

AMERICAN UNIVERSITY OF BEIRUT

AN INFORMATION THEORETIC ANALYSIS
OF TIME-LIMITED SIGNALING

by
YOUSSEF ALI JAFFAL

A dissertation
submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
to the Department of Electrical and Computer Engineering
of Maroun Semaan Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
December 2020

AMERICAN UNIVERSITY OF BEIRUT


AN INFORMATION THEORETIC ANALYSIS OF TIME-LIMITED SIGNALING

by
YOUSSEF ALI JAFFAL

Approved by:

Dr. Ibrahim Abou Faycal, Professor
Department of Electrical and Computer Engineering

Advisor



Dr. Zaher Dawy, Professor
Department of Electrical and Computer Engineering

Chair of Committee




Dr. Mohammad Mansour, Professor
Department of Electrical and Computer Engineering

Member of Committee



Dr. Louay Bazzi, Professor
Department of Electrical and Computer Engineering

Member of Committee



Dr. Bassam Shayya, Professor
Department of Mathematics

Member of Committee



Dr. Gerhard Kramer, Professor
Technical University of Munich

Member of Committee



Dr. Tobias Koch, Visiting Professor
Universidad Carlos III de Madrid

Member of Committee



Date of thesis defense: December 3, 2020

Acknowledgements

I would like to deeply thank my supervisor prof. Ibrahim Abou-Faycal for his continuous support. I really appreciate his efforts that improved and sharpened my research and teaching skills. I am proud of completing this thesis which was not possible without the guidance and efforts of prof. Abou-Faycal.

It is my honor to have professors Gerhard Kramer and Tobias Koch as external committee members and I am thankful for their constructive feedback.

I would like to express my gratitude to the following professors at the department of mathematics at the American university of Beirut for welcoming me to attend their courses: professors Bassam Shayya, Farouk Abi Khuzam, Kamal Khuri-Makdisi, Florian Bertrand, and Giuseppe Della Sala. My special thanks are for prof. Bassam Shayya who is also one of my committee members.

Also I would like to thank the faculty members and the staff at the department of electrical and computer engineering at the American University of Beirut. My special thanks are for professors Zaher Dawy, Mohammad Mansour, and Louay Bazzi for being on my thesis committee. I am also grateful for taking with them several advanced courses on communication engineering and signal processing.

I would like to thank the IT team (especially Mr. Saadallah Itani and Mr. Ali Kaafarani) for their efforts and help in HPC configuration and software installation on the server.

I would like to acknowledge the American University of Beirut and the National Council for Scientific Research of Lebanon for granting me a doctoral fellowship.

Finally, I would like to thank my family. My deepest thanks are to my father and my mother for their love and support. I am also grateful for the support of my brothers Dr. Hussein and Dr. Hamza. I would like to thank my beloved wife Hawraa for supporting me during the last four years and for taking care of our lovely son Ali.

An Abstract of the Dissertation of

Youssef Ali Jaffal for Doctor of Philosophy
Major: Electrical and Computer Engineering

Title: An Information Theoretic Analysis of Time-Limited Signaling

Practical communication systems use finite duration transmissions and signaling. As the time-limited signals have infinite bandwidth, they lose some of their energy when passed through a band-limited channel. In 1948, Shannon derived the capacity of the band-limited Gaussian channel making use of band-limited and therefore infinite-duration pulses. In 1965, Wyner derived the channel capacity when using time-limited signals that are “approximately band-limited” in *an asymptotic regime*, as the time support of the signals tends to infinity. A major breakthrough in this theory was achieved by Polyanskiy who studied the channel coding rates in the finite block-length regime, however the derived results are only for discrete-time channels. A related theory is the rate distortion theory which was introduced by Shannon in 1948, where he derived the rate distortion function for band-limited white Gaussian sources. However his results are obtained by allowing compression at once an asymptotically infinite duration piece from the source, which is not possible for practical systems. Recently, Kostina studied the rate distortion theory in the finite block-length regime for discrete-time sources.

In this thesis I study the achievable rates when using signals with finite duration over a band-limited Gaussian channel, and the main goal is to quantify these rates when finite-duration codewords are employed over a band-limited channel. First I considered two system models where infinite duration codewords are transmitted over a band-limited Gaussian channel, however the pulses are restricted to be time-limited. Second, I considered transmitting finite duration codewords over a band-limited Gaussian channel and studied the achievable rates and the available degrees of freedom. Finally I extended this work to the “dual” problem of rate distortion theory and considered the coding of a time-limited piece of a band-limited Gaussian source.

Contents

Acknowledgements	v
Abstract	vi
1 Introduction and thesis objectives	1
1.1 Introduction	1
1.2 Thesis objectives and outcomes	3
2 Preliminaries and related works	6
2.1 Prolate spheroidal wave functions	6
2.2 The capacity of band-limited channels	10
3 Combined PAM-OMM with finite duration pulses	15
3.1 System model	15
3.2 Problem formulation	19
3.2.1 Asymptotic analysis as $N \rightarrow \infty$	21
3.3 Numerical results	22
3.3.1 Degrees of freedom	23
3.3.2 Achievable rates	23
3.4 Sampling the received signal $y(t)$	25
4 PAM with finite duration pulses	29
4.1 System model and problem formulation	30
4.1.1 Optimization	32
4.2 Band-limited filters	33
4.2.1 The case where $2WT > 1$; Nyquist pulses	33
4.2.2 The case where $2WT = 1$	34
4.2.3 The case where $2WT < 1$	34
4.3 Time-limited filters	35
4.3.1 The case where $2WT > 1$; Nyquist pulses	35
4.3.2 The case where $2WT = 1$	37
4.3.3 The case where $2WT < 1$	39

5	Finite duration codewords over band-limited Gaussian channel	42
5.1	System Model and Problem Formulation	42
5.2	Bounds on the Data Rates	44
5.2.1	An Upper Bound	44
5.2.2	Lower Bounds	47
5.3	Numerical Results	52
5.3.1	Upper Bound	53
5.3.2	Lower Bounds	55
5.4	Possible Enhancements on the Bounds	56
5.4.1	A Tighter Upper Bound	56
5.4.2	Introducing a Guard Time or a Guard Band between the Codewords	57
5.5	Conclusion	58
6	Lossy coding of a time-limited piece of a band-limited white Gaussian source	59
6.1	System model	59
6.2	Bounds	61
6.2.1	Converse – Lower Bound	61
6.2.2	Achievability – Upper Bound	63
6.3	Asymptotic analysis as $T \rightarrow \infty$	64
6.4	Numerical results	65
6.5	Remarks	65
	Summary and conclusions	68
A	Properties of $k, h \alpha_{c, l, m}$	70
B	Abbreviations	75

List of Figures

2.1	Eigenvalues of PSWFs for $c = 100$ and $84 \leq l \leq 116$	9
3.1	Combined PAM-OMM System Model.	16
3.2	Histograms of the eigenvalues of $\mathbf{S}(e^{j2\pi f})$ for different c and N . . .	23
3.3	Convergence of achievable rates for $c=1$	24
3.4	Percentage decrease with Shannon's formula for different c	25
4.1	PAM System Model	31
4.2	Performance of time-limited RRC and band-limited RRC	36
4.3	Performance of the obtained filters for $2WT=1$ and $c=3$	38
4.4	Time-limited RRC vs time-limited filters with $2WT=1$	39
4.5	Performance of the obtained filters for $2WT=0.8$ and $c=5$	41
5.1	Continuous time system model.	43
5.2	Equivalent discrete time system model.	45
5.3	Single band interference.	47
5.4	Single time-slot interference.	50
5.5	Saturation of R_{UB}^N for $c = 1000$ and $\text{SNR} = 50$ dB.	54
5.6	Upper bounds on the degrees of freedom.	54
5.7	Upper bounds on the data rates.	54
5.8	Lower bounds on the degrees of freedom.	55
5.9	Upper and lower bounds on the rates.	56
5.10	Tighter upper bound on the rates vs. the old ones.	57
5.11	Upper bound on $\frac{T_G^*}{T}$ for different SNR levels.	58
6.1	$X(t)$ is encoded by independently coding $\{X_k(t)\}_{k \in \mathbb{Z}}$	60
6.2	Bounds vs. c for $\sigma = 1$, $W = 1000Hz$, $d = \frac{1}{4}2W$ and $\epsilon = 10^{-2}$ in the first figure and $\epsilon = 10^{-4}$ in the second.	66
6.3	Bounds vs. c for $\sigma = 1$, $W = 1000Hz$, $d = \frac{1}{1000}2W$, $\epsilon = 10^{-2}$	66

List of Tables

2.1	Numerical approximations for $\sum_{i=0}^{\infty} \lambda_{c,i}$, $\sum_{i=0}^{\infty} \lambda_{c,i}^2$, and $\sum_{i=0}^{\infty} \lambda_{c,i}^3$ using the first N summands	10
3.1	Maximum percentage decrease with Shannon's formula	25

Chapter 1

Introduction and thesis objectives

1.1 Introduction

Wireless communication systems that use radio waves have tight requirements on the used spectrum, where every technology should abide by a specific transmit spectral mask, a mask that allows the communication system to transmit data within a specific radio band, and guaranties an acceptable level of interference on the other technologies. As the adjacent radio bands may be used by other technologies, it is reasonable to consider that in wireless communication technologies the transmitter confines its transmitted data within its radio band, and the receiver looks for the transmitted data in this band.

A major factor that affects receiving the transmitted data is the added noise in the channel. It is intuitive that the received signal power should be high enough (with respect to the noise power) so the receiver can distinguish the transmitted signals. The added noise in the channel is often assumed to be an Additive White Gaussian Noise (AWGN) process. One reason that supports this assumption is that Gaussian noise has the “worst” distribution that maximizes the entropy for a given noise variance [1]. Additionally, the central limit theorem supports this assumption, since the thermal noise results from the additive effect of independent random motions of electrons.

Information theory determines the bounds that communication systems cannot surpass. The Legend of Albert [2] addresses the importance of information theory for communication engineers. The pioneering works of Claude Shannon [1, 3] defined the capacity of real band-limited Gaussian channel to be

$$C = W \times \log_2 \left(1 + \frac{P}{N_0 W} \right) \text{ bits per second,}$$

where W is the bandwidth of the channel, P is the transmitted signal power, and N_0 is the noise power per cycle. Any data rate below the channel capacity can

be achieved with arbitrary small probability of error, and no data rate is possible beyond this channel capacity. Shannon derived first the channel capacity for the Discrete Time (DT) channel as the block length of the codewords grows towards infinity. He then used the sampling theorem which provides a one-to-one relation between the continuous time and discrete time signals: For W -Hz band-limited signals when sampled at a rate of $2W$ samples per second, the operation is invertible and information lossless making use of the “sinc” function defined in this thesis as

$$\text{sinc}(2Wt) \triangleq \frac{\sin(2\pi Wt)}{2\pi Wt}.$$

Since strictly band-limited signals must be eternal, they cannot be used in any practical communication system. Moreover, some applications and technologies require real time communication with unnoticeable delays forcing the use of relatively short time signals. Some examples of such applications are phone calls, video conferencing, and online gaming. Additionally, the latency requirements are tighter in new generations of cellular communication systems compared to previous generations. For example, some of the goals set by the METIS project [4] to achieve in 5G systems compared to 4G systems are: to reduce the end to end latency by a factor of 5, to increase the battery life by a factor of 10, and to increase the user data rate by a factor between 10 to 100. However, we do not know a practical theoretical relation between the channel capacity, the transmission power, the transmission duration and the used radio band.

While the derived formula by Shannon is mathematically rigorous, some of the made assumptions do not hold in practical settings.

First: the conversion from DT to Continuous Time (CT) and vice versa is not practical since the “sinc” function needs an infinite time support. Moreover, the use of any band-limited function with non-zero finite energy is not possible in practice. In the literature, Wyner [5,6] and Gallager [7] tackled this issue and considered the use of T -seconds time-limited codewords. However, they derived asymptotic results as $T \rightarrow \infty$ and reached the same formula derived by Shannon.

Second: the second questionable assumption in practical settings is the use of infinite block-length codewords which is not feasible. Dobrushin [8], Strassen [9], and Polyanskiy [10,11] derived asymptotic expansions for the achievable rates at a given probability of error. However the results were derived only for DT channels. One of the results by Polyanskiy is the maximum achievable rates for L -parallel real Gaussian channels [11, Theorem 4].

Another related “dual” theory is the rate distortion theory which was first introduced by Shannon in his pioneer work [1]. One of his results is the rate distortion function for a real W -Hz band-limited white Gaussian source:

$$R_{\text{Shannon}} = W \log_2 \left(\frac{Q}{D} \right) \text{ bits per seconds}, \quad (1.1)$$

where Q is the average power of the source and D is the average distortion level measured using the mean squared error. Obtaining this expression may be achieved by using the Shannon-Nyquist sampling theorem to transform the problem to a DT one, and then applying the rate distortion function for the DT stationary Gaussian source [12, chapter 4.6.3]. Equation (1.1) is obtained by allowing compression at once an asymptotically infinite duration piece from the source, and hence the block-length in the equivalent DT model is infinite. The authors in [13] and [14] investigated the rate distortion theory in the finite block-length for stationary DT Gaussian sources regime, where they provide a closed form expression for the approximation of the minimal rate that achieves a maximum distortion level d with probability $1 - \epsilon$:

$$R(n, d, \epsilon) \approx \frac{n}{2} \log_2 \frac{\sigma^2}{d} + \sqrt{\frac{n}{2}} Q^{-1}(\epsilon) \log_2 e,$$

where n is the block-length and σ^2 is the variance of the source. Both papers showed that this approximation lies between tight upper and lower bounds.

1.2 Thesis objectives and outcomes

Using finite duration signals over a band-limited channel disagrees with some assumptions that Shannon used to derive his result: it leads to energy losses, introduces inter-codeword interference and possibly Inter-Symbol Interference (ISI) within the same codeword, and raises a fundamental question about the degrees of freedom when using simultaneously a finite time window and a finite radio band.

Since time-limited signals have infinite radio band support, when passed through a quasi band-limited filter they lose some of the energy components outside the radio band and therefore the Signal to Noise Ratio (SNR) decreases. Moreover, the resulting quasi band-limited signals will have larger time support and therefore their time support will overlap with the time support of the neighbouring signals. In this study I assume that the channel is an ideal low pass filter approximating a well-designed Radio Frequency (RF) circuitry at the front-end of the receiver as done in practical systems. Indeed, RF engineers design low pass filters to best approximate the ideal low pass filter and the approximation can be made sharper and sharper by increasing the order of a Butterworth filter for example. On a final note, Gallager [7, chapter 8] studied the channel capacity when transmitting time-limited signals over an arbitrary channel $H_1(f)$ such as the ideal low pass filter. In his work $H_1(f)$ may not be realizable and plays the role of a mathematical constraint. And in this thesis I adopt ideal low pass filters in my models to force the transmitter not to send information outside the radio band. Therefore the considered channel has infinite memory, and the issue of inter-codeword interference will be unavoidable. Introducing a

guard time between the consecutive transmissions suppresses the effect of inter-codeword interference, but at the cost of reduced usage of the time resource. The argument that Gallager [7, section 8.5] used for infinite block-length codewords cannot be applied in the finite block-length case; he stated that by introducing a very large guard time between consecutive codewords while using infinitely much larger time support for the codewords, the effect of the lost time resource and the effect of inter-codeword interference can be suppressed as much as needed. Such an argument cannot be used in my setting since the time duration of a codeword is finite and maintained fixed: A trade-off between the guard time length and the inter-codeword interference must be maintained in the finite duration transmission systems.

The main goal of this thesis is to study the information rates and the channel capacity when using T -seconds time-limited transmissions over a W -Hz band-limited Gaussian channel. A related question is that of determining the degrees of freedom in such scenario. The space of finite-energy functions that are band-limited and time-limited contains only the zero function. It is nevertheless commonly accepted in the literature that the dimension of the W-T space is approximately $2WT$. This argument is supported by results in [3, 15–18] that were also derived in the asymptotic regime as $WT \rightarrow \infty$.

The main tasks I accomplished in this thesis can be summarized as follows:

- To evaluate the achievable rates when using time-limited pulses but with arbitrarily large codes and possibly infinite transmissions. This task has been accomplished and it is presented in chapters 3 and 4.
 - i) In chapter 3 I first derived the achievable rates when sending T -seconds time-limited pulses using combined Pulse Amplitude Modulation - Orthogonal Multi-pulse Modulation (PAM-OMM) systems over a W -Hz band-limited Gaussian channel, and then performed numerical computations to evaluate them. Interestingly, the results show that these channel capacities depend on the time-frequency factor $c = 2WT$. When $c \geq 1$ the achievable rates are almost equal to the Shannon capacity, and hence the loss of energy has no significant effect. Using my system model I also established that there are at most $2WT$ independent data symbols. The results of this chapter were published in the following conference paper:

Y. Jaffal, and I. Abou-Faycal, “*Using time-limited pulses in a combined PAM-OMM system over band-limited channels*”, 2019 IEEE Wireless Communications and Networking Conference (WCNC), 2019, 1-7.
 - ii) In chapter 4 I considered a classical PAM system and studied optimal signaling when using time-limited pulses over band-limited AWGN channels. I adopted an information theoretic approach and quantified

the achievable rates of such systems. I showed that the Nyquist criterion cannot be satisfied and that signaling at faster than the Nyquist rate (where $T < \frac{1}{2W}$) is significantly closer to optimality. The derivations and results of this chapter were published in the following conference paper:

Y. Jaffal, I. Abou-Faycal, “Achievable rates using PAM time-limited pulses over band-limited channels: From Nyquist to FTN”, 2019 IEEE Wireless Communications and Networking Conference (WCNC), 2019, 1-6.

