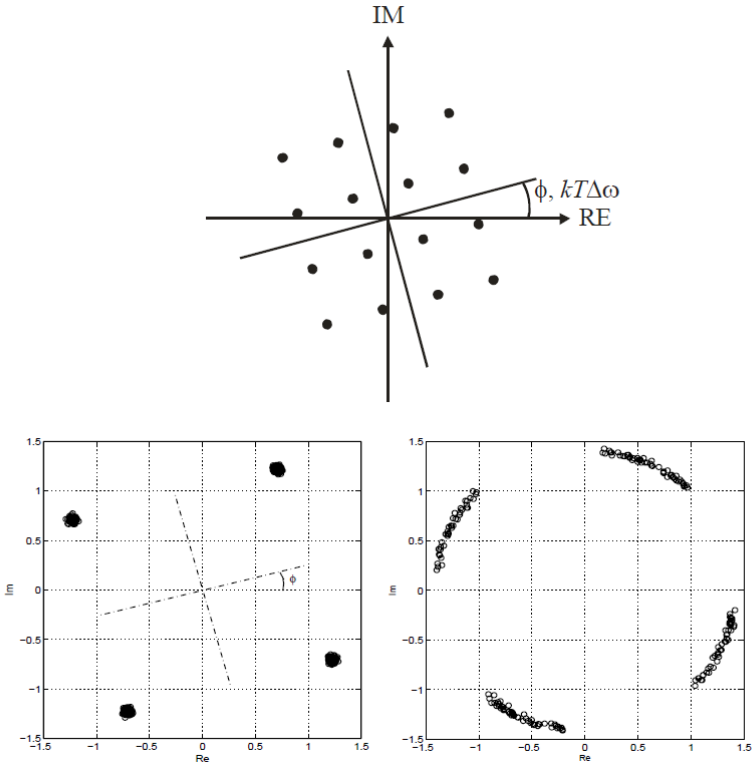


Figure 3-6: Effects of frequency and phase errors on a constellation diagram

3.1.2. Timing Error

In case of a symbol timing error, the samples are not taken at the maximum eye opening, which produces an ISI effect.

3.2. SNR Estimation

Estimation of the signal-to-noise ratio (SNR) raised more and more interest in recent past, since many existing and emerging communication systems require the SNR to be known for proper operation such as power control for adaptive coding and modulation (ACM) or iterative soft-decoding procedures. Figure 3-7 illustrates the basic principle of ACM, which is a special feature of the DVB-S2 standard [21].

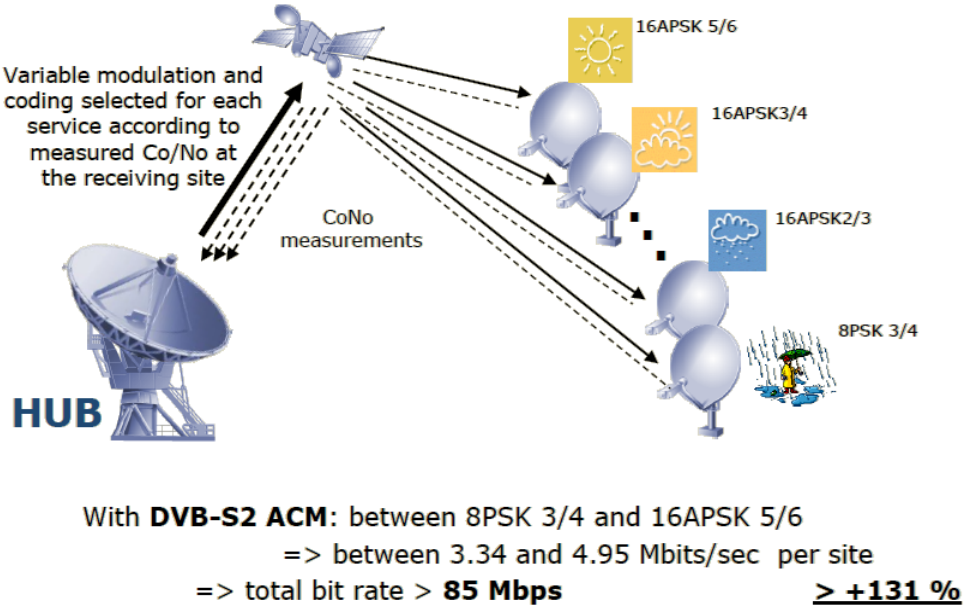


Figure 3-7: Adaptive coding and modulation (ACM) in DVB-S2 systems

ACM allows dynamical optimization of the transmission parameters by matching them according to the current channel conditions. For example, in order to enable system operation at low SNR (for bad link quality), the information rate has to be reduced by switching the current MODCOD (modulation scheme and coding) scheme to a lower modulation order and/or the coding rate.

3.2.1. Classification According to Operational Conditions

Existing SNR algorithms can be classified by a variety of criteria. It is common practice to distinguish between

- Data-aided (DA) estimation
- Decision-directed (DD) estimation

- Non-data-aided (NDA) estimation

DA estimation implies the knowledge of the unique word at the receiver side, whereas for DD estimations only a detected sequence is available. NDA estimators require neither training or pilot data nor decisions [19]. Obviously, DA techniques result in the best estimation performance, however some part of bandwidth i.e. data transmission capacity is lost for the training sequences. On the other hand, DD estimation techniques are sensitive to detection errors.

Under a different classification, SNR estimators can be

- Coherent
- Envelope-based

Coherent estimators exploit both the in-phase as well as quadrature components of the received signal, whereas envelope based estimators exploit only magnitude of the received signal.

Finally, SNR estimators can be categorized as

- Baudrate-sampled
- Oversampled

3.2.2. Classification and Assessment of Existing SNR Estimators

The following SNR estimators are described in [9]:

- Split-Symbol Moments Estimator (SSME)
  - In-service estimator which forms SNR estimation by using products and sums from the two halves of received data symbols.
  - The original SSME algorithm for a wide-band channel developed by Simon and Mileant, as well as its improved versions SSME and SSME for narrow-band channels are applicable only to BPSK-modulated signals in real AWGN channels.
- Maximum Likelihood (ML) Estimator
  - Based on ML estimation theory described before

  - Operates at the oversampled input to the MF
  - Suitable for M-PSK modulated signals in complex AWGN, but it can be extended to other coherent digital modulation schemes as well
- Squared Signal-to-Noise Variance (SNV) Estimator
  - Special case of the ML estimator operating at the optimally sampled MF output
- Second-and-Fourth-Order Moments (M2M4) Estimator
  - In-service estimator based on the calculation of signal moments which enables separate estimation of the signal and noise strength
  - No knowledge about carrier phase needed
- Signal-to-Variation (SVR) Estimator
  - NDA moment-based method
  - Originally developed for multipath fading channels, but also applied for M-PSK modulated signals in a complex AWGN channel
  - Carrier phase recovery is not required

4. Moment-Based SNR Estimation

As already mentioned before, moment-based estimators are categorized as NDA estimators, i.e. transmitted data (training and pilot sequences) do not have to be known in advance. In addition, since this sort of estimators is based on the evaluation of signal moments, knowledge of the carrier phase is not required.

The following section provides the principle of the estimator based on second- and fourth-order moments, which applies to complex AWGN channels. Using the signal model from the Chapter 2, we assume the following conditions:

- Symbols c_k are zero-mean, independent and identically distributed (i.i.d.), and normalized to unit variance such that $E\{|c_k|^2\} = 1$.
- Furthermore, frequency and timing errors are considered to be known, i.e. $\Delta f = \varepsilon = 0$, so that the received signal from (2.7) can be rewritten as:

$$r(t) = \sqrt{\rho N}\, e^{j\theta} \sum_i c_i h(t - iT) + \sqrt{N}\, w(t)\,. \tag{4.1}$$

Therefore, under the premises of perfect symbol timing recovery, i.e. $\tau = 0$, the MF output is simply given as:

$$x_{0,k} = \sqrt{S}\, e^{j\theta} c_k + \sqrt{N} n_{0,k} \tag{4.2}$$

4.1. Second- and Fourth-Order Moments Estimator

Estimators based on second- and fourth-order moments (M2M4) are of particular interest for SNR estimation, not at least because in addition to their robustness to phase variations, they are computationally very simple.

Bearing in mind that the MF output $x_{0,k}$ represents the input to an estimator, the second and fourth-order moments of $x_{0,k}$ are given by the subsequent equations:

$$\begin{aligned} M_2 &= E\{|x_{0,k}|^2\} \\ &= S\, E\{|c_k|^2\} + \sqrt{SN}\, E\{c_k n_{0,k}^*\} + \sqrt{SN}\, E\{c_k^* n_{0,k}\} + N\, E\left\{\left|n_{0,k}\right|^2\right\}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} M_4 &= E\{|x_{0,k}|^4\} \\ &= S\{|c_k|^4\} + 2S\sqrt{SN}\left(E\{|c_k|^2 c_k n_{0,k}^*\} + E\{|c_k|^2 c_k^* n_{0,k}\}\right) \\ &\quad + SN\left(E\{(c_k x_{0.k}^*)^2\} + 4E\left\{|c_k|^2 |n_{0,k}|^2\right\} + E\left\{\left(c_k^* n_{0,k}\right)^2\right\}\right) \\ &\quad + 2N\sqrt{SN}\left(E\left\{|n_{0,k}|^2 c_k n_{0,k}^*\right\} + E\left\{|n_{0,k}|^2 c_k^* n_{0,k}\right\}\right) \\ &\quad + N^2 E\left\{|n_{0,k}|^4\right\}. \end{aligned} \tag{4.4}$$

It has to be noticed that symbols c_k and noise samples $n_{0,k}$ are zero-mean and independent so that the relationships boil down to

$$M_2 = S + N \tag{4.5}$$

$$M_4 = K_c S^2 + 4\,SN + K_n N^2 \tag{4.6}$$

where K_c and K_n represent symbol and noise kurtosis, respectively, defined as $K_c = E\{|c_k|^4\}/E\{|c_k|^2\}^2$ and $K_n = E\left\{|n_{0,k}|^4\right\}/E\left\{|n_{0,k}|^2\right\}^2$. Solving this with respect to S and N, we obtain the estimates for signal and noise power:

$$\hat{S} = \frac{M_2(K_n - 2) \pm \sqrt{(4 - K_c K_n){M_2}^2 + M_4(K_c + K_n - 4)}}{K_c + K_n - 4} \tag{4.7}$$

$$\widehat{N} = M_2 - \hat{S}\,. \tag{4.8}$$

The symbol kurtosis is a function of the modulation scheme. On the other hand, the noise kurtosis depends whether the signal is transmitted over a real or a complex AWGN channel. Since we assume a complex AWGN channel with $K_n = 2$, the signal estimates are furnished as

$$\hat{S} = \frac{\sqrt{2{M_2}^2 - M_4}}{\sqrt{2 - K_c}}. \tag{4.9}$$

In case of M-PSK schemes, the symbol kurtosis $K_c = 1$, so that denominator in (4.9) is equal to one. Hence, M2M4 estimates applied on PSK signals in complex AWGN channel are finally given as:

$$\hat{\rho}_{M_2M_4} = \frac{\sqrt{2{M_2}^2 - M_4}}{M_2 - \sqrt{2{M_2}^2 - M_4}}. \tag{4.10}$$

Figure 4-1 depicts the normalized mean square error (NMSE) of the M2M4 estimates applied to QPSK for the total number of $K = 1024$ samples. As the theoretical benchmark for $\rho = \mathrm{SNR}$, we use the normalized CRLB for DA estimators given as

$$\mathrm{NCRLB}(\rho) = \frac{\mathrm{CRLB}(\rho)}{\rho^2} = \frac{1}{K}\left(1 + \frac{2}{\rho}\right). \tag{4.11}$$

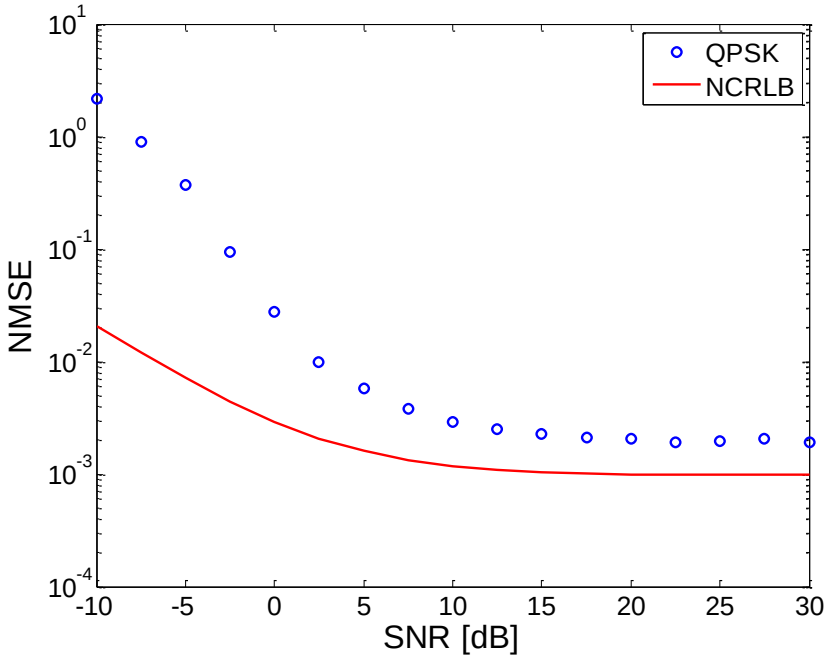


Figure 4-1: M2M4 estimator applied to QPSK (*K* = 1024)

As can been seen in Figure 4-1, the NCRLB itself does not decrease automatically with increasing SNR values; instead, we observe a floor effect in the medium-to-high SNR range. Note also that M2M4 estimators do not approach the CRLB even in asymptotical case, which is typical for envelope-based algorithms [4].