- To study the achievable rates when using finite duration codewords. This task is accomplished and it is presented in chapter 5, where I considered a communication system whereby T -seconds time-limited codewords are transmitted over a W -Hz band-limited AWGN channel. I derived upper and lower bounds for the achievable rates and the corresponding degrees of freedom and I numerically evaluated them for sample values of $2WT$. The bounds are asymptotically tight and numerical computations show the gap between them decreases as $2WT$ increases. Additionally, the possible degradation in the available degrees of freedom is upper-bounded by a logarithmic function of $2WT$. The results of chapter 5 were published in the following journal paper:

Y. Jaffal, and I. Abou-Faycal. Time-Limited Codewords over Band-Limited Channels: Data Rates and the Dimension of the WT Space. *Entropy*, 2020, 22(9).

- To extend this work to the “dual” problem of rate distortion theory, where sampling a time-limited signal is necessarily “energy-lossy”. This task is accomplished and the results are presented in chapter 6; I considered the problem of source coding a T -seconds finite duration piece of a W -Hz band-limited white Gaussian process with the \mathcal{L}^2 norm of the error as a distortion measure. I derived a lower bound and an upper bound for the smallest rate that guarantees a distortion level with a probability $1 - \epsilon$ and I numerically evaluated these bounds. I also showed that the derived bounds are asymptotically tight where they converge to Shannon’s formula. The results of chapter 6 were published in the following conference paper:

Y. Jaffal, and I. Abou-Faycal. “Lossy coding of a time-limited piece of a band-limited white Gaussian source.” IEEE International Symposium on Information Theory (ISIT 2020), 2020, 2327-2331.

Chapter 2

Preliminaries and related works

2.1 Prolate spheroidal wave functions

In [19] Slepian and Pollak showed that the Prolate Spheroidal Wave Functions (PSWFs) possess properties that make them useful in the Fourier analysis of band-limited functions and time-limited functions. Their work was motivated by solving the problem of maximum energy concentration. More specifically, among all W -Hz band-limited functions which function $f(t)$ has the maximum energy concentration over the time window $[-\frac{T}{2}, \frac{T}{2}]$:

$$\frac{\int_{-\frac{T}{2}}^{\frac{T}{2}} |f(t)|^2 dt}{\int_{-\infty}^{\infty} |f(t)|^2 dt}.$$

The authors proved that $f(t)$ is the eigenfunction with the highest eigenvalue of the integral equation

$$\lambda f(t) = \int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{\sin 2\pi W(t-s)}{\pi(t-s)} f(s) ds, \quad (2.1)$$

which is a Fredholm integral equation of the second kind [20]- [21]. While the literature provides numerical methods to solve some Fredholm integral equations, Slepian and Pollak introduced a brilliant and unique way to solve this integral equation [22]- [23]; they found the differential operator that commutes with the integral operator in (2.1), so both operators have the same eigenfunctions. Therefore the eigenfunctions of the integral equation in (2.1) are some scaled versions of the solutions of the differential equation

$$(1-t^2) \frac{d^2 u}{dt^2} - 2t \frac{du}{dt} + (\chi - c^2 t^2) u = 0,$$

where $c = \pi WT$ and the solutions of this differential equation are the angular and radial PSWFs.

For any $c = \pi WT > 0$, they defined the PSWFs as an infinitely countable set of real functions $\{\varphi_{c,l}(t)\}_{l \in \mathbb{N}}$, normalized solutions of the integral equation where for every $l \in \mathbb{N}$,

$$\lambda_{c,l} \varphi_{c,l}(t) = \int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{\sin 2\pi W(t-s)}{\pi(t-s)} \varphi_{c,l}(s) ds, \quad t \in \mathbb{R}.$$

Property 2.1.1 The PSWFs form a Complete OrthoNormal (CON) set for band-limited functions [19] with

$$\int_{-\infty}^{\infty} \varphi_{c,l}(t) \varphi_{c,m}(t) dt = \delta_{lm}, \quad \forall l, m \in \mathbb{N} \times \mathbb{N},$$

where δ_{lm} is the Kronecker delta.

Property 2.1.2 The PSWFs are orthogonal over the time window T :

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \varphi_{c,l}(t) \varphi_{c,m}(t) dt = \lambda_{c,l} \delta_{lm}, \quad \forall l, m \in \mathbb{N} \times \mathbb{N},$$

where $\{\lambda_{c,l}\}_{l \in \mathbb{N}}$ are the eigenvalues that are all in the range $0 < \lambda_{c,l} < 1$ and decreasing in l [19].

The eigenvalue $\lambda_{c,l}$ may be hence viewed as the energy concentration of $\varphi_{c,l}(t)$ in the time interval $[-\frac{T}{2}, \frac{T}{2}]$ and $\varphi_{c,0}(t)$ has the highest energy concentration. Additionally, the PSWFs are real continuous functions that are even when l is even and odd when l is odd.

In his book [7, section 8.4], Gallager considered a simple system consisting of a time-limited signal $x(t)$ that produces the output signal $y(t)$ when passed through a time-invariant filter $h(t)$. As it is desirable to have a CON basis for $x(t)$ and a CON basis for $y(t)$, Gallager determined that the members of the CON basis of $x(t)$ should satisfy the following integral equation

$$\lambda \varphi(\tau_1) = \int_{-\frac{T}{2}}^{\frac{T}{2}} \mathcal{R}(\tau_1, \tau_2) \varphi(\tau_2) d\tau_2, \quad (2.2)$$

where

$$\mathcal{R}(\tau_1, \tau_2) = \int_{-\infty}^{\infty} h(t - \tau_1) h(t - \tau_2) dt.$$

Whenever the filter $h(t)$ is the ideal W -Hz low pass filter,

$$h(t) = \frac{\sin 2\pi W t}{\pi t} = 2W \text{sinc}(2Wt),$$

and by the properties of the “sinc” function,

$$\mathcal{R}(\tau_1, \tau_2) = \frac{\sin 2\pi W(\tau_2 - \tau_1)}{\pi(\tau_2 - \tau_1)}. \quad (2.3)$$

Note that Gallager made use of one important property of the PSWFs, namely the Fourier transform of a PSWF is a scaled version of time-limited PSWF

$$\begin{aligned} \Phi_{c,l}(f) &= \begin{cases} j^l \sqrt{\frac{T}{2W\lambda_{c,l}}} \varphi_{c,l}\left(\frac{T}{2W}f\right) & \text{for } f \in [-W, W] \\ 0 & \text{otherwise} \end{cases} \\ &= j^l \sqrt{\frac{T}{2W\lambda_{c,l}}} \varphi_{c,l}\left(\frac{T}{2W}f\right) \text{rect}\left(\frac{f}{2W}\right) \end{aligned} \quad (2.4)$$

where $j = \sqrt{-1}$. These transforms solve themselves the integral equation

$$\lambda_{c,i} \Phi_{c,i}(f) = \int_{-W}^W \frac{\sin \pi T(f-s)}{\pi(f-s)} \Phi_{c,i}(s) ds, \quad \forall i \in \mathbb{N}. \quad (2.5)$$

for $-W \leq f \leq W$. By Mercer’s Theorem [24]

$$\frac{\sin \pi T(f-s)}{\pi(f-s)} = \sum_{i=0}^{\infty} \lambda_{c,i} \Phi_{c,i}(f) \Phi_{c,i}^*(s) \quad (2.6)$$

$$\frac{\sin \pi(f-s)}{\pi(f-s)} = \sum_{i=0}^{\infty} \frac{\lambda_{c,i}}{T} \Phi_{c,i}(f/T) \Phi_{c,i}^*(s/T) \quad (2.7)$$

for $-W \leq f, s \leq W$. Note that it can be shown that $\sqrt{\lambda_{c,n}} \Phi_{c,n}(f)$ is bounded, i.e. $\sqrt{\lambda_{c,n}} |\Phi_{c,n}(f)| \leq \sqrt{T}$ (proof by Cauchy Schwartz or using equation (2.6) with $s = f$).

In this thesis I use $c = 2WT$ as index for the PSWFs, which is different from the one used by Slepian and Pollak [19]: more specifically $\varphi_{1,i}(t)$ here is the same as $\varphi_{\frac{T}{2},i}(t)$ in [19]. Define the $[-\frac{T}{2}, \frac{T}{2}]$ time-limited version of a PSWF by

$$D\varphi_{c,l}(t) \hat{=} \varphi_{c,l}(t) \text{rect}\left(\frac{t}{T}\right), \quad (2.8)$$

with \mathcal{L}^2 -norm or energy $\lambda_{c,i}$.

Property 2.1.3 The normalized time-limited PSWFs $\left\{ \frac{D\varphi_{c,i}(t)}{\sqrt{\lambda_{c,i}}} \right\}_{i \in \mathbb{N}}$ form a CON set for T -seconds time-limited functions.

Denote by $\Phi_{c,l}^D(f)$ the Fourier transform of $\frac{D\varphi_{c,l}(t)}{\sqrt{\lambda_{c,l}}}$ which is equal to

$$\Phi_{c,l}^D(f) = j^l \sqrt{\frac{T}{2W}} \varphi_{c,l} \left(\frac{T}{2W} f \right), \quad f \in \mathbb{R}. \quad (2.9)$$

Based on (2.4), (2.8) and (2.9), when $\frac{D\varphi_{c,l}(t)}{\sqrt{\lambda_{c,l}}}$ is passed through an ideal low pass filter with transfer function $\text{rect}(\frac{f}{2W})$, the output is $\sqrt{\lambda_{c,l}} \varphi_{c,l}(t)$:

$$\frac{D\varphi_{c,l}(t)}{\sqrt{\lambda_{c,l}}} * 2W \text{sinc}(2Wt) = \sqrt{\lambda_{c,l}} \varphi_{c,l}(t). \quad (2.10)$$

I have used a debugged version of the software package by Adelman et al. [25] to compute the eigenvalues of the PSWFs. While it is known [7, section 8.4] that for any $\gamma > 0$

$$\lim_{c \rightarrow \infty} \lambda_{c,c(1+\gamma)} = 0, \quad (2.11)$$

$$\lim_{c \rightarrow \infty} \lambda_{c,c(1-\gamma)} = 1. \quad (2.12)$$

Fig. 2.1 shows the eigenvalues $\lambda_{c,l}$ of the PSWFs for $c = 100$ and $84 \leq l \leq 116$. Note that

$$\lambda_{100,l} < 10^{-5} \quad \text{for } l > 107, \quad \& \quad \lambda_{100,l} > 1 - 10^{-5} \quad \text{for } l < 92.$$

This transition region between the ‘‘extreme’’ eigenvalues (very close to 1 or 0) is known to have a length proportional to the logarithm of c [7, section 8.4]. For example, for $c = 2000$, $\lambda_{2000,l}$ is between 10^{-5} and $1 - 10^{-5}$ only in the range $1988 < l < 2011$.

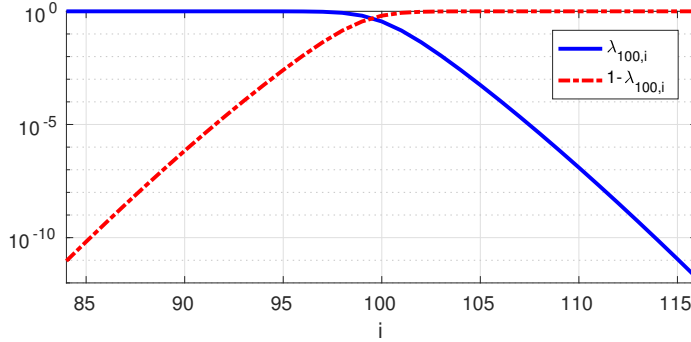


Figure 2.1: Eigenvalues of PSWFs for $c = 100$ and $84 \leq l \leq 116$.

Additionally, by Mercer’s Theorem [24],

$$\frac{\sin 2\pi W(t-s)}{\pi(t-s)} = \sum_{i=0}^{\infty} \varphi_{c,i}(t) \varphi_{c,i}(s) \quad \text{for } -\frac{T}{2} \leq t, s \leq \frac{T}{2},$$

and for $t = s$, integrating over $[-\frac{T}{2}, \frac{T}{2}]$ yields

$$\sum_{i=0}^{\infty} \lambda_{c,i} = 2WT = c. \quad (2.13)$$

In table 2.1 I list numerical approximations for sample values of $\sum_{i=0}^{\infty} \lambda_{c,i}$, $\sum_{i=0}^{\infty} \lambda_{c,i}^2$, and $\sum_{i=0}^{\infty} \lambda_{c,i}^3$ by adding only the first N terms.

	$c = 100, N = 115$	$c = 1000, N = 1015$
$\sum_{i=0}^{N-1} \lambda_{c,i}$	99.999999999988	999.999999967
$\sum_{i=0}^{N-1} \lambda_{c,i}^2$	99.187377724737	998.954076824
$\sum_{i=0}^{N-1} \lambda_{c,i}^3$	98.781143572125	998.431122935

Table 2.1: Numerical approximations for $\sum_{i=0}^{\infty} \lambda_{c,i}$, $\sum_{i=0}^{\infty} \lambda_{c,i}^2$, and $\sum_{i=0}^{\infty} \lambda_{c,i}^3$ using the first N summands

In this thesis, shifted (in time and frequency) PSWFs will come in handy. Denote by ${}_{k,h}\alpha_{c,l,m}$ the inner product between the l^{th} normalized time-limited PSWF already shifted in time and frequency by kT -seconds and $h2W$ -Hz respectively, and the band-limited version of the m^{th} normalized time-limited PSWF. By equation (2.10) and using Parseval,

$$\begin{aligned} {}_{k,h}\alpha_{c,l,m} &\hat{=} \left\langle \frac{D\varphi_{c,l}(t-kT)}{\sqrt{\lambda_{c,l}}} e^{j2\pi h2W(t-kT)}, \sqrt{\lambda_{c,m}} \varphi_{c,m}(t) \right\rangle \\ &= \left\langle \Phi_{c,l}^D(f-h2W) e^{-j2\pi kTf}, \Phi_{c,m}^D(f) \text{rect}\left(\frac{f}{2W}\right) \right\rangle \\ &= \left\langle \Phi_{c,l}^D(f-h2W) e^{-j2\pi kTf} \text{rect}\left(\frac{f}{2W}\right), \Phi_{c,m}^D(f) \right\rangle, \end{aligned} \quad (2.14)$$

which can be interpreted as the inner product between the bandlimited shifted normalized l^{th} PSWF and the m^{th} normalized time-limited PSWF. In Appendix A, I study the magnitude of ${}_{k,h}\alpha_{c,l,m}$ and derive various bounds that are used in chapter 5.

2.2 The capacity of band-limited channels

In this section, I summarize the different derivations of the channel capacity of a band-limited Gaussian channel, that are proposed by Shannon [1,3], Wyner [5,6],

and Gallager [7]. In the following the power budget is restricted to P Watts, and hence the total energy per T -seconds time window is PT . The noise is assumed to be AWGN with Power Spectral Density (PSD) $\frac{N_0}{2}$ Watts/Hz. The capacity of W -Hz band-limited channel in the presence of AWGN was first derived by Claude Shannon [1, 3] to be

$$C = W \times \log_2 \left(1 + \frac{P}{N_0 W} \right). \quad (2.15)$$

Given an absolute continuous random variable X with Probability Density Function (PDF) $p(x)$, Shannon [1] defined the differential entropy $h(X)$ as a measure of uncertainty of X , such that

$$h(X) = - \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx.$$

Shannon proved that the Gaussian distribution maximizes $h(X)$ under the constraint that the variance is fixed, where the Gaussian distribution is given by

$$p^*(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

where σ is the standard deviation. And the corresponding entropy is

$$h^*(X) = \frac{1}{2} \log_2 2\pi e\sigma^2.$$

Also, given another absolute continuous random variable Y with PDF $p(y)$, he defined the conditional entropy $h(Y/X)$ as a measure of the average uncertainty in Y when X is known, where

$$h(Y/X) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log_2 \frac{p(x, y)}{p(x)} dy dx.$$

Under a power constraint on the codewords, Shannon [1] proved that the channel capacity is given by

$$C = \max_{p(x)} (h(Y) - h(Y/X)) = \max I(X, Y), \quad (2.16)$$

where $p(x)$ is such that $E[X^2] \leq P$ and it is optimal when it is Gaussian distributed. The induced $p(y)$ is a zero mean Gaussian distribution and its variance is the sum of the variances of X and the noise.

To study the CT channels, Shannon used the sampling theorem which provides a one-to-one relation between the CT and DT for W -Hz band-limited signals when sampled no slower than $2W$ samples per second. Therefore, in any time window T there are $2WT$ independent data symbols and hence the energy per sample is $\frac{P}{2W}$. The noise is assumed to be W -Hz band-limited (if not the received signal is

passed through a low pass filter). The noise power is N_0W and the noise variance per sample is $N_0W \frac{T}{2WT} = \frac{N_0}{2}$.

Per second,

$$C = 2W \times \frac{1}{2} \log \left(1 + \frac{P}{\frac{N_0}{2}} \right),$$

and the channel capacity follows as in (2.15).

In [3], Shannon used a geometrical method to find the capacity of the band-limited Gaussian channel. Using the sampling theorem he showed that the dimension of the W - T space is no more than $2WT$, then he considered the signals in this space that have an average power less than P and mapped them to the geometrical space; Shannon showed that these signals correspond to the points that fall within a sphere of radius $r = \sqrt{2WTP}$ in the $2WT$ dimensional geometric space. Let the point q_1 to be the point that corresponds to the signal x_1 , then if the signal x_1 is transmitted through a noisy channel with a noise power N_0W , the received signal y_1 will correspond to a point in the geometrical space that falls within a sphere of radius $\sqrt{2WTN_0W}$ and centred at q_1 , and therefore any received signal will be mapped to a point within a sphere of radius $\sqrt{2WT(N_0W + P)}$. To be able to extract the sent message with very small probability of error, the received signals should map to different spheres that do not intersect. Therefore the maximum number of used messages can be calculated through dividing the volume of the sphere with radius $\sqrt{2WT(N_0W + P)}$ by the volume of the sphere with radius $\sqrt{2WTN_0W}$. The volume of a sphere of radius r in the $2WT$ dimensional space is given by

$$V = \alpha r^{2WT},$$

where α is fixed when $2WT$ is fixed. Therefore the maximum number of possible messages is

$$M = \left(\sqrt{1 + \frac{P}{N_0W}} \right)^{2WT}.$$

And hence the channel capacity is

$$C = \frac{\log_2 M}{T} = W \times \log_2 \left(1 + \frac{P}{N_0W} \right) \quad \text{bits/seconds.}$$

Note that the capacity of the complex channel is

$$C_{Shannon} = 2W \log_2 \left[1 + \frac{P}{2N_0W} \right] \quad \text{bits per seconds,} \quad (2.17)$$

where W is the bandwidth of the baseband channel, P is the average transmit power and N_0 is the spectrum of the additive circularly symmetric complex white Gaussian noise.

Although the proof is mathematically rigorous, the method implies the use of “sinc” pulses which is not practical. In the literature, Wyner [5,6] and Gallager [7] tackled this issue and considered the use of T -seconds time-limited codewords. However, they derived asymptotic results as $T \rightarrow \infty$ and reached the same formula (2.15) derived by Shannon.

In [5], Wyner considered four different physical models and derived the asymptotic channel capacity for each model. Wyner stated that the first two models suffer from some physical difficulties; the assumed noise in the first model results in an infinite noise power at the receiver, and the use of strictly band-limited signals in the second model is not practical and may produce interference between consecutive codewords. On the other hand, he proved that using a noise model with finite power results in infinite capacity. Wyner made some assumptions to avoid these issues in [6] and in the third and fourth models in [5]. He derived the channel capacity of the different models by relating the continuous time to discrete time models as Shannon did in [1], but by using the PSWFs and their property that as $2WT \rightarrow \infty$, the first $2WT$ PSWFs form asymptotically a CON set for the time-limited and approximately band-limited signals.

Gallager [7, section 8.5] considered transmitting time-limited signals over a channel with impulse response $h(t)$. He found that the desirable orthonormal basis functions for the input signal should satisfy (2.2). Using the input signal $X(t)$

$$X(t) = \sum_{i=0}^N X_i \varphi_i(t),$$

produces the signal $U(t)$ at the output of the filter,

$$U(t) = \sum_{i=0}^N X_i \sqrt{\lambda_i} \theta_i(t),$$

where $\varphi_i(t)$ and $\theta_i(t)$ are the desired basis functions of the input and the output respectively. With the presence of AWGN, the received signal will be

$$Y(t) = \sum_{i=0}^N (X_i \sqrt{\lambda_i} + N_i) \theta_i(t),$$

where N_i are independent Gaussian variables. Therefore, this system model is mapped to a DT model with parallel independent channels where

$$Y_i = \sqrt{\lambda_i} X_i + N_i \text{ for } i = 0, 1, 2, \dots, N.$$

In [7, Section 8.3] he considered the special case where $H(f)$ is the ideal low pass filter and the noise is white and found the capacity to be the same as the one derived by Shannon, while avoiding the issues of infinite noise power and infinite capacity. In that special case (where $H(f)$ is the ideal low pass filter) the transformation between continuous time and discrete time was also done through

the use of the PSWFs [7, Sections 8.4 and 8.5]. Finally, Gallager provided an intuitive argument regarding inter-codeword interference [7, Section 8.5]; one can introduce a large guard time, say $T^{1-\epsilon}$ for some $\epsilon > 0$. Asymptotically, inter-codeword interference is avoided without affecting the data rates since $T^{1-\epsilon}/T \rightarrow 0$ as $T \rightarrow \infty$.

Chapter 3

Combined PAM-OMM with finite duration pulses

In this chapter, I propose a system model that consists of sending T -seconds time-limited pulses over W -Hz band-limited Gaussian channel: I specifically consider a combined PAM-OMM system. I take an information-theoretic approach and I derive the achievable rates over this system and evaluate them numerically. These rates are naturally obtained while allowing codewords to have arbitrarily large block-length. With a view towards practical implementations, although the CT codewords have infinite duration, the pulses are restricted to be time-limited as is commonly used in practice.

In my derivations, I rely on techniques inspired by the works in [26] and [27] where Tsybakov derived the capacity of Gaussian channels under ISI in the first and Telatar derived the channel capacity of multi-antenna systems in the second. In broad terms, I conduct the analysis in the frequency domain and I modify my problem in a manner akin to what is done for multi-antenna systems to be able to formulate my optimization problem.

Also I use the PSWFs and their properties in my derivations and numerical computations (I refer the reader to [19] and [7, section 8.4] for the definition and properties of the PSWFs, some of which are summarized in section 2.1). Note that the use of the PSWFs is not fundamental in this chapter and other time-limited filters can be used. In this chapter I will use unique values for the time window T and the bandwidth W , then $c = 2WT$ is fixed in my system and therefore I denote the PSWF $\varphi_{c,i}(t)$ by $\varphi_i(t)$ and I denote its corresponding eigenvalue $\lambda_{c,i}$ by λ_i . Also I denote the Fourier transform of PSWF by $\Phi_i(f)$ and I denote the $[-\frac{T}{2}, \frac{T}{2}]$ time-limited version of PSWF by $D\varphi_i(t)$.

3.1 System model

I consider the combined PAM-OMM system illustrated in Figure 3.1, where the

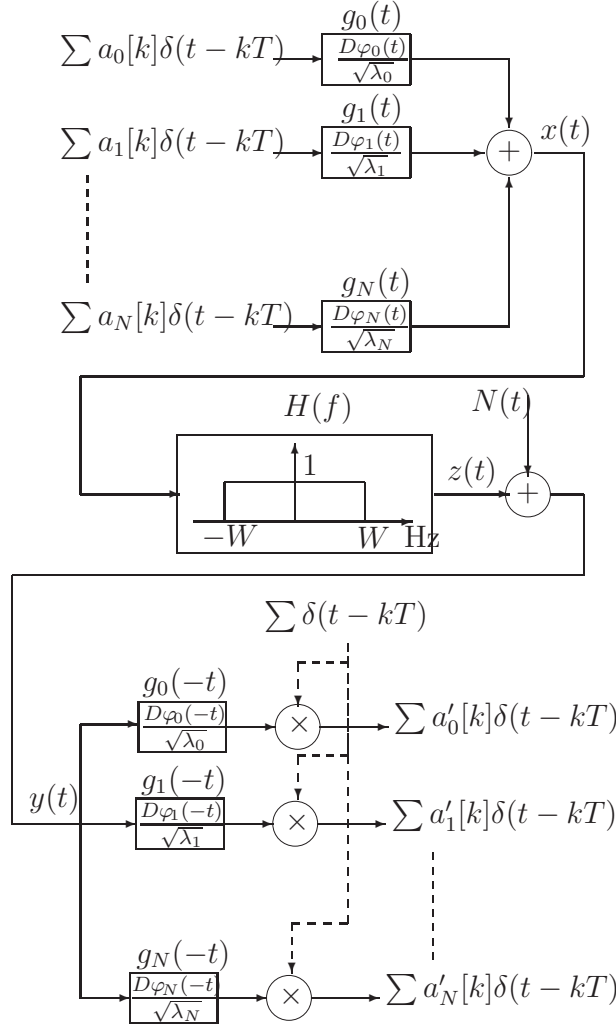


Figure 3.1: Combined PAM-OMM System Model.

channel is an ideal low pass filter with cut-off frequency W , and the additive noise $N(t)$ is a real W -Hz band-limited Wide Sense Stationary (WSS), Gaussian noise process with mean zero, and with PSD $S_N(f) = \frac{N_0}{2}$ when $f \in [-W, W]$.

The transmitted signal $x(t)$ consists of a sequence of T -seconds time-limited signals,

$$x(t) = \sum_{m=0}^N \sum_{l=-\infty}^{\infty} a_m[l] g_m(t - lT).$$

The T -seconds time-limited real pulses $\{g_m(t)\}_{m=0}^N$ are chosen from the set of normalized time-limited PSWFs $\{g_m(t) \doteq \frac{1}{\sqrt{\lambda_m}} D\varphi_m(t)\}_m$ which forms a CON set for all such signals. This choice is motivated by the work of Gallager [7,

section 8.4] who showed that the PSWFs are desirable functions when sending time-limited signals over a band-limited channel.

An average power constraint is imposed on the input whose average transmission power is

$$P_{avg} = \frac{\sum_{m=0}^N \mathbb{E}[|a_m|^2]}{T}. \quad (3.1)$$

Since the channel is an ideal low pass filter with cut-off frequency W , using the properties of the PSWF (refer to section 2.1)

$$z(t) = \sum_{m=0}^N \sum_{l=-\infty}^{\infty} a_m[l] \sqrt{\lambda_m} \varphi_m(t - lT).$$

On the receiver side I assume that the filters are also restricted to be T -seconds time-limited and I use the bank of filters consisting of the normalized time-limited PSWFs $\left\{ \frac{D\varphi_m(-t)}{\sqrt{\lambda_m}} \right\}_{m=0}^N$.

Interestingly, the choice of the normalized time-limited PSWFs at the receiver allows extracting sufficient statistics; the received signal $y(t)$ is W -Hz band-limited and hence sufficient statistics can be extracted by sampling $y(t)$ at a rate of $2W$ samples per second. Following the derivations for both cases leads to the same optimization problem as shown in section 3.4, where it is shown that in both cases the rates are computed using the water-filling solution over the eigenvalues of the same matrix.

Or by intuition, sufficient statistics can be extracted by projecting $y(t)$ on the set $\{\varphi_m(t - kT)\}_{m=0, k=-\infty}^{m=N, k=\infty}$ since the signal of interest $z(t)$ lives in the space spanned by this set. And by the properties of PSWFs, projecting a band-limited signal on $\frac{D\varphi_m(t-kT)}{\sqrt{\lambda_m}}$ is equal to its projection on $\varphi_m(t - kT)$ with a scale of $\sqrt{\lambda_m}$. In my system the information symbols and the noise symbols will be scaled by the same factor and hence the mutual information is the same in both cases. Also, performing the derivations for both cases leads to the same results and in the following I focus on the use of $\left\{ \frac{D\varphi_m(-t)}{\sqrt{\lambda_m}} \right\}_{m=0}^N$ at the receiver.

At the output of the n^{th} filter, the k^{th} sample is

$$\begin{aligned} d'_n[k] &= \sum_{m=0}^N \sum_{l=-\infty}^{+\infty} \frac{\sqrt{\lambda_m}}{\sqrt{\lambda_n}} a_m[l] \langle \varphi_m(t - lT), D\varphi_n(t - kT) \rangle \\ &\quad + \frac{1}{\sqrt{\lambda_n}} \langle N(t), D\varphi_n(t - kT) \rangle \\ &= \sum_{m=0}^N (s_{n,m} * a_m)[k] + w_n[k], \end{aligned} \quad (3.2)$$

where

$$s_{n,m}[k-l] = \frac{\sqrt{\lambda_m}}{\sqrt{\lambda_n}} \langle \varphi_m(t), D\varphi_n(t - (k-l)T) \rangle$$

and

$$w_n[k] = \frac{1}{\sqrt{\lambda_n}} \langle N(t), D\varphi_n(t - kT) \rangle.$$

The noise sequence $w_n[k]$ is obtained by passing first $N(t)$ through the bank of filters with impulse responses $\frac{D\varphi_n(-t)}{\sqrt{\lambda_n}}$, and then sampling the resultant signals $v_n(t)$ at kT . The $\{v_n(\cdot)\}$'s are Jointly WSS (JWSS), zero-mean Gaussian processes with cross-spectrum

$$S_{v_m v_n}(f) = \sqrt{\lambda_m} \sqrt{\lambda_n} \frac{N_0}{2} \Phi_m(-f) \Phi_n(f) = \frac{N_0}{2} \left[\sqrt{\lambda_m} \Phi_m^*(f) \right] \left[\sqrt{\lambda_n} \Phi_n(f) \right],$$

where I used the fact that the Fourier transform of $\frac{D\varphi_n(t)}{\sqrt{\lambda_n}}$ on the band $[-W, W]$ is $\sqrt{\lambda_n} \Phi_n(f)$. The $\{w_n[\cdot]\}$'s are therefore JWSS, zero-mean Gaussian processes with cross-PSD

$$S_{w_m w_n}(e^{j2\pi f}) = \frac{N_0}{2} \times \sum_{k \in \mathbb{Z}} \left[\sqrt{\frac{\lambda_m}{T}} \Phi_m^* \left(\frac{f-k}{T} \right) \right] \left[\sqrt{\frac{\lambda_n}{T}} \Phi_n \left(\frac{f-k}{T} \right) \right].$$

Writing equation (3.2) in the frequency domain, for f in $[-\frac{1}{2}, \frac{1}{2}]$,

$$A'_n(e^{j2\pi f}) = \sum_{m=0}^N S_{n,m}(e^{j2\pi f}) A_m(e^{j2\pi f}) + W_n(e^{j2\pi f})$$

where

$$S_{n,m}(e^{j2\pi f}) = \sum_{k \in \mathbb{Z}} \sqrt{\frac{\lambda_n}{T}} \Phi_n^* \left(\frac{f-k}{T} \right) \sqrt{\frac{\lambda_m}{T}} \Phi_m \left(\frac{f-k}{T} \right), \quad (3.3)$$

and using a vector notation

$$\mathbf{A}'(e^{j2\pi f}) = \mathbf{S}(e^{j2\pi f}) \mathbf{A}(e^{j2\pi f}) + \mathbf{W}(e^{j2\pi f}), \quad (3.4)$$

where

$$\begin{aligned}
\mathbf{A}'(e^{j2\pi f}) &= \begin{bmatrix} A'_0(e^{j2\pi f}) \\ A'_1(e^{j2\pi f}) \\ \vdots \\ A'_N(e^{j2\pi f}) \end{bmatrix}, \\
\mathbf{S}(e^{j2\pi f}) &= \begin{bmatrix} S_{0,0}(e^{j2\pi f}) & S_{0,1}(e^{j2\pi f}) & \dots & S_{0,N}(e^{j2\pi f}) \\ S_{1,0}(e^{j2\pi f}) & S_{1,1}(e^{j2\pi f}) & \dots & S_{1,N}(e^{j2\pi f}) \\ \vdots & \vdots & \ddots & \vdots \\ S_{N,0}(e^{j2\pi f}) & S_{N,1}(e^{j2\pi f}) & \dots & S_{N,N}(e^{j2\pi f}) \end{bmatrix}, \\
\mathbf{A}(e^{j2\pi f}) &= \begin{bmatrix} A_0(e^{j2\pi f}) \\ A_1(e^{j2\pi f}) \\ \vdots \\ A_N(e^{j2\pi f}) \end{bmatrix}, \quad \mathbf{W}(e^{j2\pi f}) = \begin{bmatrix} W_0(e^{j2\pi f}) \\ W_1(e^{j2\pi f}) \\ \vdots \\ W_N(e^{j2\pi f}) \end{bmatrix}
\end{aligned} \tag{3.5}$$

with the auto-covariance matrix of $\mathbf{W}(e^{j2\pi f})$

$$\Lambda_{\mathbf{W}}(e^{j2\pi f}) = \frac{N_0}{2} \mathbf{S}(e^{j2\pi f}).$$

3.2 Problem formulation

Let f be a frequency in the range $[-\frac{1}{2}, \frac{1}{2}]$ and denote by r the number of aliased components in $\mathbf{S}(e^{j2\pi f})$ at frequency f .