Unfortunately, when we apply the same procedure to 16-APSK or 16-QAM constellations, more or less useless results are provided in medium-to-high SNR region, which is exemplified in Figure 4-2. The reason for this is the jitter introduced through the varying nature of the data envelope.

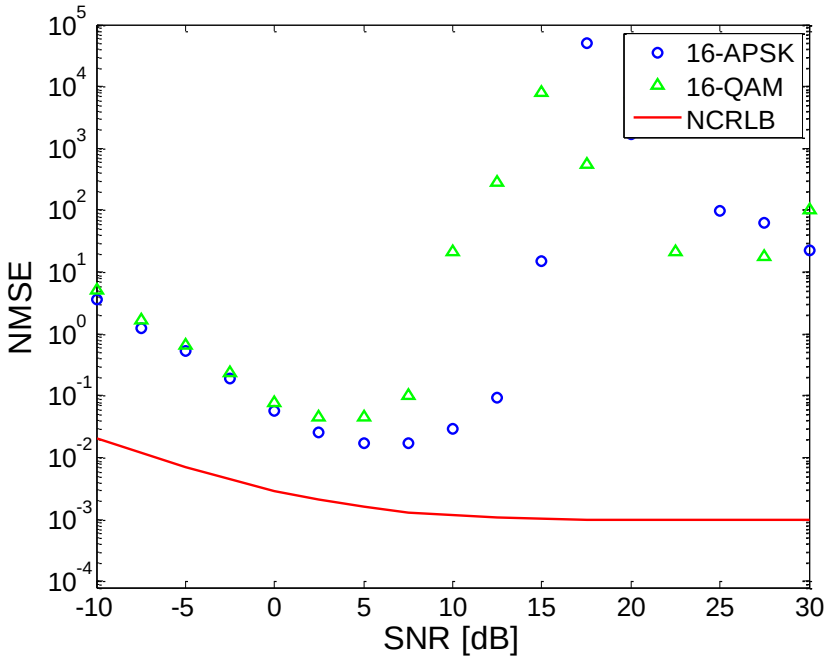


Figure 4-2: M2M4 estimator applied to 16-QAM/APSK (*K* = 1024)

4.2. Modified Estimator for Non-Constant Envelope Signals

The algorithm in [10] has been proposed to mitigate the non-negligible degradation of M2M4 estimation in medium-to-high SNR region, which arises because of the jitter introduced by the varying data amplitude. The idea is to modify the conventional M2M4 algorithm so that, by proper scaling, each of the related non-constant envelope constellation can be partitioned into domains containing only symbols with the same amplitude. In this way, the estimation procedure is reduced to PSK-like partitions.

Figure 4-3 depicts the basic idea applied to 16-APSK and 16-QAM. Blue dots represent the transmitted symbols, black dashed circles with radii R_A, R_B and R_C depict the related amplitude levels of the selected modulation scheme and red circles, R_{AB} and R_{BC}, are the partitioning radii. For comparison, the QPSK situation is shown as well.

As can be seen, non-constant modulus constellations like 16-QAM or 16-APSK contain more than one amplitude level. Hence, dividing the related modulation scheme into three (16-QAM) or two (16-APSK) domains, this enables a unique assignment of samples for partitioned SNR estimation within areas, where the symbols might be considered as PSK-modulated.

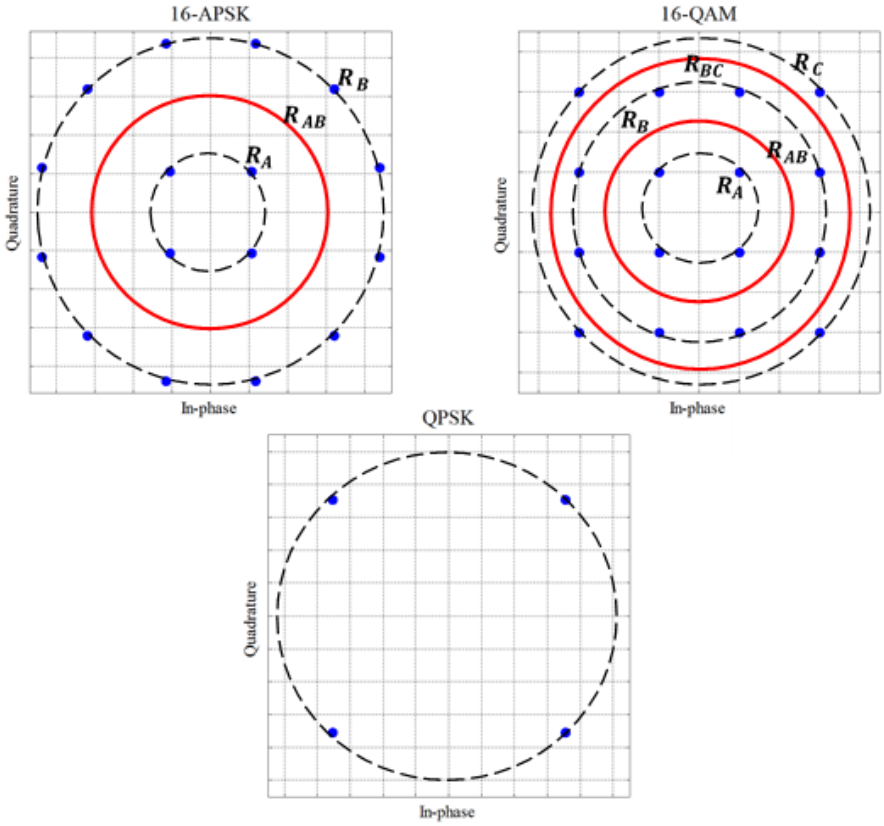


Figure 4-3: Partitioning of the signal space for 16-QAM/APSK schemes

In case of 16-QAM schemes, it is clear that four inner symbols constitute the domain A, i.e. $|c_k| = R_A$, eight center symbols constitute domain B, i.e. $|c_k| = R_{\boldsymbol{B}}$, and the rest of four symbols with $|c_k| = R_C$ represents domain C. Similarly, for 16-APSK signals, we have that the domain A consists of four symbols with $|c_k| = R_A$ and twelve outer symbols with $|c_k| = R_{\boldsymbol{B}}$ are forming domain B.

The most important task for the modified M2M4 algorithm is the proper choice of the partitioning radii R_{AB} and R_{BC}, since they are not known in advance. However, even though we have no knowledge about the true value of the signal power S , they can be computed if the signal power estimate $\hat{S}$ is achievable [10]. Fortunately, by virtue of the moment-based approach, $\hat{S}$ is established through (4.9). In order to evaluate the reliability of signal power estimates, it seems reasonable to compare it with related CRLB.

Assuming the model in (4.2), under the condition that no phase error occurred, the log-likelihood function of the signal power is given as [10]:

$$\Lambda(S,N) = -L \cdot \log(\pi N) - \frac{1}{N}\sum_{k=1}^{L}\left(|r_k|^2 - 2\sqrt{S}\,\mathrm{Re}[r_k c_k^*] + S|c_k|^2\right), \tag{4.12}$$

so that the variance of unbiased DA estimation of the signal power is lower bounded by

$$\mathrm{CRLB}(S) = \frac{E\left[\frac{\partial^2\Lambda}{\partial N^2}\right]}{E\left[\frac{\partial^2\Lambda}{\partial S\,\partial N}\right]^2 - E\left[\frac{\partial^2\Lambda}{\partial S^2}\right]E\left[\frac{\partial^2\Lambda}{\partial N^2}\right]}. \tag{4.13}$$

Therefore, the normalized CRLB can be written as

$$\mathrm{NCRLB}(S) = \frac{\mathrm{CRLB}(S)}{S^2} = \frac{1}{S^2 E\left[\frac{\partial^2\Lambda}{\partial S^2}\right]} = \frac{2}{K\rho}. \tag{4.14}$$

Figure 4-4 shows the signal power estimation performance in terms of NMSE for the M2M4 estimator. Simulation results for $K = 1024$ applied to QPSK, 16-QAM and 16-APSK modulation schemes are depicted. For comparison, the NCRLB(S) given by (4.14) is plotted as well. As expected, when applied to QPSK, the jitter variance of the signal power estimate approaches asymptotically the NCRLB, whereas for 16-QAM and 16-APSK an error floor emerges. However, it is shown in [10] that the jitter floor decreases with increasing observation lengths so that, by choosing K suitably large, we obtain reliable estimates by proper scaling and partitioning of non-constant modulus schemes.

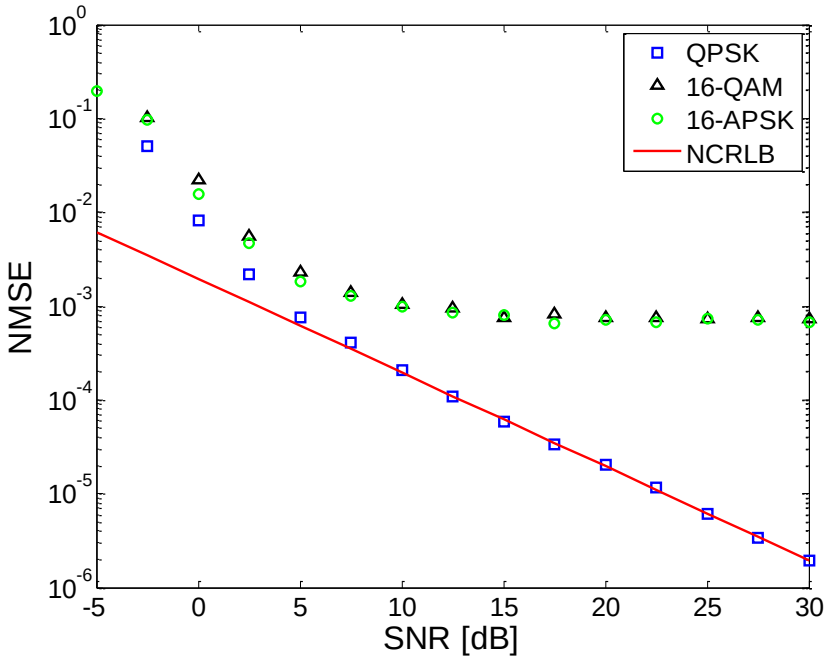


Figure 4-4: Moment-based estimation of the signal power applied to different modulation schemes (*K* = 1024)

Since partitioning radii R_{AB} and R_{BC} are functions of the symbol magnitudes, i.e. $|c_k| \in \{R_A, R_B\}$ for 16-APSK and $|c_k| \in \{R_A, R_B, R_C\}$ for 16-QAM and under the assumption that the modulation scheme is known in advance, one can choose the partitioning radii such that the probability of erroneously detected signal magnitudes c_k is minimized [10]:

$$R_{AB} = \min_R \{\Pr[|x_{0,k}| > R, |c_k| = R_A] + \Pr[|x_{0,k}| < R, |c_k| = R_B]\}, \tag{4.15}$$

$$R_{BC} = \min_R \{\Pr[|x_{0,k}| > R, |c_k| = R_B] + \Pr[|x_{0,k}| < R, |c_k| = R_C]\}. \tag{4.16}$$

From the statistical point of view, we have a Rician distribution, i.e. the PDF of the received signal's magnitude $|x_{0,k}|$ is given as:

$$p_r(s,x) = 2\rho x e^{-\rho(x^2+s^2)} I_0(2\rho s x), \tag{4.17}$$

with $\rho = SNR$ and $|c_k| = s$. Furthermore, $I_0(\cdot)$ denotes the modified zero-order Bessel function of the first kind and x is the random variable in radial direction. Therefore, the probabilities can be simply calculated as $\Pr[|c_k| = R_A] = \int_R^\infty p_r(R_A, x)\, dx$, $\Pr[|c_k| =$

$R_B] = \int_0^R p_r(R_B, x)\, dx$, so that the partitioning radius in case of 16-APSK can be obtained as

$$R_{AB} = R_A + \alpha_0(R_B - R_A) \tag{4.18}$$

representing a limit for unique assignment of the symbols within domains A and B. Similarly, for 16-QAM schemes, the partitioning radii are given as

$$R_{AB} = R_A + \alpha_0(R_B - R_A)\,, \tag{4.19}$$

$$R_{BC} = R_B + \alpha_1(R_C - R_B)\,. \tag{4.20}$$

In this context, α_0 and α_1 are partitioning parameters used for optimization purposes with respect to modulation schemes and SNR. Table 1 provides a listing of α_0 and α_1 obtained for 16-APSK and 16-QAM at different values of the selected SNR value.