The matrix $\mathbf{S}(e^{j2\pi f})$ can be seen as the sum of r dyadic products:¹

$$\mathbf{S}(e^{j2\pi f}) = \sum_{k \in \mathbb{Z}} \mathbf{v}(f - k) \mathbf{v}^\dagger(f - k),$$

where

$$\mathbf{v}(f) = \begin{bmatrix} \sqrt{\frac{\lambda_0}{T}} \Phi_0^*\left(\frac{f}{T}\right) \\ \sqrt{\frac{\lambda_1}{T}} \Phi_1^*\left(\frac{f}{T}\right) \\ \vdots \\ \sqrt{\frac{\lambda_N}{T}} \Phi_N^*\left(\frac{f}{T}\right) \end{bmatrix}, \tag{3.6}$$

and since the rank of each dyadic product is less or equal to one, then the rank of $\mathbf{S}(e^{j2\pi f})$ is less or equal r . When $2WT$ is an integer, the number of aliased components at all frequencies in $[-\frac{1}{2}, \frac{1}{2}]$ is equal to $2WT$, and therefore given the time window T and the bandwidth W , there are at most $2WT$ independent

¹I denote the conjugate transpose of a given matrix \mathbf{M} by \mathbf{M}^\dagger

symbols. In what follows, possibly more than r PSWFs are used and hence $\mathbf{S}(e^{j2\pi f})$ will not have a full rank.

Since $\mathbf{S}(e^{j2\pi f})$ is a Hermitian matrix, it can be decomposed as

$$\mathbf{S}(e^{j2\pi f}) = \mathbf{U}(e^{j2\pi f}) \mathbf{D}(e^{j2\pi f}) \mathbf{U}^\dagger(e^{j2\pi f}),$$

where $\mathbf{U}(e^{j2\pi f})$ is a unitary matrix the columns of which are the eigenvectors of $\mathbf{S}(e^{j2\pi f})$, and $\mathbf{D}(e^{j2\pi f})$ is a diagonal matrix with only r non-zero elements that correspond to the eigenvalues of $\mathbf{S}(e^{j2\pi f})$. Since $\mathbf{U}(e^{j2\pi f})$ is unitary, multiplying by $\mathbf{U}^\dagger(e^{j2\pi f})$ on both sides in equation (3.4) yields

$$\mathbf{U}^\dagger(e^{j2\pi f}) \mathbf{A}'(e^{j2\pi f}) = \mathbf{D}(e^{j2\pi f}) \mathbf{U}^\dagger(e^{j2\pi f}) \mathbf{A}(e^{j2\pi f}) + \mathbf{W}'(e^{j2\pi f}), \quad (3.7)$$

where $\mathbf{W}'(e^{j2\pi f}) = \mathbf{U}^\dagger(e^{j2\pi f}) \mathbf{W}(e^{j2\pi f})$ with an auto-covariance matrix

$$\Lambda'_{\mathbf{W}}(e^{j2\pi f}) = \frac{N_0}{2} \mathbf{D}(e^{j2\pi f}).$$

Because $\mathbf{D}(e^{j2\pi f})$ does not have a full rank, parts of the matrices in equation (3.7) are statistically irrelevant; let $\mathbf{D}_r(e^{j2\pi f})$ to be an $(r \times r)$ diagonal matrix that contains the non-zero eigenvalues of $\mathbf{S}(e^{j2\pi f})$ on its main diagonal, and let $\mathbf{U}_r^\dagger(e^{j2\pi f})$ to be the sub-matrix of $\mathbf{U}^\dagger(e^{j2\pi f})$ that contains only the r rows that correspond to the non-zero eigenvalues in $\mathbf{D}(e^{j2\pi f})$. The matrix $\mathbf{S}(e^{j2\pi f}) = \mathbf{U}_r(e^{j2\pi f}) \mathbf{D}_r(e^{j2\pi f}) \mathbf{U}_r^\dagger(e^{j2\pi f})$ and sufficient statistics in equation (3.7) are:

$$\mathbf{U}_r^\dagger(e^{j2\pi f}) \mathbf{A}'(e^{j2\pi f}) = \mathbf{D}_r(e^{j2\pi f}) \mathbf{U}_r^\dagger(e^{j2\pi f}) \mathbf{A}(e^{j2\pi f}) + \mathbf{U}_r^\dagger(e^{j2\pi f}) \mathbf{W}(e^{j2\pi f}). \quad (3.8)$$

Since $\mathbf{D}_r(e^{j2\pi f})$ is a diagonal matrix with positive elements in the main diagonal, it has a square root matrix. Define $\hat{\mathbf{D}}_r(e^{j2\pi f})$ to be the inverse of that square root matrix and multiplying equation (3.8) by $\hat{\mathbf{D}}_r(e^{j2\pi f})$ on both sides we get

$$\ddot{\mathbf{A}}'(e^{j2\pi f}) = \Sigma(e^{j2\pi f}) \ddot{\mathbf{A}}(e^{j2\pi f}) + \ddot{\mathbf{W}}(e^{j2\pi f}), \quad (3.9)$$

where

$$\begin{aligned} \ddot{\mathbf{A}}'(e^{j2\pi f}) &= \hat{\mathbf{D}}_r(e^{j2\pi f}) \mathbf{U}_r^\dagger(e^{j2\pi f}) \mathbf{A}'(e^{j2\pi f}), \\ \Sigma(e^{j2\pi f}) &= \hat{\mathbf{D}}_r(e^{j2\pi f}) \mathbf{D}_r(e^{j2\pi f}), \\ \ddot{\mathbf{A}}(e^{j2\pi f}) &= \mathbf{U}_r^\dagger(e^{j2\pi f}) \mathbf{A}(e^{j2\pi f}), \\ \ddot{\mathbf{W}}(e^{j2\pi f}) &= \hat{\mathbf{D}}_r(e^{j2\pi f}) \mathbf{U}_r^\dagger(e^{j2\pi f}) \mathbf{W}(e^{j2\pi f}), \end{aligned} \quad (3.10)$$

where the auto-covariance of $\ddot{\mathbf{W}}(e^{j2\pi f})$ is given by

$$\Lambda_{\ddot{\mathbf{W}}}(e^{j2\pi f}) = \frac{N_0}{2} \hat{\mathbf{D}}_r(e^{j2\pi f}) \mathbf{D}_r(e^{j2\pi f}) \hat{\mathbf{D}}_r^\dagger(e^{j2\pi f}) = \frac{N_0}{2} \mathbf{1}.$$

Note that the trace of the auto-covariance of $\ddot{\mathbf{A}}(e^{j2\pi f})$ is

$$\text{tr}(\Lambda_{\ddot{\mathbf{A}}}(e^{j2\pi f})) = \text{tr}(\Lambda_{\mathbf{A}}(e^{j2\pi f})).$$

Additionally, the average power in equation (3.1) being less than the total available or allowed transmission power P , can be written as:

$$\frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{tr}(\Lambda_{\ddot{\mathbf{A}}}(e^{j2\pi f})) df \leq P. \quad (3.11)$$

It is clear that $\Sigma(e^{j2\pi f})$ is diagonal and it contains the square root of the non-zero eigenvalues of $\mathbf{S}(e^{j2\pi f})$ on its main diagonal. Therefore the entries of $\Sigma(e^{j2\pi f})$ are bounded since the entries of $\mathbf{S}(e^{j2\pi f})$ are bounded [28]. Moreover, $\Sigma_{(i,i)}(e^{j2\pi f})$ are continuous functions of f since the entries of $\mathbf{S}(e^{j2\pi f})$ are continuous [29] and have consequently inverse DT Fourier transforms $\sigma_i[k]$ that are square-summable. Writing (3.9) back in time domain we get

$$\ddot{a}'_i[k] = (\sigma_i * \ddot{a}_i)[k] + \ddot{w}_i[k], \quad 1 \leq i \leq r. \quad (3.12)$$

The problem at hand is that of the achievable rates of channel (3.12) under (3.11). By Tsybakov [26],

$$R = \max_{\Lambda_{\ddot{\mathbf{A}}}(e^{j2\pi f})} \frac{1}{2T} \sum_{i=1}^r \int_{-\frac{1}{2}}^{\frac{1}{2}} \log_2 \left(\frac{\Sigma_{(i,i)}^2(e^{j2\pi f}) \Lambda_{\ddot{\mathbf{A}}(i,i)}(e^{j2\pi f})}{\frac{N_0}{2}} + 1 \right) df \quad (3.13)$$

subject to $\frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{tr}(\Lambda_{\ddot{\mathbf{A}}}(e^{j2\pi f})) df \leq P.$

The optimization problem in equation (3.13) is convex and it has the well-known water-filling solution:

$$\Lambda_{\ddot{\mathbf{A}}(i,i)}(e^{j2\pi f}) = \left(\nu - \frac{N_0}{2} \frac{1}{\Sigma_{(i,i)}^2(e^{j2\pi f})} \right)^+,$$

and the water level ν is selected such that

$$\frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{tr}(\Lambda_{\ddot{\mathbf{A}}}(e^{j2\pi f})) df = P.$$

3.2.1 Asymptotic analysis as $N \rightarrow \infty$

For any practical system N must be finite. But at this stage I consider the asymptotic case to find an upper bound on the achievable rates; in the following theorem I show that the achievable rates converge to Shannon's capacity as N increases towards infinity.

Using Mercer's theorem (2.7)

$$\sum_{i=0}^{\infty} \frac{\lambda_i}{T} \Phi_i \left(\frac{f}{T} \right) \Phi_i^* \left(\frac{f}{T} \right) = 1 \quad (3.14)$$

and

$$\sum_{i=0}^{\infty} \frac{\lambda_i}{T} \Phi_i \left(\frac{f}{T} \right) \Phi_i^* \left(\frac{f-k}{T} \right) = 0, \quad (3.15)$$

for any non-zero integer k . This indicates that asymptotically the vectors $\{\mathbf{v}(f-k)\}_k$ are orthonormal.

Theorem 1. *For integer values of $c = 2WT$ such that $c \geq 1$, there are $2WT$ degrees of freedom and the achievable rates converge to Shannon's capacity as $N \rightarrow \infty$ in the considered system model.*

Proof. Asymptotically and as $N \rightarrow \infty$, let $c = 2WT$ be a strictly positive integer. The matrix $\mathbf{S}(e^{j2\pi f})$ is the sum of c dyadic products at every $f \in [-0.5, 0.5]$. Moreover, by (3.15) $\{\mathbf{v}(f-k)\}_k$ forms an orthonormal set for $|k| \leq c/2$ and therefore $\mathbf{S}(e^{j2\pi f})$ has c non-zero eigenvalues that are equal to 1, i.e.,

$$\sum_{(i,i)}^2 (e^{j2\pi f}) = 1 \quad i \leq c.$$

Therefore, the water-filling solution in (3.13) is the equal-power allocation solution such that $\Lambda_{\mathbf{A}(i,i)}(e^{j2\pi f}) = \frac{PT}{c}$ and consequently, as $N \rightarrow \infty$

$$R = \frac{c}{2T} \log_2 \left(1 + \frac{2PT}{N_0 c} \right) = W \log_2 \left(1 + \frac{P}{N_0 W} \right),$$

which is equal to Shannon's formula. \square

Moreover, it can be shown that the rates converge to Shannon's capacity for any strictly positive real values of c ; there will be $\lfloor c \rfloor$ degrees of freedom over an interval of measure $\lfloor c \rfloor - c$, and $\lceil c \rceil$ degrees of freedom on the remaining part of the band $[-0.5, 0.5]$. Therefore there will be c degrees of freedom on average, and the non-zero eigenvalues of $\mathbf{S}(e^{j2\pi f})$ will be equal to 1.

3.3 Numerical results

I perform the numerical computations to find the available degrees of freedom and the achievable data rates. The integral in equation (3.13) is computed using Riemann integral with 2097152 uniformly distributed samples for $-0.5 \leq f \leq 0.5$.

3.3.1 Degrees of freedom

As presented earlier in section 3.2, the degrees of freedom are determined by the rank of the matrix $\mathbf{S}(e^{j2\pi f})$. For $c \in \{1, 2, 3, 6\}$, my numerical computations show that when using at least the first c PSWFs, the rank of the matrix $\mathbf{S}(e^{j2\pi f})$ is equal to c for all the 2097152 samples. Moreover, the computed non-zero eigenvalues of $\mathbf{S}(e^{j2\pi f})$ are not close to zero. Therefore one can claim that the rank of the matrix $\mathbf{S}(e^{j2\pi f})$ is c almost everywhere for $-0.5 \leq f \leq 0.5$ (since the eigenvalues are continuous as proved in section 3.2).

Fig. 3.2 shows the histogram of the computed eigenvalues of $\mathbf{S}(e^{j2\pi f})$ when using the first $N + 1$ PSWFs for different combinations of c and N . Note that the eigenvalues get closer to 1 as N increases, therefore the achievable rates are expected to approach Shannon's capacity as N increases and this is validated in the next section.

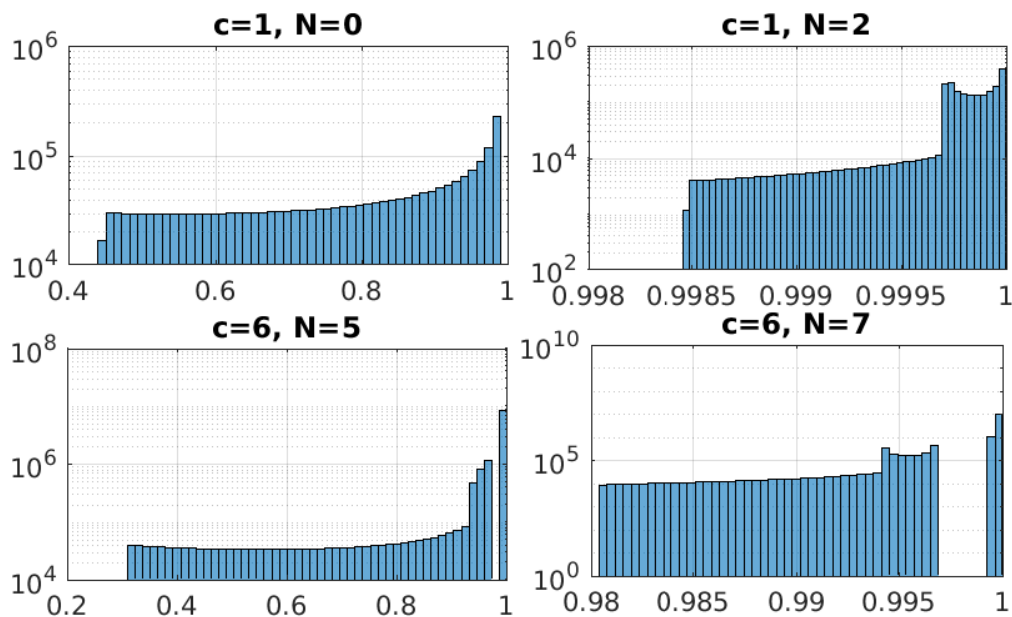


Figure 3.2: Histograms of the eigenvalues of $\mathbf{S}(e^{j2\pi f})$ for different c and N

3.3.2 Achievable rates

In the following results, I fixed the radio band to 1000Hz, and I focused on the percentage decrease between the achieved rates and Shannon's formula. I note from equation (3.13) that the percentage decrease is the same for fixed $\frac{P}{N_0W} = \frac{\text{SNR}}{2W}$, and therefore, the results for $W \neq 1\text{KHz}$ can be extracted from my presented results over some range of SNR.

The question that will be addressed is how many PSWFs should be used in the proposed system. Although the rank of $\Sigma(f)$ is equal to $c = 2WT$, the column vectors of $U_r^\dagger(f)$ given in equation (3.10) lead to the use of the available PSWFs with different weights. By intuition, the use of PSWFs with almost zero eigenvalues does not achieve a valuable improvement on the achievable rates of the system. In my computations I calculated the information rates when using different numbers of PSWFs; and as expected, the rates saturate after the use of a certain number of PSWFs, beyond which adding more PSWFs has negligible effect on the achievable data rates. Interestingly, and as I proved in 3.2.1, the computed rates approach the Shannon's capacity as the number of used PSWFs increases.

Fig. 3.3 shows the percentage decrease between the calculated rates and Shannon's formula for $c = 1$ when using the first $N + 1$ PSWFs. And as shown in this figure, using the first four PWSFs achieves more than 99.999% of the Shannon formula in my selected SNR range.

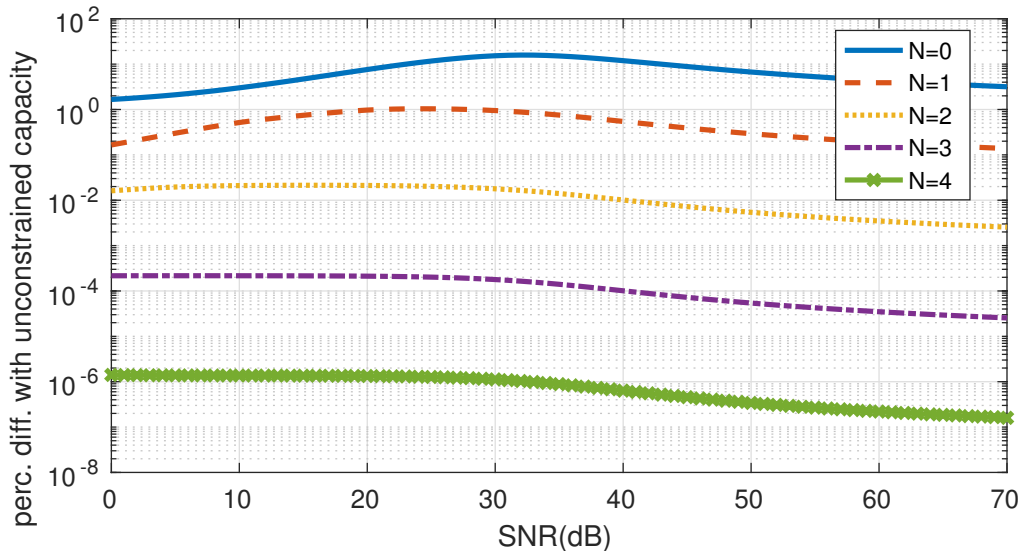


Figure 3.3: Convergence of achievable rates for $c=1$

Table 3.1 presents the maximum percentage decrease between the calculated achievable rates and Shannon's capacity for different values of c and when using the first $N + 1$ time-limited PSWFs as parallel filters. Note that, for the selected values of c and over the considered SNR range, the achievable rates can be made arbitrary close to the channel capacity by adding finite number of filters. Interestingly, note that the maximum percentage difference is approximately 1% for $c = 1$ when using just the first two time-limited PSWFs as parallel filters. However, as c increases, the convergence towards Shannon's capacity becomes

slower.

	$c = 1$	$c = 2$	$c = 3$	$c = 6$
$N = c - 1$	15.84%	9.16%	6.63%	3.77%
$N = c$	1.04%	1.12%	1.05%	0.84%
$N = c + 1$	0.021%	0.053%	0.073%	0.099%
$N = c + 2$	0.00022%	0.0014%	0.0031%	0.0078%

Table 3.1: Maximum percentage decrease with Shannon's formula

Fig. 3.4 presents the difference between the calculated achievable rates and Shannon's capacity for a range of SNR and different values of c when using only the first c PSWFs. The results show that the achieved rates increase with c and they become closer to Shannon's capacity, which is consistent with the results by Wyner [5]- [6]; Wyner proved that as $c \rightarrow \infty$, it is sufficient to use the first c time-limited PSWFs (using an OMM system) to achieve Shannon's capacity.

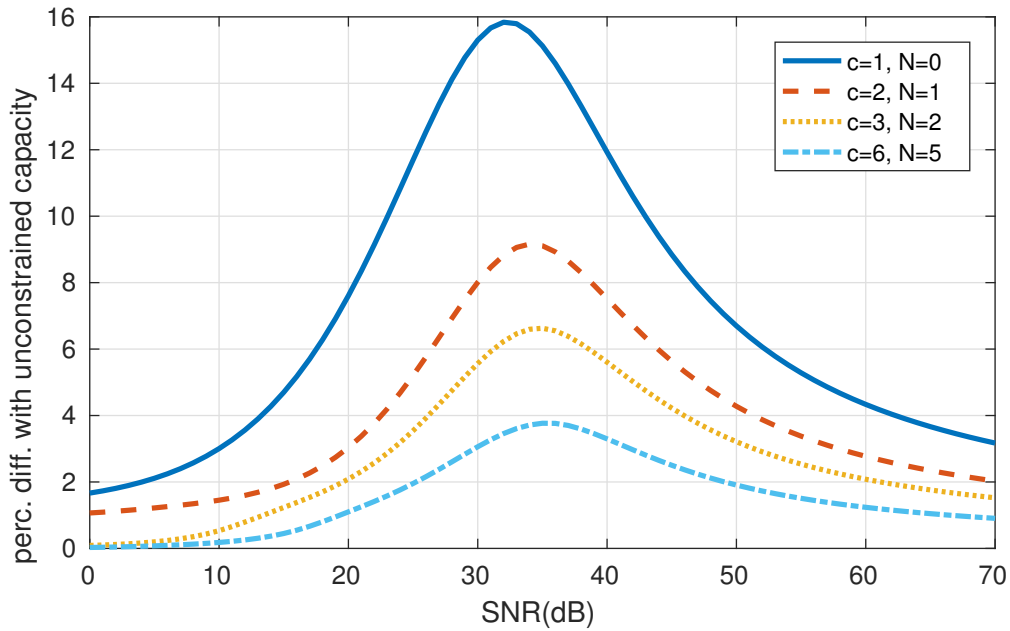


Figure 3.4: Percentage decrease with Shannon's formula for different c

3.4 Sampling the received signal $y(t)$

Being band-limited, extracting sufficient statistics from the received signal may be done through sampling it at a rate of $2W$ samples per second.

For $k \in \mathbb{Z}$ and $n \in \{0, 1, \dots, c-1\}$, $\{kc+n\}_{k,c}$ span the set of all integers. The samples at time $\frac{kc+n}{2W}$ are

$$\begin{aligned} a'_n[k] &= \sum_{m=0}^N \sum_{l=-\infty}^{+\infty} a_m[l] \sqrt{\lambda_m} \varphi_m \left(\left[k - l + \frac{n}{c} \right] T \right) + w_n[k] \\ &= \sum_{m=0}^N (b_{n,m} * a_m)[k] + w_n[k], \end{aligned} \quad (3.16)$$

where

$$\begin{aligned} b_{n,m}[k-l] &= \sqrt{\lambda_m} \varphi_m \left([k-l]T + \frac{n}{c}T \right), \\ w_n[k] &= N \left(kT + \frac{n}{2W} \right). \end{aligned}$$

Writing equation (3.16) in the frequency domain, for f in $[-\frac{1}{2}, \frac{1}{2}]$,

$$A'_n(e^{j2\pi f}) = \sum_{m=0}^N B_{n,m}(e^{j2\pi f}) A_m(e^{j2\pi f}) + W_n(e^{j2\pi f})$$

where,

$$\begin{aligned} B_{n,m}(e^{j2\pi f}) &= \sum_{k \in \mathbb{Z}} \sqrt{\lambda_m} \varphi_m \left(kT + \frac{nT}{c} \right) e^{-j2\pi f k} \\ &= \sqrt{\lambda_m} \int_{-\infty}^{+\infty} \varphi_m(t) e^{-j2\pi \frac{f}{T} t} e^{j2\pi f \frac{n}{c}} \sum_{k \in \mathbb{Z}} \delta \left[t - kT - \frac{nT}{c} \right] dt \\ &= e^{j2\pi f \frac{n}{c}} \sqrt{\lambda_m} \int_{-\infty}^{+\infty} \varphi_m(t) \sum_{k \in \mathbb{Z}} \delta \left[t - kT - \frac{nT}{c} \right] e^{-j2\pi \frac{f}{T} t} dt \\ &= e^{j2\pi f \frac{n}{c}} \sqrt{\lambda_m} \mathcal{FT} \left\{ \varphi_m(t) \sum_{k \in \mathbb{Z}} \delta \left[t - kT - \frac{nT}{c} \right] \right\} \Big|_{\frac{f}{T}} \\ &= e^{j2\pi f \frac{n}{c}} \frac{\sqrt{\lambda_m}}{T} \left\{ \Phi_m(f) * \sum_{k \in \mathbb{Z}} \delta \left[f - \frac{k}{T} \right] e^{-j2\pi k \frac{n}{c}} \right\} \Big|_{\frac{f}{T}} \\ &= e^{j2\pi f \frac{n}{c}} \frac{\sqrt{\lambda_m}}{T} \left\{ \sum_{k \in \mathbb{Z}} \Phi_m \left(\frac{f}{T} - \frac{k}{T} \right) e^{-j2\pi k \frac{n}{c}} \right\} \\ &= \frac{\sqrt{\lambda_m}}{T} \sum_{k \in \mathbb{Z}} \Phi_m \left(\frac{f-k}{T} \right) e^{j2\pi(f-k)\frac{n}{c}}. \end{aligned}$$

Using a vector notation

$$\mathbf{A}'(e^{j2\pi f}) = \mathbf{B}(e^{j2\pi f}) \mathbf{A}(e^{j2\pi f}) + \mathbf{W}(e^{j2\pi f}), \quad (3.17)$$

where

$$\begin{aligned} \mathbf{A}'(e^{j2\pi f}) &= \begin{bmatrix} A'_0(e^{j2\pi f}) \\ A'_1(e^{j2\pi f}) \\ \vdots \\ A'_{c-1}(e^{j2\pi f}) \end{bmatrix} & \mathbf{A}(e^{j2\pi f}) &= \begin{bmatrix} A_0(e^{j2\pi f}) \\ A_1(e^{j2\pi f}) \\ \vdots \\ A_N(e^{j2\pi f}) \end{bmatrix} \\ \mathbf{B}(e^{j2\pi f}) &= \begin{bmatrix} B_{0,0}(e^{j2\pi f}) & \dots & B_{0,N}(e^{j2\pi f}) \\ B_{1,0}(e^{j2\pi f}) & \dots & B_{1,N}(e^{j2\pi f}) \\ \vdots & \ddots & \vdots \\ B_{c-1,0}(e^{j2\pi f}) & \dots & B_{c-1,N}(e^{j2\pi f}) \end{bmatrix} \\ \& \mathbf{W}(e^{j2\pi f}) &= \begin{bmatrix} W_0(e^{j2\pi f}) \\ W_1(e^{j2\pi f}) \\ \vdots \\ W_{c-1}(e^{j2\pi f}) \end{bmatrix}. \end{aligned}$$

Note that

$$\mathbf{B}(e^{j2\pi f}) = \frac{1}{\sqrt{T}} \sum_{k \in \mathbb{Z}} \mathbf{e}(f-k) \mathbf{v}^\dagger(f-k),$$

where

$$\mathbf{e}(f) = \begin{bmatrix} 1 \\ e^{j2\pi f \frac{1}{c}} \\ \vdots \\ e^{j2\pi f \frac{c-1}{c}} \end{bmatrix} \quad \mathbf{v}(f) = \begin{bmatrix} \sqrt{\frac{\lambda_0}{T}} \Phi_0^*\left(\frac{f}{T}\right) \\ \sqrt{\frac{\lambda_1}{T}} \Phi_1^*\left(\frac{f}{T}\right) \\ \vdots \\ \sqrt{\frac{\lambda_N}{T}} \Phi_N^*\left(\frac{f}{T}\right) \end{bmatrix}$$

Since

$$\mathbf{e}^\dagger(f-k) \mathbf{e}(f-l) = \sum_{n=0}^{c-1} e^{j2\pi(k-l)\frac{n}{c}} = c \delta[k-l],$$

then

$$\begin{aligned} \mathbf{B}^\dagger(e^{j2\pi f}) \mathbf{B}(e^{j2\pi f}) &= \frac{1}{T} \sum_{k \in \mathbb{Z}} \mathbf{v}(f-k) \mathbf{e}^\dagger(f-k) \sum_{l \in \mathbb{Z}} \mathbf{e}(f-l) \mathbf{v}^\dagger(f-l) \\ &= \frac{1}{T} \sum_{k, l \in \mathbb{Z}} \mathbf{v}(f-k) \mathbf{e}^\dagger(f-k) \mathbf{e}(f-l) \mathbf{v}^\dagger(f-l) \\ &= \frac{c}{T} \sum_{k \in \mathbb{Z}} \mathbf{v}(f-k) \mathbf{v}^\dagger(f-k). \end{aligned}$$

The auto-covariance matrix of $\mathbf{W}(e^{j2\pi f})$ is diagonal and equal to

$$\Lambda_{\mathbf{w}}(e^{j2\pi f}) = W N_0 I_c = \frac{cN_0}{2T} I_c.$$

Note that $\mathbf{S}(e^{j2\pi f}) = \frac{T}{c} \mathbf{B}^\dagger(e^{j2\pi f}) \mathbf{B}(e^{j2\pi f})$ where $\mathbf{S}(e^{j2\pi f})$ is given by equation (3.5). Therefore, from an information theoretic perspective, the system given here by equation (3.17) is equivalent to the one derived in equation (3.9).

Chapter 4

PAM with finite duration pulses

In this chapter I study and evaluate the performance over the band-limited channel using *fixed* finite-duration transmit pulses. More specifically I limit my study to PAM systems and I evaluate the achievable information rates by optimizing over the two major design components: the pulse shape/receive filter and the power allocation with which the symbols are transmitted. I assume that the channel is an ideal low pass filter as explained in Section 1.2.

Perhaps the most popular PAM pulses/receive filters –both studied and widely used in practice– are the “Nyquist pulses”, i.e., ones where the Nyquist criterion is satisfied. This is due to the fact that they eliminate ISI and simplify consequently the receiver’s design. A necessary condition to satisfy the ISI-free criterion is to have $T \geq \frac{1}{2W}$, where T is the symbol rate and W is the bandwidth of the pulse.

A well known family of pulses that satisfies this criterion is the family of Raised Cosine (RC) filters with $2WT = 1 + \beta$ such that $0 \leq \beta \leq 1$. As β increases the decay of the pulse in time domain becomes faster, and hence truncating it will have lower effects. However, and for a fixed T , as β increases the required bandwidth increases.

I consider pulses in three cases of the signaling rate: $2WT > 1$, $2WT = 1$, and $2WT < 1$ and evaluate the information-theoretic achievable rates when using time-limited pulses. In each case, I compare to the achievable rates when using band-limited –and hence infinite duration– pulses, which I derive as well. Considering the case $2WT < 1$ is tightly related to the Faster Than Nyquist (FTN) signaling concept that was first proposed by Mazo [30] where he proved that speeding up the signaling of the “sinc” pulses by a factor of 25% does not affect the minimum distance between the codewords when using binary input [31]. This naturally introduces ISI and requires more sophisticated receivers. Part of my objective in this chapter is to study finite-duration pulses as well as the corresponding signaling rates. I will answer the fundamental question: “When using finite-duration pulses, is signaling FTN beneficial?”

4.1 System model and problem formulation

Let $g(t)$ and $\hat{g}(t)$ be two unit-energy real-valued functions that are time-limited to $[-\frac{c}{2}, \frac{c}{2}]$ for some positive scalar c . The interest is in the set of functions

$$\{g(t - k), \quad k \in \mathbb{Z}\},$$

which does not necessarily form an orthonormal set as the inner product

$$\begin{aligned} \int_{-\infty}^{+\infty} g(t - k) g(t - l) dt &= \int_{-\infty}^{+\infty} g(t - (k - l))g(t) dt \\ &= g(t) * g(-t) \Big|_{(k-l)}, \end{aligned}$$

is possibly only zero for $|k - l| \geq c$ and is generally not necessarily a “delta” function unless $c \leq 1$. In what follows I denote by $R_g[n]$ this sampled autocorrelation and by $R_g(e^{j2\pi f})$ its Fourier transform:

$$\begin{aligned} R_g[n] &= (g(t) * g(-t)) \Big|_{(n)} = \begin{cases} 1 & n = 0 \\ 0 & |n| \geq c \end{cases} \quad (4.1) \\ R_g(e^{j2\pi f}) &= \sum_{k \in \mathbb{Z}} |G(f - k)|^2. \end{aligned}$$

The objective of this chapter is the analysis of the use of PAM systems that are parametrized by a symbol interval T and transmit pulses and receive filters that are of a finite duration. More precisely, I use the PAM pulse-shaping filter $(1/\sqrt{T})g(t/T)$ (i.e., $\sqrt{T}G(fT)$ in the Fourier domain) and the receive filter $(1/\sqrt{T})\hat{g}(-t/T)$. Naturally, these functions are $[-\frac{cT}{2}, \frac{cT}{2}]$ time-limited and have unit-energy.

Figure 4.1 shows a classical PAM system when used over a band-limited linear Gaussian channel; symbols from a signal constellation are used to modulate the amplitude of the chosen pulse every T seconds, and $H(f)$ is the channel transfer function.