SNR	16-APSK: α_0	16-QAM: α_0	16-QAM: α_1
10 dB	0.458	0.457	0.860
15 dB	0.486	0.485	0.614
20 dB	0.496	0.495	0.536
25 dB	0.499	0.498	0.511
∞	0.5	0.5	0.5

Table 1: Optimized partitioning parameters

Intuitively, it is clear that a problem may arise from the fact that, in order to properly optimize R_{AB} and R_{BC}, the optimization of the partitioning parameters α_0 and α_1 has to be available in advance, i.e. true SNR values have to be known at the receiver. Fortunately, by detailed inspection of Table 1, it can be observed that the parameters $\alpha_0 = \alpha_1 \to 0.5$ for $\mathrm{SNR} \to \infty$, so that these partitioning radii may be computed without knowledge of the true SNR. Therefore, equations (4.19) and (4.20) straightforwardly reduce to

$$R_{AB} = \frac{1}{2}(R_B + R_A)\,, \tag{4.21}$$

$$R_{BC} = \frac{1}{2}(R_C + R_B)\,. \tag{4.22}$$

Finally, it has to be mentioned that R_{AB} and R_{BC} have to be scaled by the estimate of the signal power $\hat{S}$, i.e. ${R_{AB}}' = \sqrt{\hat{S}}\, R_{AB}$ and ${R_{BC}}' = \sqrt{\hat{S}}\, R_{BC}$.

As already pointed out, all moment-based SNR estimators are functions of the sample moments. Hence, in practice the moments must be estimated by their time averages:

$$\widehat{M_2} \approx \frac{1}{K}\sum_{k=0}^{K}\left|x_{0,k}\right|^2, \qquad \widehat{M_4} \approx \frac{1}{K}\sum_{k=0}^{K}\left|x_{0,k}\right|^4. \tag{4.23}$$

4.2.1. Tuning of the 16-APSK Scheme

The 16-APSK constellation has two amplitude levels, so that only one partitioning radius ${R_{AB}}'$ has to be introduced for proper separation of the regions containing only symbols with constant amplitude.

For domain A, where $\left|x_{0,k}\right| \leq {R_{AB}}'$, second- and fourth-order moments are obtained as:

$$\widehat{M_2'} \approx \frac{1}{K'}\sum_{k=0}^{K'}\left|x_{0,k}\right|^2, \qquad \widehat{M_4'} \approx \frac{1}{K'}\sum_{k=0}^{K'}\left|x_{0,k}\right|^4. \tag{4.24}$$

The difference to (4.23) is that now only samples are considered that fall into the related region, which means that $K' \approx K/4$. Therefore, for SNR estimates related to domain A we get:

$$\hat{\rho}'_{16-APSK} = \frac{1}{R_A^2}\frac{\sqrt{2\widehat{M_2'}^2 - \widehat{M_4'}}}{\widehat{M_2'} - \sqrt{2\widehat{M_2'}^2 - \widehat{M_4'}}}. \tag{4.25}$$

Hence, for domain B, where $\left|x_{0,k}\right| > {R_{AB}}'$, with $K'' \approx 3K/4$, moments and SNR estimate are given by

$$\widehat{M_2''} \approx \frac{1}{K''}\sum_{k=0}^{K''}\left|x_{0,k}\right|^2, \qquad \widehat{M_4''} \approx \frac{1}{K''}\sum_{k=0}^{K''}\left|x_{0,k}\right|^4 \tag{4.26}$$

$$\hat{\rho}''_{16-APSK} = \frac{1}{R_B^2} \frac{\sqrt{2\widehat{M_2''}^2 - \widehat{M_4''}}}{\widehat{M_2''} - \sqrt{2\widehat{M_2''}^2 - \widehat{M_4''}}} \tag{4.27}$$

Figures 4-5 and 4-6 illustrate the simulation results for $K = 1024$, when the proposed estimation algorithm is applied to 16-APSK schemes with ring ratio[1] $R_B/R_A = 3.15$ for both non-perfect and perfect 16-APSK assignments of the symbols to the selected domain, and for cases when domain A or B is chosen for estimation. Perfect assignment assumes implementation of an ideal detector, so that the symbols c_k are known at the receiver. For comparison purpose, we use again the NCRLB given by (4.11).

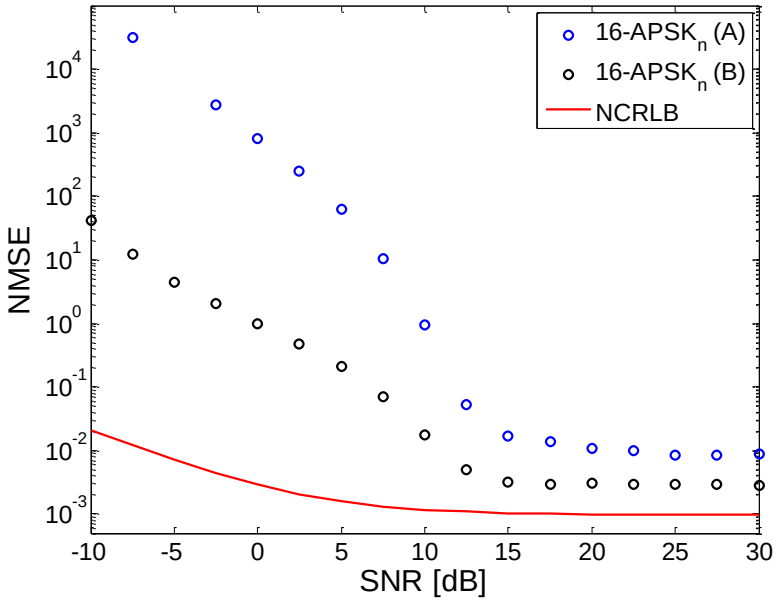


Figure 4-5 : Modified M2M4 estimation applied to 16-APSK with non-perfect assignment to partition A and B (*K* = 1024)

[1] The DVB-S2 standard specifies six possible ring ratios for 16-APSK schemes that range from 2.57 to 3.15 depending on the code rate.

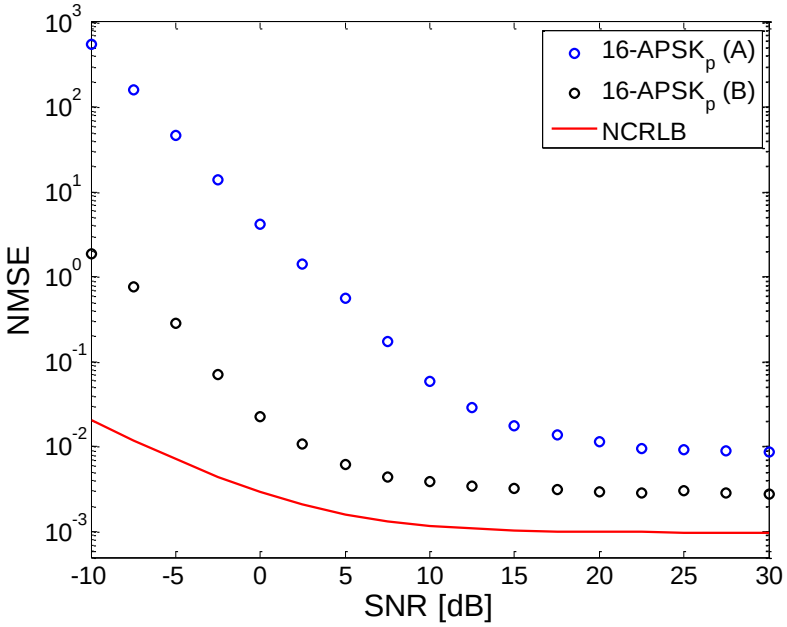


Figure 4-6: Modified M2M4 estimation applied to 16-APSK with perfect assignment to partition A and B (K = 1024)

As it can be seen, in terms of NMSE, both perfect as well as non-perfect assignment of symbols exhibit a similar behavior at higher SNR values, whereas for lower SNRs some performance degradation is observed. This loss arises due to an increased number of wrong assignments of samples to the corresponding domains.

Note also that choosing domain A instead of domain B results in a degraded performance. This has been expected, since on average the observation subset K' within domain A is smaller than K'' within domain B.

4.2.2. Tuning of the 16-QAM Scheme

In case of 16-QAM schemes, two partitioning radii R_{AB}' and R_{BC}' are introduced for proper partitioning into three regions. Therefore, second- and fourth-order moments as well as SNR estimates for domain A, B and C are given in the following.

For domain A: $\left|x_{0,k}\right| \leq R_{AB}'$, where $K' \approx K/4$:

$$\widehat{M_2'} \approx \frac{1}{K'}\sum_{k=0}^{K'}\left|x_{0,k}\right|^2, \qquad \widehat{M_4'} \approx \frac{1}{K'}\sum_{k=0}^{K'}\left|x_{0,k}\right|^4, \tag{4.28}$$

$$\hat{\rho}'_{16-QAM} = \frac{1}{R_A^2}\frac{\sqrt{2\widehat{M_2'}^2 - \widehat{M_4'}}}{\widehat{M_2'} - \sqrt{2\widehat{M_2'}^2 - \widehat{M_4'}}}. \tag{4.29}$$

For domain B: ${R_{AB}}' < \left|x_{0,k}\right| \leq {R_{BC}}'$, with $K'' \approx K/2$:

$$\widehat{M_2''} \approx \frac{1}{K''}\sum_{k=0}^{K''}\left|x_{0,k}\right|^2, \qquad \widehat{M_4''} \approx \frac{1}{K''}\sum_{k=0}^{K''}\left|x_{0,k}\right|^4, \tag{4.30}$$

$$\hat{\rho}''_{16-QAM} = \frac{1}{R_B^2}\frac{\sqrt{2\widehat{M_2''}^2 - \widehat{M_4''}}}{\widehat{M_2''} - \sqrt{2\widehat{M_2''}^2 - \widehat{M_4''}}}. \tag{4.31}$$

And finally, for domain C, where $\left|x_{0,k}\right| > {R_{BC}}'$ and $K''' \approx K/4$:

$$\widehat{M_2'''} \approx \frac{1}{K'''}\sum_{k=0}^{K'''}\left|x_{0,k}\right|^2, \qquad \widehat{M_4'''} \approx \frac{1}{K'''}\sum_{k=0}^{K'''}\left|x_{0,k}\right|^4, \tag{4.32}$$

$$\hat{\rho}'''_{16-QAM} = \frac{1}{R_C^2}\frac{\sqrt{2\widehat{M_2'''}^2 - \widehat{M_4'''}}}{\widehat{M_2''} - \sqrt{2\widehat{M_2'''}^2 - \widehat{M_4'''}}}. \tag{4.33}$$

Simulation runs are repeated for the 16-QAM scheme and $K = 1024$. In this case as well, non-perfect and perfect assignment of samples is investigated. Figures 4-7 and 4-8 show the evolution of NMSE when the estimation procedure is performed on domain A, domain B and domain C.