The transmitted signal can be written as:

$$X(t) = \frac{1}{\sqrt{T}} \sum_{k=-\infty}^{+\infty} A[k] g\left(\frac{t}{T} - k\right),$$

where the $\{A[k]\}_k$'s are chosen from a given real¹ signal constellation. I impose

¹For the complex AWGN channel, $\{A[k]\}_k$'s can be complex and my derivation accommodates the complex case

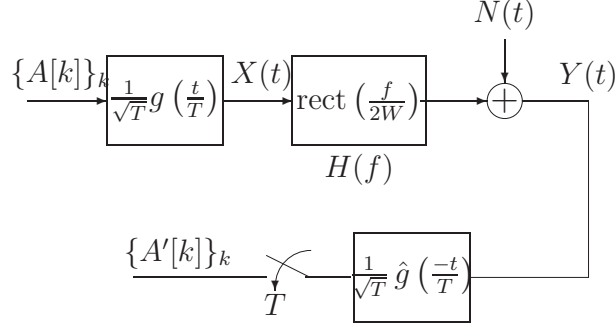


Figure 4.1: PAM System Model

on this transmitted signal an average power constraint P :

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{cT + 2nT} \int_{-nT - \frac{cT}{2}}^{nT + \frac{cT}{2}} \left| \sum_{k=-n}^n A[k] \frac{1}{\sqrt{T}} g\left(\frac{t}{T} - k\right) \right|^2 dt \leq P \\
& \Leftrightarrow \lim_{n \rightarrow \infty} \frac{1}{cT + 2nT} \sum_{k=-n}^n \sum_{l=-n}^n A[k] A^*[l] R_g[k-l] \leq P \\
& \Leftrightarrow \frac{1}{T} \sum_{m=-\infty}^{\infty} R_A[m] R_g[m] \leq P, \tag{4.2}
\end{aligned}$$

where I assumed the data symbols $\{A[k]\}$ form a stationary process with auto-correlation function $R_A[\cdot]$, and that the limit in (4.2) exists.

When it comes to the channel, I assume that it is an ideal W -Hz band-limited AWGN channel where the additive noise $N(t)$ is a stationary “White” Gaussian process with mean zero and PSD $S_N(f) = \frac{N_0}{2}$ for $f \in [-W, W]$. The output of the channel is

$$Y(t) = \frac{1}{\sqrt{T}} \sum_{k=-\infty}^{+\infty} A[k] b\left(\frac{t}{T} - k\right) + N(t).$$

where $b(t)$ is the function $g(t)$ “band-limited” to WT -Hz, i.e.,

$$\begin{aligned}
B(f) &= G(f) \text{rect}\left(\frac{f}{2WT}\right) \\
&\Leftrightarrow b(t) = 2WT \int_{-\frac{c}{2}}^{\frac{c}{2}} g(\tau) \text{sinc}(2WT(t - \tau)) d\tau.
\end{aligned}$$

A filter that is matched to the function $\frac{1}{\sqrt{T}} \hat{g}\left(\frac{t}{T}\right)$ is used at the receiver, the output of which is sampled every T seconds. The k^{th} sample is

$$A'[k] = \sum_{l=-\infty}^{+\infty} A[l] s[k-l] + V[k] = (s * A)[k] + V[k], \tag{4.3}$$

where the discrete sequence $\{s[\cdot]\}$ is

$$s[k] = \frac{1}{T} \left\langle b\left(\frac{t}{T}\right), \hat{g}\left(\frac{t}{T} - k\right) \right\rangle = b(t) * \hat{g}(-t) \Big|_{k-l} \quad (4.4)$$

$$\begin{aligned} S(e^{j2\pi f}) &= \sum_{k \in \mathbb{Z}} B(f - k) \hat{G}^*(f - k) \\ &= \sum_{k \in \mathbb{Z}} G(f - k) \hat{G}^*(f - k) \text{rect}\left(\frac{f - k}{2WT}\right). \end{aligned} \quad (4.5)$$

The DT process $\{V[\cdot]\}$ is the filtered noise $N(t)$ sampled every T seconds:

$$V[k] = \frac{1}{\sqrt{T}} \left\langle N(t), \hat{g}\left(\frac{t}{T} - k\right) \right\rangle,$$

and is hence a DT stationary zero-mean Gaussian process with PSD

$$S_V(e^{j2\pi f}) = \frac{N_o}{2} \sum_{k \in \mathbb{Z}} \left| \hat{G}(f - k) \right|^2 \text{rect}\left(\frac{f - k}{2WT}\right). \quad (4.6)$$

4.1.1 Optimization

My objective is to maximize the mutual information when the system described in section 4.1 is used, subject to a transmitted average power constraint (4.2) that can be written using Parseval as,

$$\frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} S_A(e^{j2\pi f}) \sum_{k \in \mathbb{Z}} |G(f - k)|^2 df \leq P. \quad (4.7)$$

The optimal solution is known to be achieved by a zero-mean DT Gaussian² process $\{A[n]\}_n$ and the remaining optimization I tackle is over the pulse shaping filter $g(t)$ and over the spectrum of $\{A[n]\}$. By Tsybakov [26] and equation (4.3) this optimization problem can be formulated as³

$$\begin{aligned} R &= \max_{S_A(\cdot), g(\cdot)} \frac{1}{2T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left[1 + \frac{S_A(e^{j2\pi f}) |S(e^{j2\pi f})|^2}{S_V(e^{j2\pi f})} \right] df \\ \text{subject to} & \quad \frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} S_A(e^{j2\pi f}) \sum_{k \in \mathbb{Z}} |G(f - k)|^2 df \leq P \\ & \quad g(t) \text{ time-limited to } \left[-\frac{c}{2}, \frac{c}{2} \right] \\ & \quad \int_{-\frac{c}{2}}^{\frac{c}{2}} |g(t)|^2 dt = 1. \end{aligned} \quad (4.8)$$

²Circular complex whenever the channel is complex

³If the channel is complex, the objective function is twice the one in (4.8)

In the numerical results presented hereafter, I fix the radio band to 1000 Hz and quantify the percentage difference between the achieved data rates and Shannon's expression (2.15). Note that the achievable rates are only dependent on P, N_o and W through the ratio $\frac{P}{N_o W}$ and the product $2WT$, then the results for $W \neq 1$ KHz can be extracted from my presented results over some range of $\text{SNR} = 2P/N_o$.

In the remainder of this chapter I study three different cases for the relationship between W and T . In order to set benchmarks, I consider first the case where the transmit and receive filters are possibly of infinite duration.

4.2 Band-limited filters

I study first the system at hand whenever the transmit pulse and receive filter are allowed to be of infinite duration –and hence possibly band-limited. Given that the channel is band-limited, an optimal transmit pulse will naturally be band-limited as well.

4.2.1 The case where $2WT > 1$; Nyquist pulses

As customary in the literature, let $T = \frac{1+\beta}{2W}$ where β denotes “the excess bandwidth factor”.

The PAM system satisfies the Nyquist criterion if and only if $s[\cdot]$ defined in (4.4) is a unit sample, or equivalently

$$\sum_{k \in \mathbb{Z}} G(f-k) \hat{G}^*(f-k) \text{rect}\left(\frac{f-k}{2WT}\right) = 1. \quad (4.9)$$

When $\sqrt{T}G(fT)$ is band-limited to W -Hz, $G(f)$ is band-limited to $WT = \frac{1+\beta}{2}$ and $B(f) = G(f)$. The use of a matched filter to $b(t)$ at the receiver is information lossless, in which case equations (4.5) and (4.6) become

$$\begin{aligned} S(e^{j2\pi f}) &= \sum_{k \in \mathbb{Z}} G(f-k) G^*(f-k) \text{rect}\left(\frac{f-k}{1+\beta}\right) = 1 \\ S_V(e^{j2\pi f}) &= \frac{N_o}{2}. \end{aligned}$$

The optimal power allocation is known to be flat: $S_A(e^{j2\pi f}) = PT$ resulting in achievable rates equal to

$$\frac{1}{2T} \log \left[1 + \frac{2PT}{N_o} \right] = \frac{W}{1+\beta} \log \left[1 + \frac{P(1+\beta)}{N_o W} \right], \quad (4.10)$$

a result that is consistent with equation (20) in [32].

Equation (4.10) shows that the achievable rates decrease as β increases, and the largest value is equal to the Shannon capacity when $\beta \rightarrow 0$. This is consistent with the findings in [32] where the authors found that using Nyquist pulses imposes some “penalty” on the achievable rates, and that this penalty can be diminished by using FTN signaling, a subject I investigate further in Sections 4.2.3 and 4.3.3.

4.2.2 The case where $2WT = 1$

In this case, the only satisfying Nyquist’s criterion pulse and receive filter is the pair of “sinc” functions that achieve the well known formula by Shannon (see equation (4.10) with $\beta = 0$).

4.2.3 The case where $2WT < 1$

Now consider the case $W = \frac{1-\beta}{2T}$ where $0 < \beta < 1$. If one uses the W -Hz band-limited transmit pulse

$$\frac{1}{\sqrt{T}}g\left(\frac{t}{T}\right) = \sqrt{2W}\text{sinc}(2Wt),$$

and $\hat{g}(t) = g(t)$, then $\hat{G}(f) = G(f) = \frac{1}{\sqrt{1-\beta}}\text{rect}\left(\frac{f}{1-\beta}\right)$ and equations (4.5) and (4.6) become

$$S(e^{j2\pi f}) = \sum_{k \in \mathbb{Z}} \frac{1}{1-\beta} \text{rect}\left(\frac{f-k}{1-\beta}\right),$$

$$S_V(e^{j2\pi f}) = \frac{N_o}{2} \sum_{k \in \mathbb{Z}} \frac{1}{1-\beta} \text{rect}\left(\frac{f-k}{1-\beta}\right).$$

The optimizing spectrum is $S_A(e^{j2\pi f}) = PT \text{rect}\left(\frac{f}{1-\beta}\right)$ and achieves the rate

$$\frac{1-\beta}{2T} \log \left[1 + \frac{2PT}{(1-\beta)N_o} \right] = W \log \left[1 + \frac{P}{N_o W} \right],$$

which is the same as Shannon’s capacity expression, a result that is consistent with the results in [33].

In the following theorem I prove the strong statement that, provided some mild conditions, any band-limited pulse/filter pair achieves Shannon’s capacity (2.15). The “sinc” filters as well as the Root-Raised Cosine (RRC) filters can be such choices.

Theorem 2. *If $2WT \leq 1$, Shannon’s capacity can be achieved by any W -Hz band-limited transmit pulse and receive filter pair, provided that their Fourier transforms on $[-W, W]$ are zero only on a set of measure zero.*

Proof. Let $(1/\sqrt{T})g(t/T)$ and $(1/\sqrt{T})\hat{g}(t/T)$ to be W -Hz band-limited filters. No aliasing occurs in (4.5), (4.6) and (4.7), and for $-0.5 \leq f \leq 0.5$,

$$\begin{aligned} S(e^{j2\pi f}) &= G(f)\hat{G}^*(f) \operatorname{rect}\left(\frac{f}{1-\beta}\right), \\ S_V(e^{j2\pi f}) &= \frac{N_o}{2} \left|\hat{G}(f)\right|^2 \operatorname{rect}\left(\frac{f}{1-\beta}\right), \\ \frac{1}{T} \int_{-\frac{1-\beta}{2}}^{\frac{1-\beta}{2}} S_A(e^{j2\pi f}) |G(f)|^2 df &\leq P. \end{aligned}$$

If the subsets of $[-\frac{1-\beta}{2}, \frac{1-\beta}{2}]$ where $G(f)$ and $\hat{G}(f)$ are zero have zero measure, using $S_A(e^{j2\pi f}) = \frac{PT}{(1-\beta)|G(f)|^2}$ almost everywhere yields an objective function in (4.8) that equals the Shannon capacity (2.15). \square

4.3 Time-limited filters

I study now the achievable rates whenever the transmit pulse and receive filter are of finite durations.

4.3.1 The case where $2WT > 1$; Nyquist pulses

As in section 4.2.1, let $2WT = 1 + \beta$. When the pulse and receive filter are constrained to be of a finite duration, I prove that the problem is essentially “ill-posed” and the Nyquist criterion cannot be satisfied. More specifically, in the following theorem I show that the identity (4.9) is not possible.

Theorem 3. *For any finite non-negative β , if the transmit pulse and receive filter are time-limited and the channel is band-limited, then it is impossible to satisfy the Nyquist criterion.*

Proof. First note that for all $k \in \mathbb{Z}$, $G(f-k)$ and $\hat{G}^*(f-k)$ are analytic as they are the Fourier transforms of compactly supported functions which are in $\mathcal{L}^1(\mathbb{R})$ by the Cauchy-Schwarz inequality. Consequently, $\left\{G(f-k)\hat{G}^*(f-k)\right\}_k$ are analytic for all $k \in \mathbb{Z}$ and $\sum_{k \in \mathcal{A}} G(f-k)\hat{G}^*(f-k)$ is analytic for any finite set of integers \mathcal{A} . Since β is finite, one can find an open interval where $\sum_{k \in \mathbb{Z}} G(f-k)\hat{G}^*(f-k) \operatorname{rect}\left(\frac{f-k}{2WT}\right)$ contains only a finite number of aliased components indexed by \mathcal{A} . Consequently, $\sum_{k \in \mathcal{A}} G(f-k)\hat{G}^*(f-k)$ cannot be constant over any open interval for otherwise, by the identity theorem, it is constant everywhere and will not have non-zero finite energy. In conclusion, the sum cannot be equal to one for all f . \square

In light of the previous theorem, I consider in what follows time-limited versions of Nyquist pulses such as RRC filters. More precisely, I consider $\hat{g}(t) = g(t)$ to be a truncated (to c -seconds) RRC filters.

In Figure 4.2 I numerically evaluate the achievable rates and plot the degradation in performance of RRC filters for sample values of β and c . This degradation is measured as the percentage decrease from Shannon's formula, and hence the closer the curve is to zero the better. One can make the following observations:

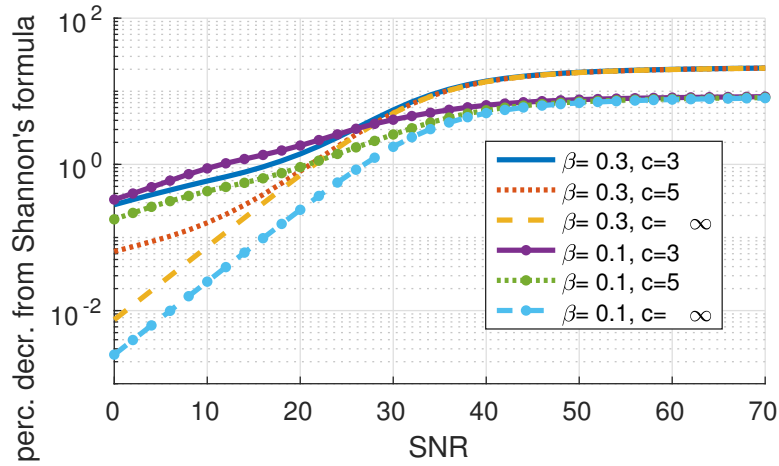


Figure 4.2: Performance of time-limited RRC and band-limited RRC

- For a fixed β note that as expected, as c increases the performance of the time-limited RRC improves toward that of the band-limited RRC (i.e. RRC with infinite time span) given by (4.10).

Additionally, the larger the value of β , the faster this convergence is.

- As the SNR increases the percentage decrease from Shannon's expression is larger, converging to $100 \frac{\beta}{1+\beta}$. Keeping in mind that the percentage decrease is the same for a given value of $\frac{\text{SNR}}{2W}$, one can conclude that the RRC pulse achieves a relatively good performance for low $\frac{\text{SNR}}{2W}$. As an example, for $W = 1$ MHz and at SNR = 30 dB, the percentage decrease for $c = 5$ and $\beta = 0.3$ is less than 0.065%.

4.3.2 The case where $2WT = 1$

Since $WT = (1/2)$, no aliasing occurs in (4.5) and (4.6), then

$$\begin{aligned} S(e^{j2\pi f}) &= G(f)\hat{G}^*(f) & f \in \left[-\frac{1}{2}, \frac{1}{2}\right] \\ S_V(e^{j2\pi f}) &= \frac{N_o}{2} |\hat{G}(f)|^2 & f \in \left[-\frac{1}{2}, \frac{1}{2}\right], \end{aligned}$$

and hence

$$\begin{aligned} R &= \max_{S_A(\cdot), g(\cdot)} \frac{1}{2T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left[1 + \frac{2}{N_o} S_A(e^{j2\pi f}) |G(f)|^2 \right] df \\ \text{subject to} & \quad \frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} S_A(e^{j2\pi f}) \sum_{k \in \mathbb{Z}} |G(f-k)|^2 df \leq P \\ & \quad g(t) \text{ time-limited to } \left[-\frac{c}{2}, \frac{c}{2}\right] \\ & \quad \int_{-\frac{c}{2}}^{\frac{c}{2}} |g(t)|^2 dt = 1, \end{aligned}$$

where I simplified the fraction in the “log” by removing $|\hat{G}(f)|^2$ from the numerator and the denominator. This is possible since $\hat{G}(f)$ is the Fourier transform of a time-limited function with non-zero finite energy (and hence it is in $\mathcal{L}^1(\mathbb{R})$), then it is analytic and the set of frequencies where $|\hat{G}(f)|^2$ is zero has zero measure, and it can be cancelled. Perhaps surprisingly, the information rates depend only on the transmit filter $g(t)$.

The next question is “what is the optimal finite-duration pulse $g(t)$ ”? Using calculus of variation leads to a non-linear Fredholm integral equation. While for $c \leq 1$ the integral equation is reasonable, for any other value of c the integral equation is rather intractable. In what follows I present the numerical approach I adopted to search for the optimal solution. The approach is based on using the PSWFs and their corresponding eigenvalues.