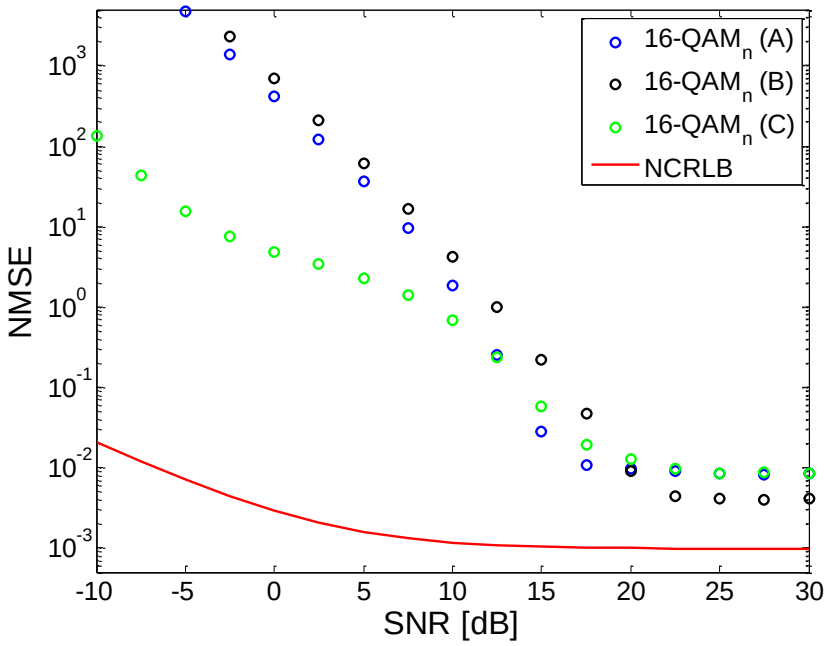


Figure 4-7 : Modified M2M4 estimation applied to 16-QAM with non-perfect assignment to partition A, B and C (K = 1024)

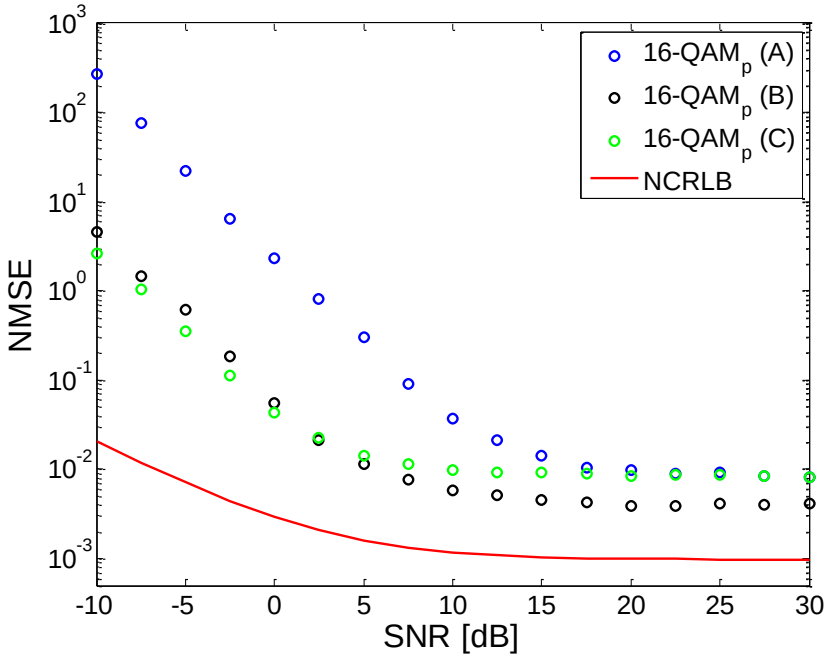


Figure 4-8 : Modified M2M4 estimation applied to 16-QAM with perfect assignment to partition A, B and C (K = 1024)

Finally, Figure 4-9 depicts the evolution of the mean estimator output. As it can be seen, in case of non-perfect assignment to partition B, some loss in performance for non-constant modulus schemes is only noticeable for SNR $\leq$ 10 dB. This arises due to increasing number of erroneous assignments when the SNR is not sufficiently high, i.e. related estimators are biased in the lower SNR range. In that context, if the SNR value where the performance begins to deteriorate, shall be decreased, then the number of samples used for estimation have to be increased [10].

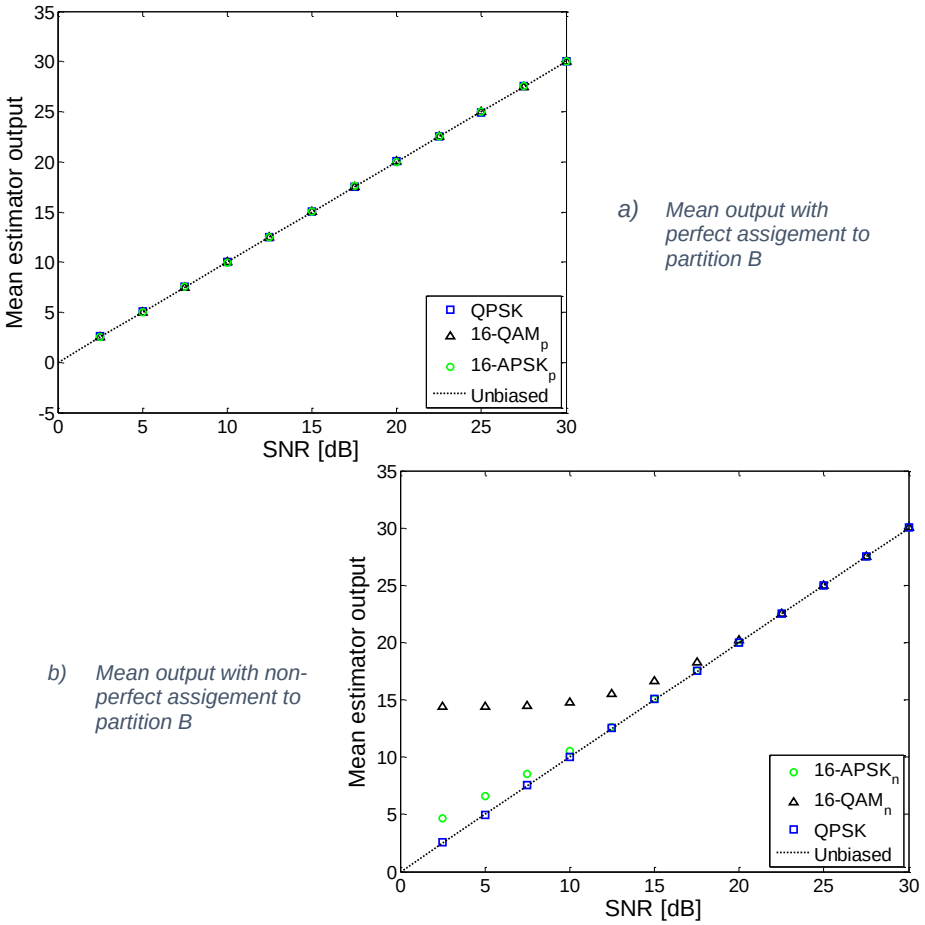


Figure 4-9: Mean estimator output for partition based SNR estimation applied to different modulation schemes (K = 1024)

5. Dual-Filter Framework

In this Chapter, completely new approach for SNR estimation will be presented, which has been proposed in [4] in case of linearly modulated signals. The basic idea is sketched in Figure 5-1: The received signal $r(t)$ passes not only the matched filter (MF), but also a second one operated in parallel to the MF. The outputs $x_0(t)$ and $x_1(t)$ represent two inputs to the SNR estimator, which is to be detailed in the sequel.

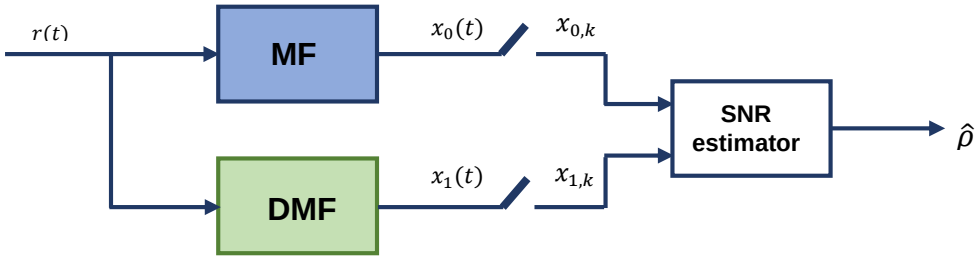


Figure 5-1: Dual-filter framework for SNR estimation

The MF output $x_{0,k}$ is given by (4.2). In order to calculate the output of the second filter, i.e. $x_{1,k}$, we have to define suitable properties the related filter must satisfy. By detailed inspection of the results obtained in the previous chapter, where advantages and disadvantages of different estimation procedures were discussed, we could make some conclusions about the filter design:

- First of all, if the new filter would produce signal as well as noise components at its output, the same non-negligible performance degradation would occur as it had been observed with M2M4 estimators. Therefore, we want to nullify all signal components, or in another words: the new filter must be designed in a way so that it generates only a noise component at its output.
- On the other hand, the goal is to gain an additional relationship between signal and noise power such that an SNR estimate can be extracted. Therefore, the equations on S and N have to be linearly independent, i.e. the second filter must not be the scaled version of the MF.

With this in mind, the desired output of the second filter, sampled at $t = kT$, yields

$$x_{1,k} \doteq r(t) \otimes h_1(t)|_{t=kT,\ \tau=0} = \sqrt{N}\, n_{1,k}\ , \tag{5.1}$$

where $h_1(t)$ denotes the impulse response of the second filter and $n_{1,k}$ represents the noise samples given as $n_{1,k} = n_1(kT)$, where $n_1(t) = w(t) \otimes h_1(t)$.

Since we wish to remove the signal part, the convolution between the auxiliary filter and the received signal must have zero-crossing at all sampling instants $t = kT$:

$$g_1(kT) = h(t) \otimes h_1(t)|_{t=kT} = 0\ . \tag{5.2}$$

It is clear that a RCos pulse satisfies this constraint for all samples, except for samples at instant $t = 0$. Hence, our next step is to find a proper modification which enables the cancellation of the signal part at $t = 0$.

Let assume that the desired relation is established in form of $g_1(t) = t \cdot g(t)$, where $g_1(t)$ represents the convolution of $h(t)$ with $h_1(t)$, which means that

$$h(t) \otimes h_1(t) = t \cdot g(t)\ . \tag{5.3}$$

Because solving (5.3) with respect to $h_1(t)$ in the time domain is fairly complicated, the problem will be shifted to the frequency domain:

$$H(f)H_1(f) = -\frac{1}{j2\pi}\frac{\partial G(f)}{\partial f}, \tag{5.4}$$

where $G(f)$, $H(f)$ and $H_1(f)$ are the Fourier transforms of $g(t)$, $h(t)$ and $h_1(t)$, respectively, i.e. $G(f) = \mathcal{F}\{g(t)\}$, $H(f) = \mathcal{F}\{h(t)\}$ and $H_1(f) = \mathcal{F}\{h_1(t)\}$. Bearing in mind that $G(f) = H^2(f)$, the transfer function of related filter boils down to

$$H_1(f) = -\frac{1}{H(f)} \cdot \frac{2H(f)}{j2\pi}\frac{\partial H(f)}{\partial f} = -\frac{1}{j\pi}\frac{\partial H(f)}{\partial f}\ . \tag{5.5}$$

Clearly, $H_1(f)$ is the first-order derivate of $H(f)$, therefore it make sense to denote it as derivate matched filter (DMF). Focusing on the DMF design, we can observe the following:

- From the fact that it is the first-order derivate of $H(f)$, which is band-limited to $\pm(1+\alpha)/2T$, it follows that the DMF is band-limited as well.
- Since the MF has a flat spectral support apart from the roll-off region, the DMF must be zero in related frequency range.
- With $G_1(f) = \mathcal{F}\{g_1(t)\}$, the zero-crossing condition can be rewritten as:

$$g_1(t) \cdot \sum_k \delta(t - kT) = 0 \quad \Leftrightarrow \quad G_1(f) \otimes \sum_k \delta\left(f - \frac{k}{T}\right) = 0\,. \tag{5.6}$$

It is clear from the first two statements that the DMF has no spectral support for $|f| < (1-\alpha)/2T$ as well as for $|f| > (1+\alpha)/2T$, which means that its overall transfer function is non-zero only inside of the roll-off band. Since only two replicas overlap in related band, we have from (5.6):

$$G_1(f) + G_1\left(f - \frac{1}{T}\right) = 0, \qquad \frac{1-\alpha}{2T} \le f \le \frac{1+\alpha}{2T}, \tag{5.7}$$

or equivalently:

$$G_1\left(f + \frac{1}{2T}\right) + G_1\left(f - \frac{1}{2T}\right) = 0, \qquad |f| \le \frac{\alpha}{2T}\,. \tag{5.8}$$

Intuitively, one observes that non-differentiability at the band edges $f = \pm(1+\alpha)/2T$ may raise serious implementation issues, so that the following section will be devoted to the detailed analysis of this particular problem.