Numerical solution

Denote the duration of the pulse shaping filter by $T_s = cT$ and let $c_s = 2WT_s = 2WTc$. The PSWF⁴ parametrized by T_s and W is denoted $\varphi_{c_s, i}(t)$, $\lambda_{c_s, i}$ is its

⁴When it comes to the index of the PSWFs, I will use $c_s = 2WT_s$ throughout this chapter. Note that the parameter “ c_s ” here is different from the one used by Slepian and Pollak [19]. More specifically $\varphi_{1, i}(t)$ here is the same as $\varphi_{\frac{1}{2}, i}(t)$ in [19].

corresponding eigenvalue and $D\varphi_{c_s,i}(t)$ denotes the $[-\frac{T_s}{2}, \frac{T_s}{2}]$ time-limited PSWF $\varphi_{c_s,i}(t)$. Since the time-limited PSWFs form a complete orthogonal set for T_s seconds time-limited functions, then $g(t)$ and $b(t)$ can be uniquely written as a linear combination of time-limited PSWFs and band-limited PSWFs respectively: $g(t) = \sum_{i=0}^{\infty} \alpha_i \frac{\sqrt{T}}{\sqrt{\lambda_{c_s,i}}} D\varphi_{c_s,i}(tT)$ and $b(t) = \sum_{i=0}^{\infty} \alpha_i \sqrt{T\lambda_{c_s,i}} \varphi_{c_s,i}(tT)$, where $\frac{D\varphi_{c_s,i}(t)}{\sqrt{\lambda_{c_s,i}}}$ and $\varphi_{c_s,i}(t)$ are unit-norm functions and the constraint on the norm of $g(t)$ can be written as $\sum_{i=0}^{\infty} \alpha_i^2 = 1$.

I use the first $(N + 1)$ PSWFs in my computations as an approximate sub-optimal solution and I use the optimization toolbox in ‘‘MATLAB’’ to optimize for the values of $\{\alpha_i\}_i$'s.

It is worth noting that in [34, chapter 7] the author used the PSWFs in the context of FTN signaling. The author considered the PSWF with index zero only and studied the power-out-of-band and the minimum distance between the shifted pulses. In this work, I study the achievable rates and I search for the optimal combination of PSWFs which span time-limited pulses.

Figure 4.3 presents the results when using the first $(N + 1)$ PSWFs in the solution for $c = 3$. As expected, the obtained filters are even and differ for various levels of SNR.

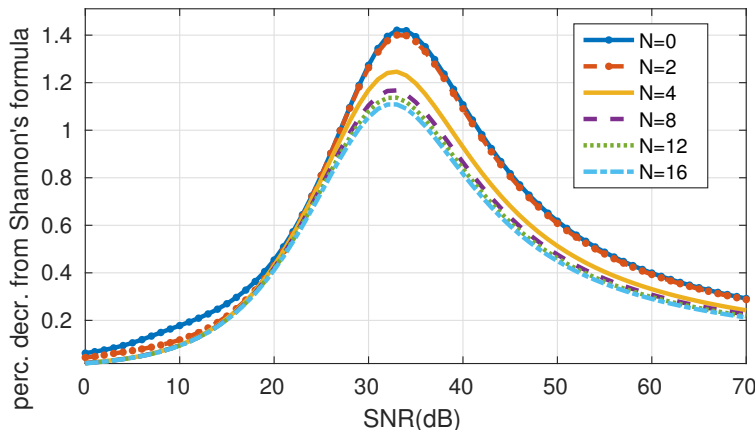


Figure 4.3: Performance of the obtained filters for $2WT=1$ and $c=3$

Two different ranges of SNR are noted, ones where the results behave differently: $SNR < 20$ dB and $SNR > 30$ dB. For $SNR < 20$ dB, numerical results show that using the first 5, 9, 13 or 17 PSWFs achieves almost the same performance, and one can conclude that this numerical solution is quasi-optimal over the considered SNR range.

On the other hand, in the range $SNR > 30$ dB using more PSWFs achieves tangible improvements despite the fact that the corresponding eigenvalues of

these added PSWFs are very small; for example $\lambda_{3,13} \approx 1.299 \times 10^{-17}$ and by adding $\{D\varphi_{3,13}(t), D\varphi_{3,14}(t), D\varphi_{3,15}(t), D\varphi_{3,16}(t)\}$ sizeable improvements are still achieved in the results. One may conclude that this numerical solution is further from optimality over this SNR range. I conjecture that the lower values of c and SNR, the closer are the approximations using the first $(N + 1)$ -PSWFs to optimality.

In Figure 4.4 I plot the percentage decrease of the time-limited RRC for $c = 5$ with two different values of β , namely 0.1 and 0.3. I also show on the same graph the percentage decrease of the numerically obtained optimal filter for $c = 5$ when using the first 17 PSWFs. Naturally, even with 17 PSWFs the optimized time-limited pulse outperforms the time-limited RRC over the whole SNR range. This agrees with the fact that the dimension of the $\mathcal{W}\text{-}\mathcal{T}$ space is approximately $2WT$ ([3, 15–17]), which indicates that one has at most $2WT$ independent symbols every \mathcal{T} seconds. In the case where $2WT = 1$ as here, one has at most one independent symbol every $\frac{1}{2W}$ seconds and by using Nyquist pulses with $2WT > 1$ one ends up under-using the available degrees of freedom.

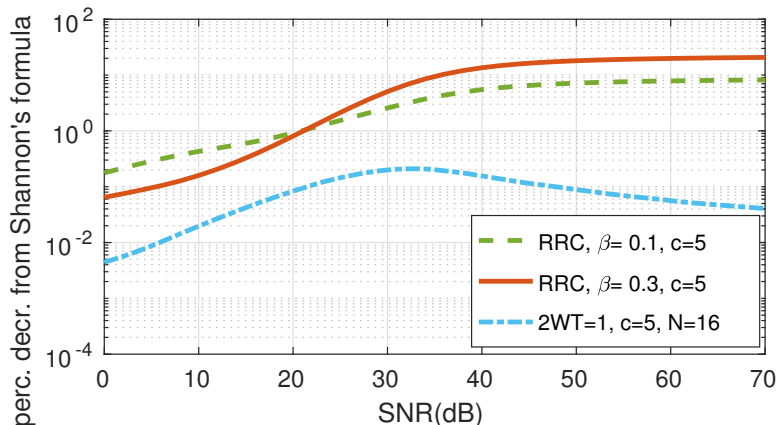


Figure 4.4: Time-limited RRC vs time-limited filters with $2WT=1$

4.3.3 The case where $2WT < 1$

Let $2WT = 1 - \beta$ as in section 4.2.3. Since no aliasing occurs in (4.5) and (4.6)

$$S(e^{j2\pi f}) = G(f)\hat{G}^*(f) \operatorname{rect}\left(\frac{f}{1-\beta}\right),$$

$$S_V(e^{j2\pi f}) = \frac{N_o}{2} |\hat{G}(f)|^2 \operatorname{rect}\left(\frac{f}{1-\beta}\right),$$

and the achievable rates are

$$\begin{aligned}
R &= \max_{S_A(\cdot), g(\cdot)} \frac{1}{2T} \int_{-\frac{1-\beta}{2}}^{\frac{1-\beta}{2}} \log \left[1 + \frac{2}{N_o} S_A(e^{j2\pi f}) |G(f)|^2 \right] df \\
\text{subject to} \quad & \frac{1}{T} \int_{-\frac{1}{2}}^{\frac{1}{2}} S_A(e^{j2\pi f}) \sum_{k \in \mathbb{Z}} |G(f-k)|^2 df \leq P \\
& g(t) \text{ time-limited to } \left[-\frac{c}{2}, \frac{c}{2} \right] \\
& \int_{-\frac{c}{2}}^{\frac{c}{2}} |g(t)|^2 dt = 1.
\end{aligned}$$

As in section 4.3.2 and by using the same proof, R does not depend on $\hat{g}(t)$. I summarize these findings in the following theorem.

Theorem 4. *If the pulses are time-limited and the channel is band-limited such that $2WT \leq 1$, then the achievable rates are independent of the time-limited receive filter.*

I applied the numerical solution proposed in section 4.3.2, with $c_s = 2WT_s = 2WTc = (1 - \beta)c$. I considered the values $\beta = 0.2$ and $c = 5$ and used the PSWFs $\{\varphi_{4,i}(t)\}_i$. In Figure 4.5, I present the results when using the first $(N+1)$ PSWFs in the numerical solution. Surprisingly, the obtained filters when using the first 9 PSWFs almost achieve the Shannon capacity as the losses are less than $3.1 \times 10^{-5}\%$ over the considered SNR range. This comes at the expense of the system operating at a faster rate than the $2WT = 1$ case. For example in Figure 4.5, the receiver generates 25% more samples than case B. Also note that the obtained filters are even and vary with the SNR level (as in section 4.3.2).

It is apparent that signaling at a rate faster than the Nyquist rate $(\frac{1}{2W})$ is beneficial, and this is due to the additional diversity which allows better control of the continuous time codewords. However this improvement in the achievable rates requires more complicated transmitter and receiver designs.

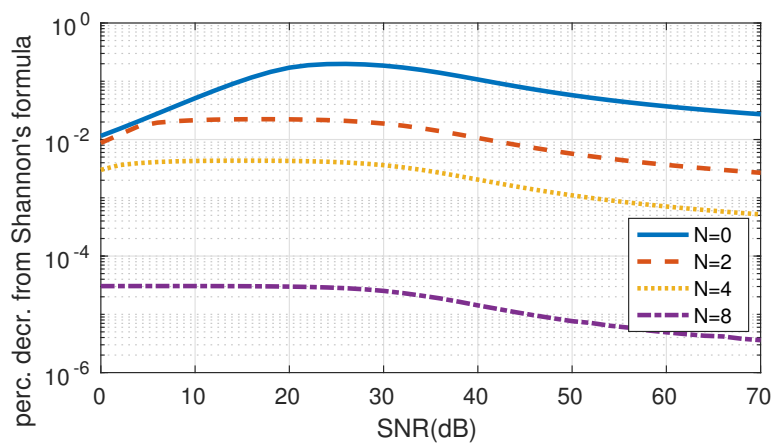


Figure 4.5: Performance of the obtained filters for $2WT=0.8$ and $c=5$

Chapter 5

Finite duration codewords over band-limited Gaussian channel

In this chapter, I consider transmitting CT finite duration codewords over a band-limited Gaussian channel. My main goal is to investigate the degrees of freedom and the achievable pairs of data rates and probability of error. I use a similar approach to Wyner [5, 6] and Gallager [7] (using the PSWFs) to transform the problem from CT to DT and vice versa, and then I apply the (adapted) results by Polyanskiy [11] for parallel DT AWGN channels.

5.1 System Model and Problem Formulation

I consider a system model where a T -seconds time-limited codeword is transmitted over a linear Gaussian channel with transfer function $H(f)$ —assumed to be an ideal W -Hz low-pass filter, and an additive complex Gaussian noise $N(t)$, assumed to be a stationary W -Hz band-limited “white” process with mean zero and PSD $S_N(f) = N_o$ for $f \in [-W, W]$. With a system in mind whereby other codewords may be transmitted—possibly by other users—consecutively and/or in neighboring bands, I denote by $C_{0,0}(t)$ the codeword carrying the data packet of interest, and by $\{C_{k,h}(t)\}_{(k,h) \in \mathbb{Z}^2 \setminus \{(0,0)\}}$ those carrying other data packets, possibly transmitted by other devices and interfering with the message of interest as illustrated in Figure 5.1. My model is based on the reasonable assumption that all codewords follow the same modulation techniques, since every frequency band is usually allocated to a unique technology which abides by specific standards, and the neighboring bands are more likely to be used by the same technology.

In what follows, I consider various scenarios where some or all of those interfering codewords are present and I denote by $\mathcal{I} \subset \mathbb{Z}^2 \setminus \{(0,0)\}$ the set of other present codewords. The overall signal going through the channel can hence be written as the sum of the codeword of interest and the other interfering codewords:

$$x(t) = C_{0,0}(t) + \sum_{(k,h) \in \mathcal{I}} C_{k,h}(t),$$

where $C_{k,h}(t)$ is non-zero only over $t \in [-T/2 + kT, T/2 + kT]$. On the receiver side, the data packet of interest is to be recovered from $y(t)$, $t \in [-T/2, T/2]$, a T -seconds time-limited version of the output of the channel $r(t)$.

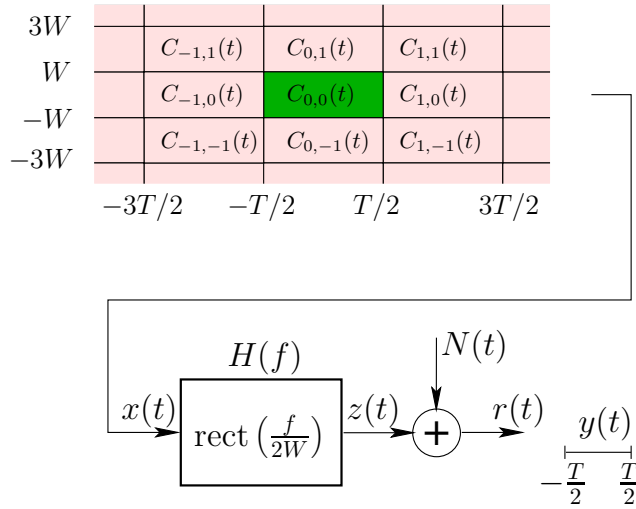


Figure 5.1: Continuous time system model.

I assume that, whenever present, a transmitted codeword satisfies the power constraint,

$$\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |C_{k,h}(t + kT)|^2 dt \leq P. \quad (5.1)$$

As Gallager proved that the PSWFs are the desirable CON set when sending a time-limited codeword over a band-limited channel [7, Section 8.4], I use the normalized time-limited PSWF as orthogonal pulses to send the data symbols. Hence the codewords can be written as

$$C_{k,h}(t) = \sum_{l=0}^{\infty} a_{k,h,l} \frac{D\varphi_{c,l}(t - kT)}{\sqrt{\lambda_{c,l}}} e^{j2\pi h 2W(t - kT)}$$

which are non-zero only on $t \in [-\frac{T}{2} + kT, \frac{T}{2} + kT]$, where $c = 2WT$. Representing the continuous time signal $C_{k,h}(t)$ by the symbols $\{a_{k,h,l}\}$ is known as “signal space representation” in the context of digital communications. The $\{a_{k,h,l}\}$ ’s

are chosen from a given complex signal constellation and by Plancherel and (5.1) they satisfy

$$\frac{1}{T} \sum_{l=0}^{\infty} |a_{k,h,l}|^2 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |C_{k,h}(t + kT)|^2 dt \leq P. \quad (5.2)$$

When it comes to the noise, $N(t)$ is band-limited and can be hence decomposed as

$$N(t) = \sum_{l=0}^{\infty} n_l \varphi_{c,l}(t),$$

where $\{n_l\}_{l \in \mathbb{Z}}$ are independent zero-mean complex circular Gaussian random variables with variance N_0 . Note that the use of PSWFs is necessary here to get independent random variables.

At the receiver, sufficient statistics are clearly obtained by projecting $y(t)$ on the set of normalized time-limited PSWF to extract the data symbols. It is worth noting that since $r(t)$ is band-limited and has finite energy, it is necessarily analytic and it is therefore sufficient to know $r(t)$ over any open interval to determine it fully. As a consequence, from an information-theoretic perspective, whether $r(t)$ as whole is available or only $y(t)$, the information rates are identical.

The problem at hand is to maximize the information rates given a maximum probability of error. This is naturally related to the available degrees of freedom when sending time-limited codewords over a band-limited channel, which is the maximum number of independent data symbols that can be transmitted to the receiver.

In the following section, I consider various scenarios and derive upper and lower bounds for the data rates and the degrees of freedom.

5.2 Bounds on the Data Rates

5.2.1 An Upper Bound

To derive an upper bound, I consider the case where only $C_{0,0}(t)$ is transmitted over the channel. By ignoring the other transmitted codewords I ignore the effect of inter-codeword interference, and I obtain upper bounds on the rates and the degrees of freedom since interference can only be harmful. In this scenario, the input to the channel can be written as

$$x(t) = \sum_{m=0}^{\infty} a_{0,0,m} \frac{D\varphi_{c,m}(t)}{\sqrt{\lambda_{c,m}}},$$

and the received signal $r(t)$ is band-limited and can be written as

$$r(t) = \sum_{m=0}^{\infty} a_{0,0,m} \sqrt{\lambda_{c,m}} \varphi_{c,m}(t) + N(t) = \sum_{m=0}^{\infty} \left[a_{0,0,m} \sqrt{\lambda_{c,m}} + n_m \right] \varphi_{c,m}(t)$$

$$\Rightarrow y(t) = r(t) \operatorname{rect}\left(\frac{t}{T}\right) = \sum_{m=0}^{\infty} \left[a_{0,0,m} \sqrt{\lambda_{c,m}} + n_m \right] D\varphi_{c,m}(t) = \sum_{m=0}^{\infty} y_m D\varphi_{c,m}(t),$$

where $y_m = \sqrt{\lambda_{c,m}} a_{0,0,m} + n_m$. From an information theoretic perspective, the considered system model is equivalent to the discrete time system model in Figure 5.2 where $r'_m = y_m / \sqrt{\lambda_{c,m}} = a_{0,0,m} + n_m / \sqrt{\lambda_{c,m}}$.