5.1. DMF Shaping

Returning back to time domain, i.e. applying the inverse Fourier transform to (5.5), we obtain the impulse response of the DMF furnished by:

$$h_1(t) = \mathcal{F}^{-1}\{H_1(f)\} = 2t \cdot h(t)\,. \tag{5.9}$$

As already mentioned, when the MF output is sampled at symbol rate, aliasing will occur in the roll-off band only. Since the desired symbol-rate sequence at DMF output is aimed to compensate the corresponding distortion, the filter must not eliminate any

content in the related region [4]. Thus, assuming that $H_1(f) \neq 0$ inside of the roll-off band, we enforce the following normalization step:

$$\int_{-\infty}^{\infty} {h_1}^2(t)dt = 1\,. \tag{5.10}$$

Figure 5-2 shows the DMF (blue line) and the RRCos (green line) response in the time domain for $\alpha = 0.25$, when the related FIR filter is spanned over a period of $N_S = 21$ symbols. In Figure 5-3 depicts the evolution of the DMF for different values of α. It is important to observe that the impulse response of the DMF experiences a fairly slow roll-off behavior. This kind of behavior is expected, since, as mentioned previously, non-differentiability at $f = \pm(1+\alpha)/2T$ results in sharp edges of the corresponding filters transfer function. In addition, higher amplitude levels of side lobes is observed for smaller values of α.

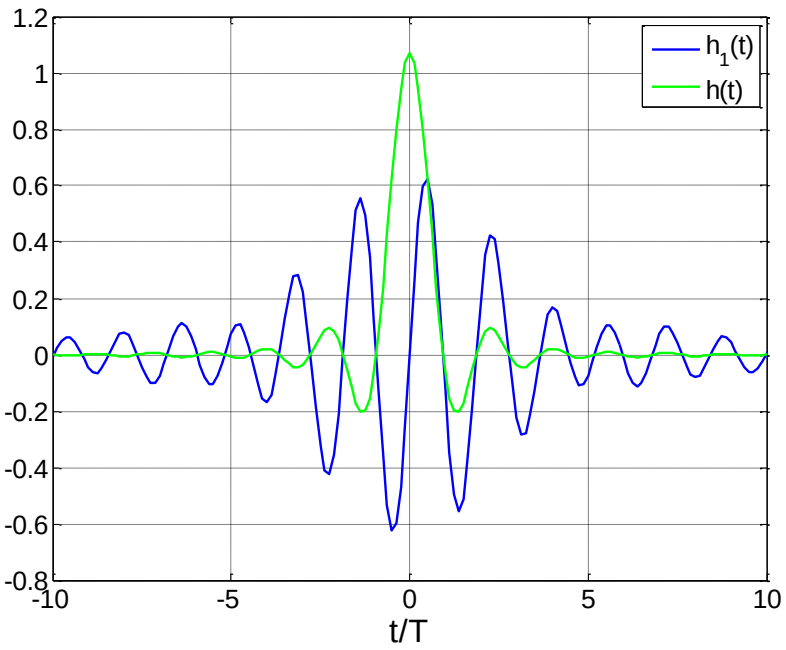


Figure 5-2 : Impulse response of DMF and RRCos shape for $\alpha = 0.25$

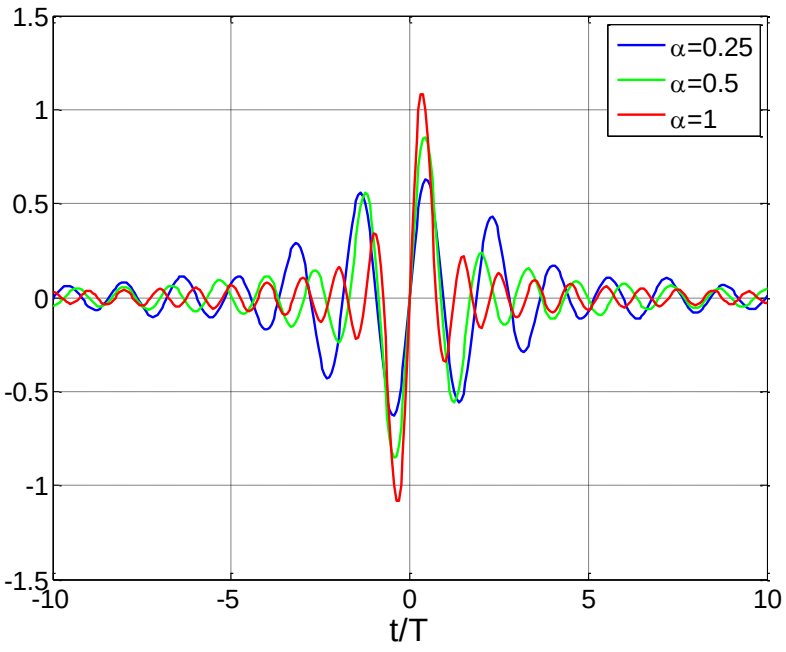


Figure 5-3 : Evolution of the DMF for different roll-off factors

Nevertheless, even for $\alpha = 1$, the DMF has to cover a large number of symbols periods, i.e. it is to be realized with a large number of taps. As already exemplified, noise power estimates degrade in the higher SNR regime, if the number of taps is not selected large enough.

For completeness, the evolution of $g_1(t)$ and $g(t)$ is depicted in Figure 5-4 for $\alpha = 0.25$. As it can be seen, the convolution of the DMF with the transmit pulse $h(t)$ results in the desired zero-crossing pattern, which confirms that $h_1(t)$ is a valid approach for signal cancellation purposes.

Finally, Figure 5-5 illustrates the magnitude response of MF and DMF for different values of α. It is to be noticed that the normalization step in (5.10) causes an increase of the amplitude of $|H_1(f)|$, when α decreases. Furthermore, as already discussed before, the ringing artifacts due to the Gibbs phenomenon can be observed at band edges.

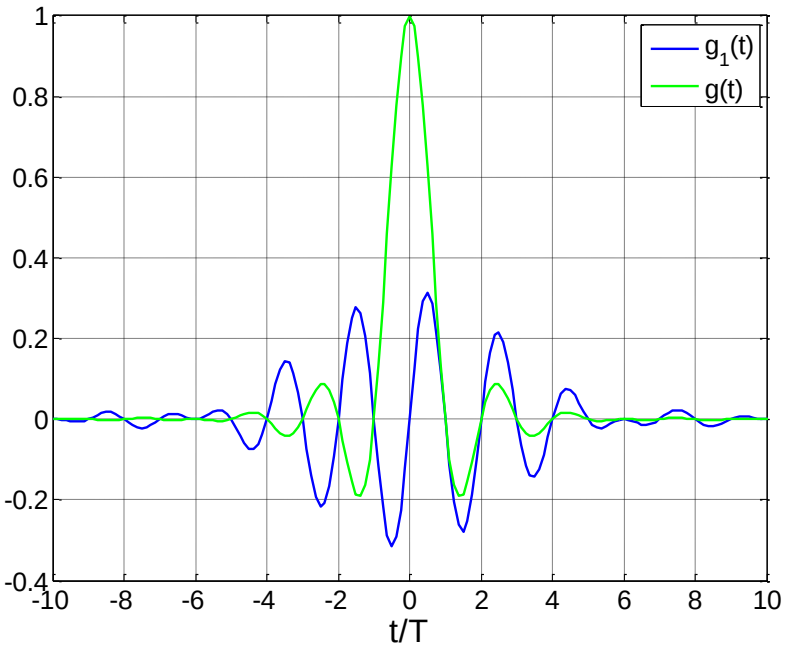


Figure 5-4 : Evolution of $g_1(t)$ and RCos pulse for $\alpha = 0.25$

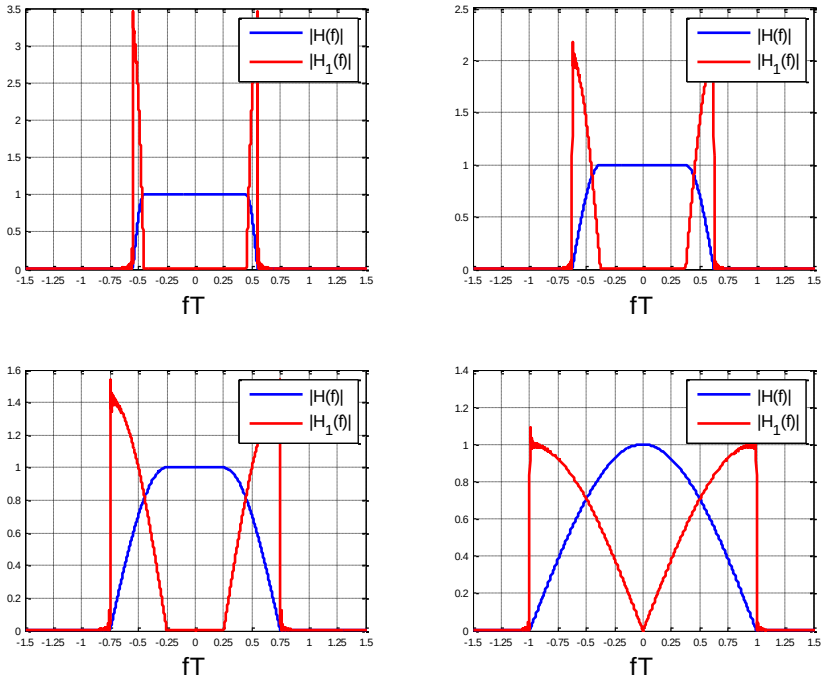


Figure 5-5 : Spectral evolution of MF and DMF for $\alpha = 0.1$ (top left), $\alpha = 0.25$ (top right), $\alpha = 0.5$ (bottom left) and $\alpha = 1$ (bottom right)

5.2. Requirements for the Additional Filter Function

As already mentioned, in order to prevent performance loss when using the dual-filter framework for SNR estimation, the DMF has to be realized with a large number of taps. Obviously this requirement comes to the price of a high computational complexity. However, by shaping the filter $H_1(f)$ via an additional filter function $P(f) \neq 0$, which avoids a sharp cut-off of $H_1(f)$ at $f = \pm(1+\alpha)/2T$, the overall computational load required by auxiliary filter can be significantly reduced as will be shown in the sequel.

Let $\overline{H_1}(f)$ denote the shaped version of the DMF $H_1(f)$, i.e.

$$\overline{H_1}(f) = H_1(f)\, P(f)\,. \tag{5.11}$$

In the following, we will focus on the requirements that $P(f)$ must satisfy in order to be a valid filter function in terms of DMF. Clearly, $P(f)$ must have the same spectral support as $H_1(f)$:

$$P(f) \neq 0, \qquad f < \left|\frac{1+\alpha}{2T}\right|. \tag{5.12}$$

It is necessary to design $P(f)$ such that the zero-crossing property is not violated, which means that

$$h(t) \otimes \overline{h_1}(t)|_{t=kT} = 0\,, \tag{5.13}$$

where $\overline{h_1}(t)$ denotes the inverse Fourier transform of $\overline{H_1}(f)$. With $G_1(f) = H(f)H_1(f) \neq 0$ for $(1-\alpha)/2T < |f| < (1+\alpha)/2T$, the requirement in (5.8) can be rewritten as

$$\overline{G_1}\left(f + \frac{1}{2T}\right) + \overline{G_1}\left(f - \frac{1}{2T}\right) = 0\,, \qquad |f| \leq \frac{\alpha}{2T} \tag{5.14}$$

which is equivalent to

$$P\left(f-\frac{1}{2T}\right)G_1\left(f-\frac{1}{2T}\right)+P\left(f+\frac{1}{2T}\right)G_1\left(f+\frac{1}{2T}\right) = \left[P\left(f-\frac{1}{2T}\right)-P\left(f+\frac{1}{2T}\right)\right]\cdot G_1\left(f-\frac{1}{2T}\right)=0\,. \tag{5.15}$$

Since we know that $G_1(f)$ must be non-zero in the roll-off band, this leads to a necessary property required when choosing $P(f)$ [1]:

$$P\left(f+\frac{1}{2T}\right)-P\left(f-\frac{1}{2T}\right)=0, \qquad |f|\le\frac{\alpha}{2T}\,. \tag{5.16}$$

5.2.1. Special Cases

Considering the shape of our window function $P(f)$, we can make the following observations with respect to $H_1(f)$:

- No shaping of $H_1(f)$ implies that $P(f)=1$.
- For real-valued $H_1(f)$ shapes, $P\left(f\pm\frac{1}{2T}\right)$ is an even-symmetric function, i.e. $P\left(f+\frac{1}{2T}\right)=P\left(f-\frac{1}{2T}\right)$.
- One of the possible choices for $P(f)$, suggested in [4], is to use a raised cosine shape $H_1(f)$ given as:

$$P\left(f\pm\frac{1}{2T}\right)=\cos^2\left(\frac{\pi fT}{\beta}\right), \quad |f|<\frac{\beta}{2T}\,,\beta\le\alpha. \tag{5.17}$$

5.2.2. Shaping with Raised Cosines

By inspection of the Figure 5-6, it can be noticed that when $H_1(f)$ is shaped via $P(f)$ specified by (5.17), the DMF can be realized by a FIR filter spanned over a much shorter window compared to the non-shaped version of the DMF. Therefore, employing (5.17) for shaping, this reduces significantly the required computational load when implementing a DMF function. The bottom plot in Figure 5-6 confirms that raised cosines are an appropriate choice for $P(f)$, since the resulting impulse response in convolution with the transmitter pulse provides the desired zero-crossing pattern.

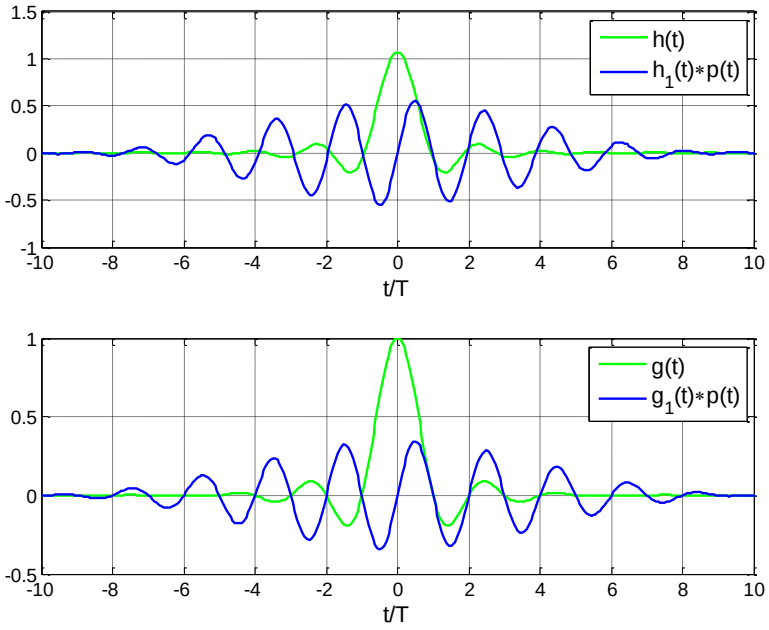


Figure 5-6 : Evolution of $\overline{h_1}(t)$ and RRCos i.e. $\overline{g_1}(t)$ and RCos for $\alpha = \beta = 0.25$

Accordingly, as it can be seen from Figure 5-7, the transfer function $\overline{H_1}(f)$ of such a filter is continuous at the band edges. Therefore, we do not have to deal any more with the ringing artifacts that arise due to abrupt transition between pass-band and stop-band. However, it is shown in subsequent chapters that the estimation performance slightly degrades compared to the non-shaped version of the DMF, which is apparently the price to be paid for the reduced computational complexity.

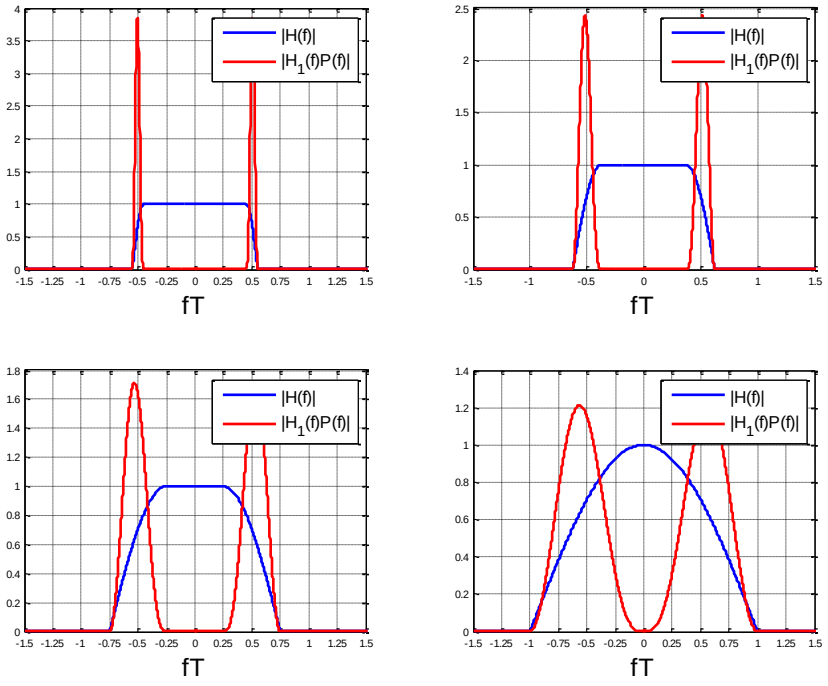


Figure 5-7 : Spectral evolution of MF and DMF shapes with additional filter function $P(f \pm 1/2T) = \cos^2(\pi fT/\alpha)$ for $\alpha = 0.1$ (top left), $\alpha = 0.25$ (top right), $\alpha = 0.5$ (bottom left) and $\alpha = 1$ (bottom right)

6. SNR Estimation of Oversampled Linearly Modulated Signals

In communication scenarios with linear modulation schemes, it is common practice to use estimators based on baudrate samples available at MF output. As already mentioned previously, unless excess bandwidth is zero, i.e. $\alpha = 0$, this approach will normally lead to aliasing effects. Although these output samples do not preserve all the information about the original signal, it turns out that under AWGN conditions they provide sufficient statistics for data detection[2]. In the sequel it will be shown that the avoidance of alias effects, by oversampling the received signal, can improve SNR estimation performance.

Estimation analysis for the oversampled scenario can be considerably simplified, if we employ the dual-filter framework introduced in Chapter 5. In this way, instead of to directly dealing with the oversampled MF output, we can exploit the baudrate samples of both filters, so that only the useful signal band is processed. It is shown that these outputs preserve all information about the received signal [1] [4].

6.1. Employing the Dual-Filter Framework

When employing an auxiliary filter in parallel to the MF for estimation purpose, it is very important to notice following facts:

- From (4.2) it should be clear that the process $\{x_{0,k}\}$ at the MF output is zero-mean Gaussian with statistically independent samples $x_{0,k}$.
- On the other hand, the DMF output $\{x_{1,k}\}$ is a colored zero-mean Gaussian process with autocorrelation $R_{n_1}(mT) = E\{n_{1,k} n_{1,k-m}{}^*\}$.
- Due to (5.2), the noise processes $\{n_{0,k}\}$ and $\{n_{1,k}\}$ are uncorrelated, i.e. $E\{n_{0,k} n_{1,k}{}^*\} = 0$, so that the output samples $x_{0,k}$ and $x_{1,k}$ are statistically independent. Hence, the cross-correlation can be expressed by the inverse Fourier transform of $H(f)\bar{H}_1(f)$, respectively.

[2] An exception is the problem of symbol timing estimation, which requires alias-free samples

- The normalization step imposed by (5.10.) implies that $R_{n_1}(0) = 1$.

The $K \times K$ autocorrelation matrix $\widetilde{\boldsymbol{C}}$ of the nose process $\{n_{1,k}\}$ is given by:

$$\widetilde{\boldsymbol{C}} = E\{\boldsymbol{n_1}\boldsymbol{n_1}^H\} = \begin{bmatrix} R_{n_1}(0) & \cdots & R_{n_1}{}^*((K-1)T) \\ \vdots & \ddots & \vdots \\ R_{n_1}((K-1)T) & \cdots & R_{n_1}(0) \end{bmatrix}, \tag{6.1}$$

with K as the total length of observations. In this context, $\boldsymbol{n_1}$ is the $1 \times K$ vector that collects all noise samples $n_{1,k}$ at the DMF output, i.e. $\boldsymbol{n_1} = [n_{1,1}, n_{1,2}, \ldots, n_{1,K}]^T$. Applying the Fourier transform to $R_{n_1}(mT)$, we get for the PSD of the noise process :

$$S_{n_1}\left(e^{j2\pi fT}\right) = \sum_k R_{n_1}(kT)\, e^{-j2\pi kfT}\,. \tag{6.2}$$

Due to the fact that $R_{n_1}(\tau) = \bar{h}_1(\tau) \otimes \bar{h}_1{}^*(-\tau)$, which is the inverse Fourier transform of $|\bar{H}_1(f)|^2$, the PSD of $\{n_{1,k}\}$ is completely given by only exploiting the transfer function of the DMF:

$$S_{n_1}\left(e^{j2\pi fT}\right) = \frac{1}{T}\sum_n \left|\bar{H}_1\left(f - \frac{n}{T}\right)\right|^2, \tag{6.3}$$

with $n \in (0,1)$, since we have only two replicas that overlap in the roll-off band. It is to note that since $|\bar{H}_1(f)|^2 = 0$ for $|f| < (1-\alpha)/2T$ and $|f| > (1+\alpha)/2T$.

Therefore, when using the eigenvalue decomposition of $\widetilde{\boldsymbol{C}}$, i.e.

$$\widetilde{\boldsymbol{C}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\boldsymbol{H}}\,, \tag{6.4}$$

the autocorrelation matrix of the noise process $\{n_{1,k}\}$ can be simply obtained from the amplitude of each spectral component of $|\bar{H}_1(f)|^2$ and from the frequency position to which the related component is assigned. The former ones, called eigenvalues of $\widetilde{\boldsymbol{C}}$, are collected in $K \times K$ diagonal matrix $\boldsymbol{\Lambda}$, such that

$$\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_0, \lambda_1 \ldots \lambda_{K-1}\}\,. \tag{6.5}$$

Since the autocorrelation matrix $\widetilde{\boldsymbol{C}}$ is Hermitian Toeplitz, each term of $\boldsymbol{\Lambda}$, i.e. the eigenvalues of $\widetilde{\boldsymbol{C}}$, can be approximated for $K \to \infty$ as [1]

$$\lambda_k = S_{n_1}\left(e^{j2\pi k/K}\right). \tag{6.6}$$

As already mentioned, the $K \times K$ unitary matrix $\boldsymbol{U}$ contains the position of each spectral component of $|\bar{H}_1(f)|^2$ given as

$$\boldsymbol{U} = \frac{1}{\sqrt{K}}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{-j2\pi/K} & e^{-j4\pi/K} & \cdots & e^{-j2\pi(K-1)/K} \\ 1 & e^{-j4\pi/K} & e^{-j8\pi/K} & \cdots & e^{-j4\pi(K-1)/K} \\ 1 & e^{-j6\pi/K} & e^{-j12\pi/K} & \cdots & e^{-j6\pi(K-1)/K} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & e^{-j2\pi(K-1)/K} & e^{-j4\pi(K-1)/K} & \cdots & e^{-j2\pi(K-1)(K-1)/K} \end{bmatrix}. \tag{6.7}$$

From (6.6) is clear that, as $K \to \infty$, $\widetilde{\boldsymbol{C}}$ becomes rank deficient with $L = \operatorname{rank}(\widetilde{\boldsymbol{C}}) = \alpha K$. Therefore, we can use the "economy size" eigenvalue decomposition of $\widetilde{\boldsymbol{C}}$**,** i.e. $\widetilde{\boldsymbol{C}} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\boldsymbol{H}}$, where the $L \times L$ matrix $\widetilde{\boldsymbol{\Lambda}}$ collects only the nonzero eigenvalues and the columns of the $K \times L$ matrix $\widetilde{\mathbf{U}}$ are the corresponding eigenvectors [1].

6.1.1. DA Case

Using the dual-filter framework, the Cramer-Rao lower bound for the DA SNR estimation is given as [1]

$$\mathrm{CRLB}(\rho) = \frac{\rho}{K}\left(\frac{\rho}{1+\alpha} + 2\right). \tag{6.8}$$

Comparing the relationship with (4.11), one can see that the bounds coincide for $\alpha = 0$. This has been expected, since in this particular case the alias-free output of the MF preserves all information about received signal.

On the other hand, the information loss by sampling the received signal below Nyquist rate can be retrieved through the DMF output. The gap decreases as α decreases, i.e. for small α the benefit contributed by using the DMF becomes less significant. This might be expressed by

$$\lim_{\rho\to\infty}\frac{\mathrm{CRLB}(\rho)}{\mathrm{CRLB}_{\mathrm{MF}}(\rho)}=\frac{1}{1+\alpha}\in\left[\frac{1}{2},1\right]. \tag{6.9}$$

For simulation purposes, DA ML estimator adopted to the dual-filter model will be used. In this context, the estimates of phase, signal and noise power are given as:

$$\hat{\theta}=\arg\{\boldsymbol{c}^H\boldsymbol{x_0}\},\qquad \hat{S}=\frac{|\boldsymbol{c}^H\boldsymbol{x_0}|^2}{K^2}, \tag{6.10}$$

$$\widehat{N}=\frac{\|\boldsymbol{x_0}\|^2-K\hat{S}+\boldsymbol{x_1}^H\widetilde{\boldsymbol{C}}^{\#}\boldsymbol{x_1}}{K(1+\alpha)}, \tag{6.11}$$

where $\boldsymbol{c}$, $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ represent the length-*K* vectors of transmitted symbols, MF and DMF output, respectively. Furthermore, $\widetilde{\boldsymbol{C}}^{\#}$ denotes the pseudo-inverse of $\widetilde{\boldsymbol{C}}$. Hence, related SNR estimates are given as $\hat{\rho}=\frac{\hat{S}}{\widehat{N}}$.

When shaping the DMF via a raised cosine filter function from, it has to be noticed, that $\widetilde{\boldsymbol{\Lambda}}$ collects eigenvalues of $\widetilde{\boldsymbol{C}}$ that are due to $|\bar{H}_1(f)|^2$ very close to zero. Unfortunately, this can affect the estimation performance, since $\widetilde{\boldsymbol{C}}^{\#}$ tends towards infinity. In order to mitigate the related issue, the first idea was to add a floor with respect to α, however, it turned out that this solution resulted in a severely biased ML estimator.

Figures 6-1 and 6-2 illustrate the jitter performance of the normalized SNR estimates as a function of the true SNR value. A sequence of $K=1024$ symbols is used for simulation purposes. Spanning a period of $N_s=81$ symbols, no additional shaping is applied to the DMF. As can be seen, for low SNR the availability of the DMF output does not improve the estimation performance. However, as already stated before, in the medium-to-high SNR region the benefit of DMF is obvious. Moreover, as α increases, the additional information about the excess band, contributed by the DMF, becomes more relevant. The mismatch in simulation results can be explained by the fact that, in order to reduce hardware requirements, estimates are averaged by a reduced number of iterations.

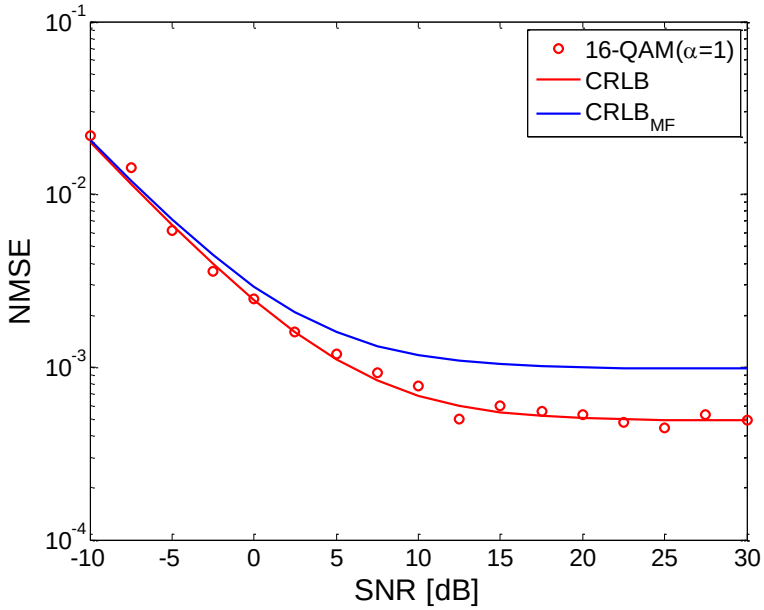


Figure 6-1: DA ML estimator applied to 16-QAM for $H_1(t)$ and $\alpha = 1$

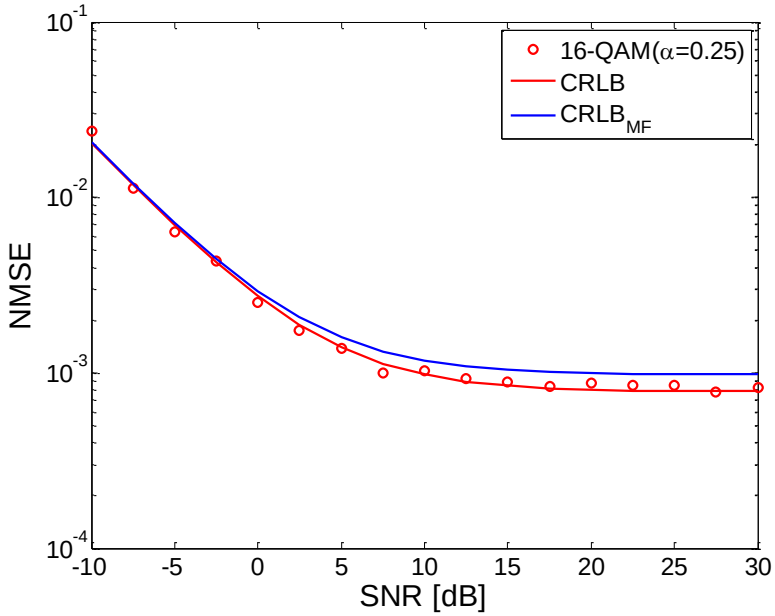


Figure 6-2: DA ML estimator applied to 16-QAM for $H_1(t)$ and $\alpha = 0.25$

The simulation runs are repeated also for 16-APSK modulation scheme and depicted in Figures 6-3 and 6-4. As expected, since the DA procedure was performed, we get similar results as in case of 16-QAM.

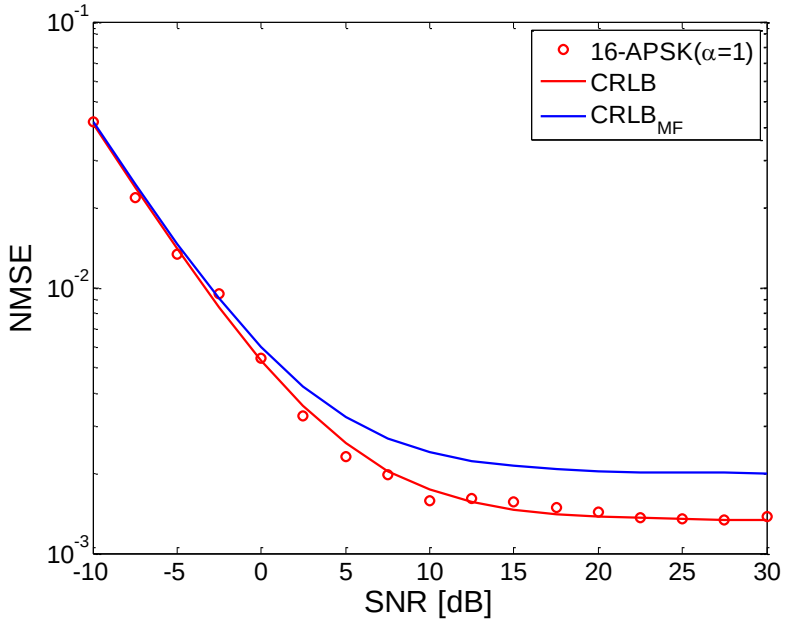


Figure 6-3: DA ML estimator applied to 16-APSK for $H_1(t)$ and $\alpha = 1$

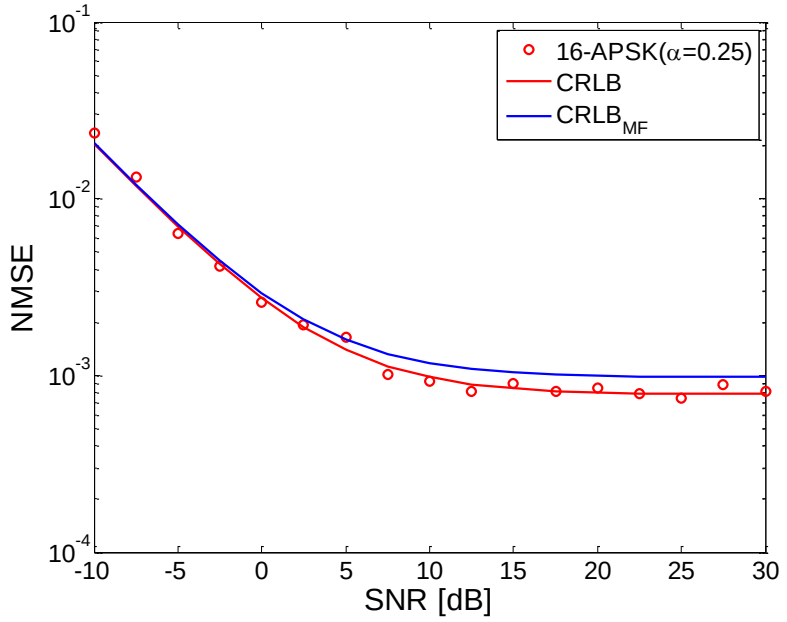


Figure 6-4: DA ML estimator applied to 16-APSK for $H_1(t)$ and $\alpha = 0.25$

6.1.2. NDA Case

Following the ML principle, authors in [1] [4] provided a derivation of the CRLB for several cases in the context of non-data-aided SNR estimation. However, in this thesis

we will focus only on one particular case, more precisely, on the quadratically constrained estimators that exploit only second-order statistics of data. Without going into details, NCRLB for this sort of SNR estimators is found as [1]:

$$\mathrm{NCRLB}(\rho) = \frac{\mathrm{CRLB}(\rho)}{\rho^2} = \frac{1}{K}\left(1+\frac{1}{\alpha}\right)\left(1+\frac{1}{\rho}\right)^2 + \frac{1}{K}(K_c - 2)\,, \tag{6.12}$$

with K_c denoting the symbol kurtosis, which has been introduced in the Chapter 4. It has to be mentioned that in case of Gaussian symbols, where $K_c = 2$, (6.12) provides the largest result, i.e. it represent the "worst case" for quadratic estimators, whereas constant modulus (CM) constellations with $K_c = 1$ yield to the lowest one. Hence, for non-CM constellations the CRLB is expected to lie in between [1].

The CRLB in (6.12) can be achieved by the so-called noise subspace estimator (NSE) presented in [4], which is an envelope-based scheme given as:

$$\hat{\rho} = \frac{\|\boldsymbol{x_0}\|^2}{\|\boldsymbol{x_1}\|^2} - 1\,, \tag{6.13}$$

where $\|\boldsymbol{x_0}\|^2 = \boldsymbol{x_0}^H\boldsymbol{x_0}$ and $\|\boldsymbol{x_1}\|^2 = \boldsymbol{x_1}^H\boldsymbol{x_1}$. It is shown in [4] that, if the PSD of the baudrate-sampled DMF output is constant within the roll-off band, then the NSE is optimal irrespective of the selected signal constellation.

Using 16-QAM signals for different roll-off factors, $K = 1024$ symbols and without shaping of the DMF, Figure 6-5 illustrates the evolution of the normalized mean square error (NMSE) for the noise subspace estimator (NSE) from (6.13). The normalized CRLB given by (6.12), w.r.t. the related modulation scheme, is depicted in green line, whereas the CRLB for Gaussian symbols and for the DA case, based only on the MF output, are depicted in magenta and blue color, respectively. As it can be seen, for $\alpha = 0.25$, up to SNR ≤ 10 dB, the NMSE is very close to the CRLB. However, for SNR > 10 dB, a significant degradation in performance is to be noticed. This can be explained by the fact that the impulse response of the FIR DMF is truncated to a window of N_s=21 symbol periods only. Therefore, due to slow roll-off behavior in the time domain, the DMF does not provide signal cancellation. For $\alpha > 0.25$, this phenomenon is also noticeable, but shifted to higher SNR regions. Nevertheless, the degradation in

performance can be mitigated by choosing a larger value of N_s, as it is observed from Figure 6-6 where N_s=81.

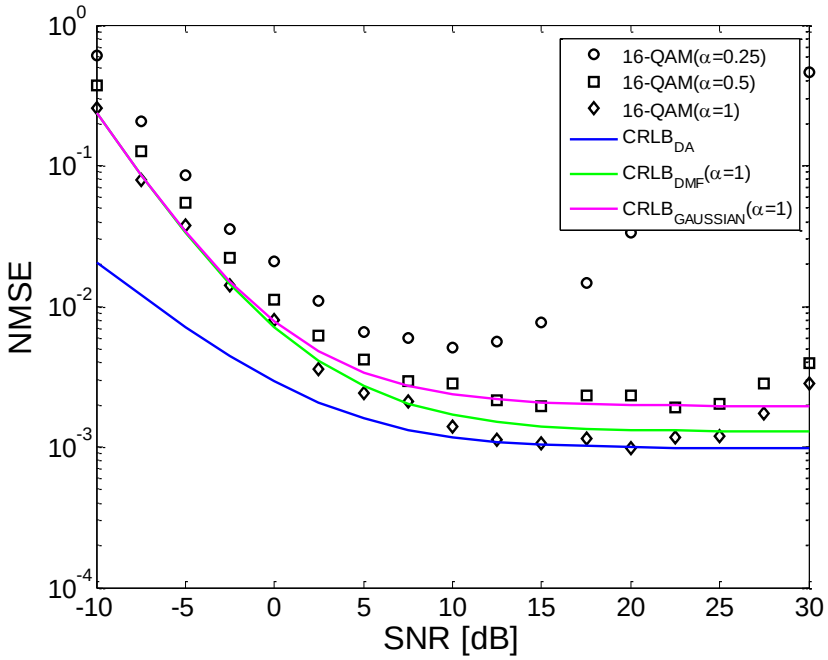


Figure 6-5: NSE applied to 16-QAM for· $N_S = 21$, *H₁(t)* and different α

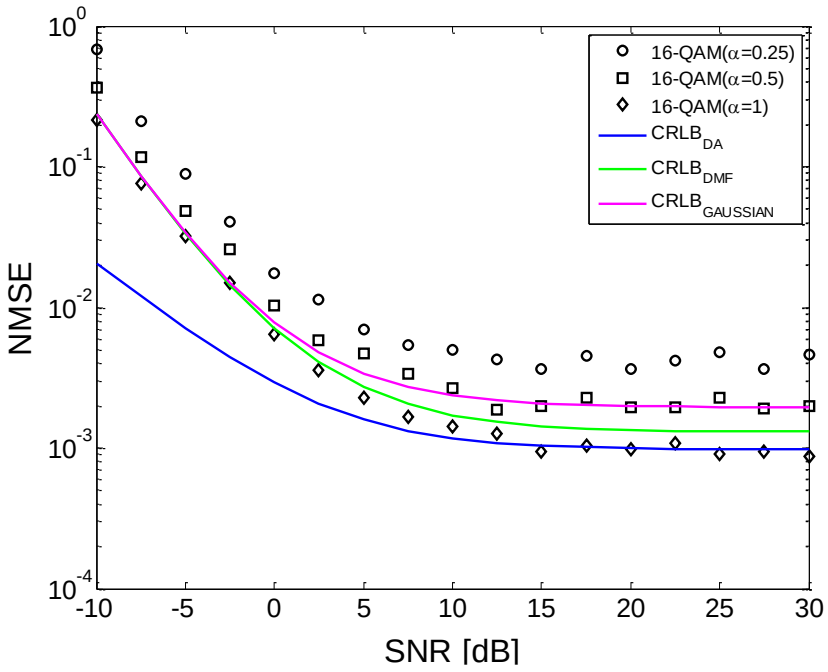


Figure 6-6: NSE applied to 16-QAM for· $N_S = 81$, *H₁(t)* and different α

For completeness, the mean estimator output with respect to the roll-off factor, is depicted in Figure 6-6, which confirms the results already discussed.

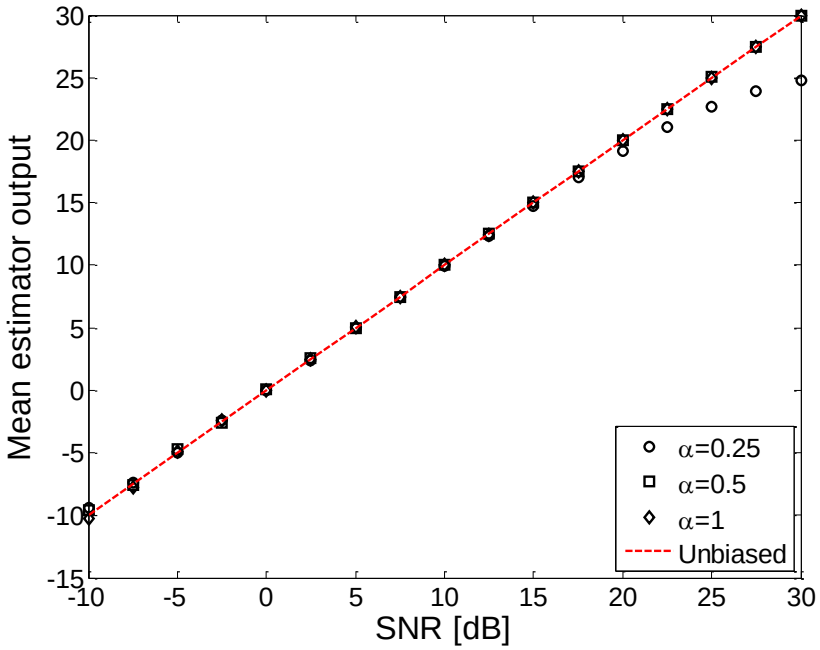


Figure 6-7: Mean output of NSE for different roll-off factors and *H₁(t)*

The performance results in Figures 6-5, 6-6 and 6-7 are achieved for a DMF realized without additional shaping. As already mentioned, performance loss in the higher SNR regime can be avoided by choosing N_S sufficiently large, but this unfortunately increases the computational load. However, when shaped by an additional filter function $P(f)$ satisfying the condition introduced in the Chapter 5, the DMF impulse response can be implemented with a significantly reduced number of taps [4].

Figure 6-8 depicts the NMSE of the NSE for $K = 1024$ and different values of α. The DMF is now smoothed by a raised cosine shape given by (5.17) and spanned over a period of $N_S = 21$ symbols. Obviously, compared to Figure 6-6, we notice a slight degradation in performance, which is the price to be paid for the reduced computational complexity. This kind of drawback is intuitively clear, since, as we already stated, optimality of the NSE is guaranteed only if the PSD of the DMF output is considered to be flat, which is not provided by shaping the DMF via a raised cosine shape.

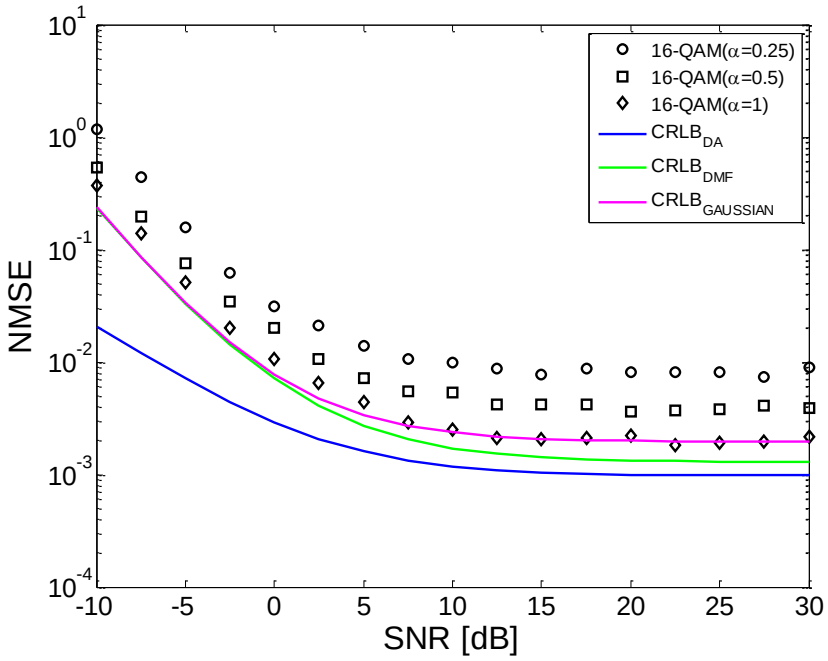


Figure 6-8: NSE applied to 16-QAM for $N_S = 21$, $\bar{H}_1(t)$ and different α

Simulation runs are repeated for 16-APSK constellations , with the same parameter settings used for 16-QAM. As in the DA case, basically the same results are achieved as for 16-QPSK, which was expected, since the NSE should be robust to different modulation schemes.

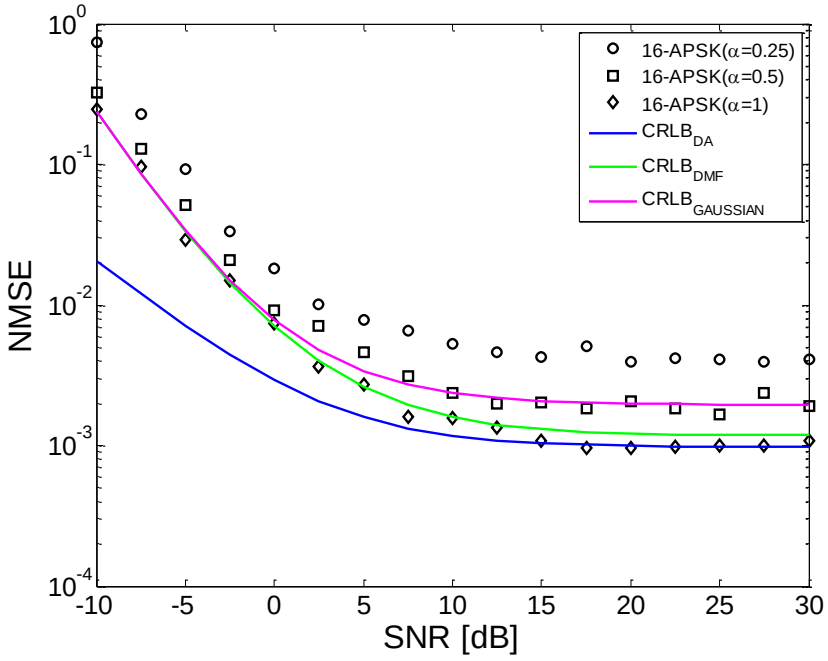


Figure 6-9: NSE applied to 16-APSK for $N_S = 81$, $H_1(t)$ and different α

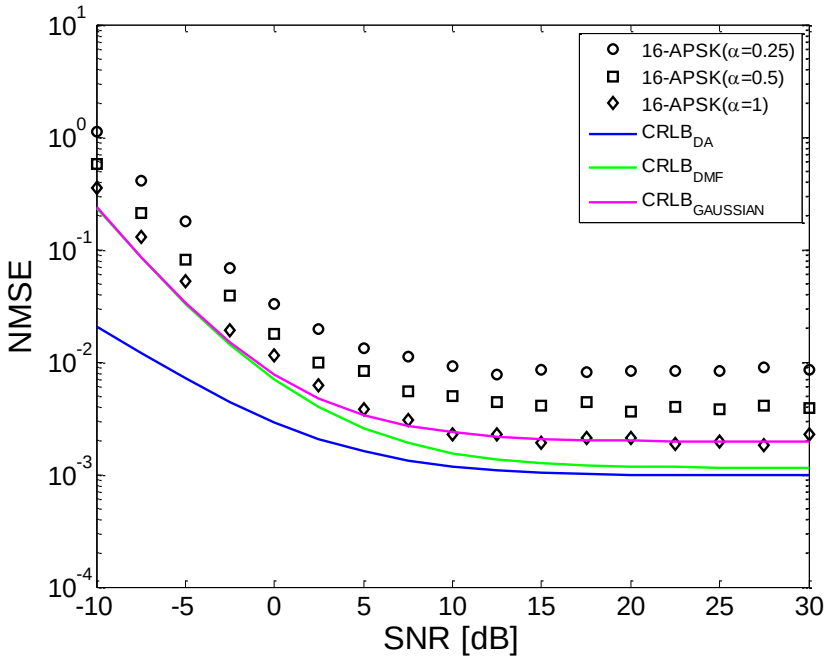


Figure 6-10: NSE applied to 16-APSK for $N_S = 21$, $\bar{H}_1(t)$ and different α

7. Conclusions and Outlook

Algorithms for moment-based estimation of the signal-to-noise ratio (SNR) are most welcome in the context of non-CM constellations, which are in the meantime specified for many next-generation communication systems. What makes this sort of estimators so attractive is the fact that they do not require any knowledge about data, they are robust to phase variations and, on top of that, they are very simple from the computational point of view.

When applied to non-CM signals, it could be shown that the modified version of the estimator based on second- and fourth-order moments is a powerful and attractive solution.

Moreover, by means of a dual-filter framework, it has been confirmed that the additional information within the roll-off band improves SNR estimation with respect to that case, where only baudrate samples at the matched filter (MF) output are exploited. The improvement depends on the excess bandwidth and can reach up to 3 dB for the DA case. In this respect, the noise subspace estimator (NSE) has been examined, which outperformed all moment-based solutions available from the open literature. Although the second filter in parallel to MF requires some extra complexity, it could be verified that the computational load reduces significantly, if this filter is shaped appropriately by an additional filter function.

Further investigations on the shaping function as well as joint estimation of SNR and carrier frequency/phase are subject to future work.