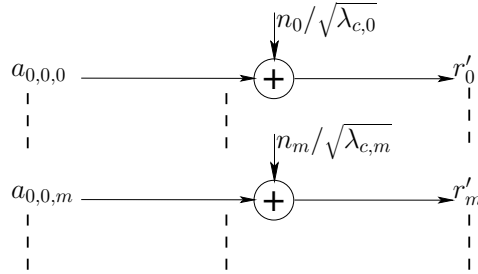


Figure 5.2: Equivalent discrete time system model.

The noise components $\{n_m / \sqrt{\lambda_{c,m}}\}$ are independent, zero-mean complex circular Gaussian random variables and each complex channel is equivalent to two usages of independent real (real and imaginary) channels with additive Gaussian noise with variance $N_0/2\lambda_{c,m}$ per dimension. If $\mathbf{E}[|a_{0,0,m}|^2] = 2P_m T$ (where $P_m T$ is the second moment per dimension), the power constraint in (5.2) can be written as $2 \sum_{m=0}^{\infty} P_m \leq P$ and the signal to noise ratio per dimension for r'_m is $\frac{2\lambda_{c,m} P_m T}{N_0}$.

One can notice that it is possible to send infinitely many independent symbols in such a system. However, only a finite number of them, say L , is useful because the energy per symbol is finite and the noise energy of the m^{th} channel is increasing towards infinity as m increases to infinity ($\lambda_{c,m}$ tends to 0 as m tends to ∞ as shown in Section 2.1). Note that L depends on $c = 2WT$ since there are approximately c eigenvalues $\lambda_{c,m}$ that are close to 1, and naturally L grows to infinity with c .

Polyanskiy [10] derived upper and lower bounds and an approximation for the achievable rates at a given probability of error in the finite block-length regime. In [11], he studied the parallel Gaussian channel set-up where N memoryless parallel channels of different noise power are used each n times. Following the methodology in [10] and applying the Berry–Esseen inequality [10] Lemma 14

over $n \times N$ real independent variables shows that the maximum number of bits that can be transmitted is

$$\frac{n}{2} \sum_{m=0}^{N-1} \log_2 \left[1 + \frac{2\lambda_{c,m} P_m T}{N_0} \right] - \sqrt{n \sum_{m=0}^{N-1} V_1 \left[\frac{2\lambda_{c,m} P_m T}{N_0} \right]} Q^{-1}(\epsilon) + O(\log_2(nL)) \quad \text{bits} \quad (5.3)$$

where

- $\epsilon \in (0, 1)$ is the probability of error,
- $V_1[\theta] = \frac{\theta}{2} \frac{\theta+2}{(\theta+1)^2} \log_2^2 e$,
- $\{P_m\}$ is the water-filling solution such that $P_m = \left[\mu - \frac{N_0}{2\lambda_{c,m} T} \right]^+$ and $n \sum_{m=0}^{N-1} P_m = P$,
- and L is the number of non-zero $\{P_m\}$'s, which is less or equal N .

It is worth noting here that the asymptotic expansion in (5.3) was first derived by Strassen in 1962 for discrete memoryless channels [9, Theorem 1.2].

In Polyanskiy's work [11] N is constant and n grows towards infinity and hence [11, Theorem 4] shows an $O(\log_2 n)$ term instead of the $O(\log_2 nL)$ term here in (5.3). In my scenario, each channel m in Figure 5.2 has a different noise power, but each has two dimensions with the same noise power. In my scenario therefore $n = 2$, L is of the order of (and grows with) c and the term $O(\log_2 nL)$ becomes $O(\log_2 L)$ as n is constant and equal to 2. These derivations yield the following lemma:

Lemma 1. *An upper bound on the data rates is given by,*

$$\begin{aligned} R_{UB}(\epsilon, P, c) &= \frac{1}{T} \sum_{m=0}^{\infty} \log_2 \left[1 + \frac{2\lambda_{c,m} P_m T}{N_0} \right] - \frac{1}{T} \sqrt{2 \sum_{m=0}^{\infty} V_1 \left[\frac{2\lambda_{c,m} P_m T}{N_0} \right]} Q^{-1}(\epsilon) \\ &\quad + \frac{1}{T} O(\log_2(L)) \quad \text{b/s.} \\ &= \frac{1}{T} \sum_{m=0}^{L-1} \log_2 \left[1 + \frac{2\lambda_{c,m} P_m T}{N_0} \right] - \frac{1}{T} \sqrt{2 \sum_{m=0}^{L-1} V_1 \left[\frac{2\lambda_{c,m} P_m T}{N_0} \right]} Q^{-1}(\epsilon) \\ &\quad + \frac{1}{T} O(\log_2(L)) \quad \text{b/s.} \end{aligned} \quad (5.4)$$

where

- $\epsilon \in (0, 1)$ is the probability of error,

- $V_1[\theta] = \frac{\theta}{2} \frac{\theta+2}{(\theta+1)^2} \log_2^2 e$,
- $\{P_m\}$ is the water-filling solution such that $P_m = \left[\mu - \frac{N_0}{2\lambda_{c,m}T} \right]^+$ and $2 \sum_{m=0}^{\infty} P_m =$
 $2 \sum_{m=0}^{L-1} P_m = P$,
- and L is the number of non-zero $\{P_m\}$ s.

5.2.2 Lower Bounds

In what follows, I derive a lower bound on the rates by jointly:

- Finding an upper bound on the interference.
- Using only the first N PSWFs to transmit data.

Subsequently, I optimize over the value of N to obtain tighter bounds as well as lower bounds on the degrees of freedom.

I study below three scenarios for the interference and derive a lower bound for these scenarios.

Consecutive Single Band Codewords (CSB)

I consider first the case where a single band is used and only the codewords $\{C_{k,0}(t)\}_{k \in \mathbb{Z}}$ are transmitted over the channel, i.e., $\mathcal{I} = \{(k, 0), k \in \mathbb{Z}^* \triangleq \mathbb{Z} \setminus \{0\}\}$ as shown in Figure 5.3.

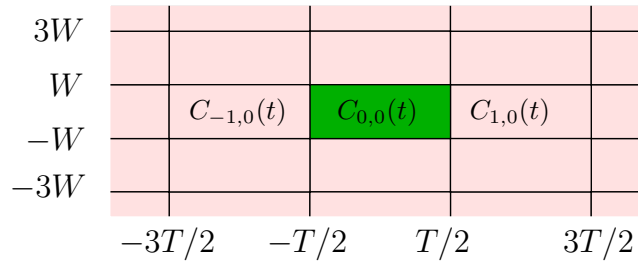


Figure 5.3: Single band interference.

In the case where only the first N PSWFs are used,

$$x(t) = C_{(0,0)}(t) + \sum_{k \in \mathbb{Z}^*} C_{k,0}(t) = \sum_{l=0}^{N-1} a_{0,0,l} \frac{D\varphi_{c,l}(t)}{\sqrt{\lambda_{c,l}}} + \sum_{k \in \mathbb{Z}^*} \sum_{l=0}^{N-1} a_{k,0,l} \frac{D\varphi_{c,l}(t - kT)}{\sqrt{\lambda_{c,l}}},$$

and the received signal is the sum of an information bearing signal, an interfering signal and additive channel noise,

$$r(t) = \left[\sum_{l=0}^{N-1} a_{0,0,l} \sqrt{\lambda_{c,l}} \varphi_{c,l}(t) \right] + \left[\sum_{k \in \mathbb{Z}^*} \sum_{l=0}^{N-1} a_{k,0,l} \sqrt{\lambda_{c,l}} \varphi_{c,l}(t - kT) \right] + N(t).$$

Projecting $r(t)$ on the m^{th} normalized time-limited PSWF—which can be done using an appropriate matched filter—results in

$$y_m = \lambda_{c,m} a_{0,0,m} + w_m,$$

where the interference plus noise term w_m is

$$w_m = \sum_{k \in \mathbb{Z}^*} \sum_{l=0}^{N-1} k_{,0} \alpha_{c,l,m} a_{k,0,l} + \sqrt{\lambda_{c,m}} n_m,$$

where $k_{,h} \alpha_{c,l,m}$ is defined in Equation (2.14). To be able to apply Polyanskiy's theorem for this lower bound and the following lower bounds, the received symbols $\{y_m\}$ should be independent Gaussian variables, and $\{w_m\}$ should also be independent Gaussian variables. Note that I can treat $\{w_m\}$ as independent here because I am deriving a lower bound (being dependent will help to increase the possible data rates).

Next, I upper bound its second moment,

$$\begin{aligned} \mathbb{E} [|w_m|^2] &= \mathbb{E} \left[\left| \sum_{k \in \mathbb{Z}^*} \sum_{l=0}^{N-1} k_{,0} \alpha_{c,l,m} a_{k,0,l} \right|^2 \right] + \lambda_{c,m} N_0 \\ &= \sum_{k \in \mathbb{Z}^*} \mathbb{E} \left[\left| \sum_{l=0}^{N-1} k_{,0} \alpha_{c,l,m} a_{k,0,l} \right|^2 \right] + \lambda_{c,m} N_0. \end{aligned}$$

Since the $\{a_{k,0,l}\}_{l=0}^{N-1}$ are not necessarily uncorrelated for a fixed k , I use the upper bound

$$\left| \sum_{l=1}^N c_l \right|^2 \leq N \sum_{l=1}^N |c_l|^2, \quad (5.5)$$

and the fact that $\mathbb{E} [|a_{k,0,l}|^2] = 2P_l T$ to upper bound the second moment

$$\begin{aligned} \mathbb{E} [|w_m|^2] &\leq \sum_{k \in \mathbb{Z}^*} N \sum_{l=0}^{N-1} |k_{,0} \alpha_{c,l,m}|^2 \mathbb{E} [|a_{k,0,l}|^2] + \lambda_{c,m} N_0 \\ &= N \sum_{l=0}^{N-1} 2P_l T \sum_{k \in \mathbb{Z}^*} |k_{,0} \alpha_{c,l,m}|^2 + \lambda_{c,m} N_0. \end{aligned}$$

Using the bound (A.11) derived in Appendix A

$$\sum_{k \in \mathbb{Z}^*} |_{k,0} \alpha_{c,l,m}|^2 \leq \lambda_{c,l}(1 - \lambda_{c,l}) \implies \mathbf{E} [|w_m|^2] \leq N \sum_{l=0}^{N-1} 2P_l T \lambda_{c,l}(1 - \lambda_{c,l}) + \lambda_{c,m} N_0.$$

Using the alternative bound (A.7) and the fact that $\sum_{l=0}^{N-1} 2P_l T \leq PT$,

$$\sum_{k \in \mathbb{Z}^*} |_{k,0} \alpha_{c,l,m}|^2 \leq \lambda_{c,m}(1 - \lambda_{c,m}) \implies \mathbf{E} [|w_m|^2] \leq \lambda_{c,m} [N(1 - \lambda_{c,m})PT + N_0].$$

Therefore the second moment of the interference term is upper-bounded by

$$I_{CSB}[m] \triangleq \min \left(N \sum_{l=0}^{N-1} 2P_l T \lambda_{c,l}(1 - \lambda_{c,l}), \lambda_{c,m} N(1 - \lambda_{c,m})PT \right), \quad (5.6)$$

and the signal to noise and interference ratio per dimension in y_m is lower-bounded by

$$S_{CSB}[m] \triangleq \frac{\lambda_{c,m}^2 2P_m T}{I_{CSB}[m] + \lambda_{c,m} N_0}.$$

These derivations yield the following lemma:

Lemma 2. A lower bound on the data rates is given by

$$R_{CSB}(\epsilon, P, c) = \max_{N, \{P_m\}_m} \frac{1}{T} \left[\sum_{m=0}^{N-1} \log_2 [1 + S_{CSB}[m]] - \sqrt{2 \sum_{m=0}^{N-1} V_1[S_{CSB}[m]] Q^{-1}(\epsilon) + O(\log_2(L))} \right], \quad (5.7)$$

where L is the number of non-zero $\{P_m\}$ s.

Single Time-Slot Multi-Band Codewords (STMB)

In this scenario multiple bands in a single time-slot are used and only the codewords $\{C_{0,h}(t)\}_{h \in \mathbb{Z}}$ are transmitted over the channel, i.e., $\mathcal{I} = \{(0, h), h \in \mathbb{Z}^*\}$ as shown in Figure 5.4 below.

As above, the output of the m^{th} matched filter is

$$y_m = \lambda_{c,m} a_{0,0,m} + w_m,$$

where the interference plus noise term is now

$$w_m = \sum_{h \in \mathbb{Z}^*} \sum_{l=0}^{N-1} \alpha_{c,l,m} a_{0,h,l} + \sqrt{\lambda_{c,m}} n_m.$$

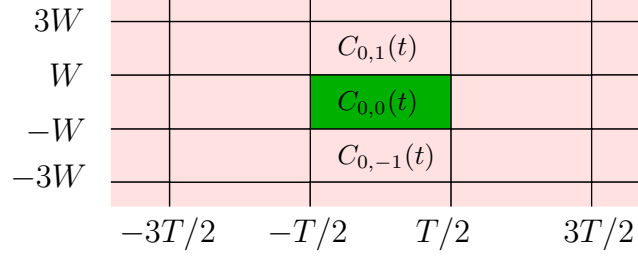


Figure 5.4: Single time-slot interference.

Using the bound (A.12), one can upper bound the second moment of w_m ;

$$\sum_{h \in \mathbb{Z}^*} |{}_{0,h} \alpha_{c,l,m}|^2 \leq \frac{\lambda_{c,m}^2 (1 - \lambda_{c,m})}{\lambda_{c,l}},$$

then

$$\begin{aligned} \mathbb{E} [|w_m|^2] &= \mathbb{E} \left[\left| \sum_{h \in \mathbb{Z}^*} \sum_{l=0}^{N-1} {}_{0,h} \alpha_{c,l,m} a_{0,h,l} \right|^2 \right] + \lambda_{c,m} N_0 \\ &\leq \lambda_{c,m}^2 (1 - \lambda_{c,m}) N \sum_{l=0}^{N-1} \frac{2P_l T}{\lambda_{c,l}} + \lambda_{c,m} N_0. \end{aligned}$$

Alternatively, using the bound (A.8)

$$\sum_{h \in \mathbb{Z}^*} |{}_{0,h} \alpha_{c,l,m}|^2 \leq \lambda_{c,m} (1 - \lambda_{c,l}),$$

the second moment of w_m can be upper bounded by

$$\begin{aligned} \mathbb{E} [|w_m|^2] &= \mathbb{E} \left[\left| \sum_{h \in \mathbb{Z}^*} \sum_{l=0}^{N-1} {}_{0,h} \alpha_{c,l,m} a_{0,h,l} \right|^2 \right] + \lambda_{c,m} N_0 \\ &\leq \lambda_{c,m} N \sum_{l=0}^{N-1} (1 - \lambda_{c,l}) 2P_l T + \lambda_{c,m} N_0. \end{aligned}$$

Therefore the second moment of the interference term is upper-bounded by

$$I_{STMB}[m] \triangleq \min \left(\lambda_{c,m}^2 (1 - \lambda_{c,m}) N \sum_{l=0}^{N-1} \frac{2P_l T}{\lambda_{c,l}}, \quad \lambda_{c,m} N \sum_{l=0}^{N-1} (1 - \lambda_{c,l}) 2P_l T \right). \quad (5.8)$$

These derivations yield the following lemma:

Lemma 3. *The corresponding lower bound on the data rates is given by*

$$R_{STMB}(\epsilon, P, c) = \max_{N, \{P_m\}_m} \frac{1}{T} \left[\sum_{m=0}^{N-1} \log_2[1 + S_{STMB}[m]] - \sqrt{2 \sum_{m=0}^{N-1} V_1[S_{STMB}[m]] Q^{-1}(\epsilon) + O(\log_2(L))} \right], \quad (5.9)$$

where L is the number of non-zero $\{P_m\}$'s and

$$S_{STMB}[m] = \frac{\lambda_{c,m}^2 2P_m T}{I_{STMB}[m] + \lambda_{c,m} N_0}.$$

Consecutive Multi-Band codewords (CMB)

I consider now the case where all the codewords $\{C_{k,h}\}_{(k,h) \in \mathbb{Z}^2}$ are transmitted over the channel. The analysis follows as above and the interference plus noise term w_m is

$$\begin{aligned} w_m &= \sum_{(k,h) \in \mathbb{Z}^2 \setminus (0,0)} \sum_{l=0}^{N-1} k,h \alpha_{c,l,m} a_{k,h,l} + \sqrt{\lambda_{c,m}} n_m \\ &= \left[\sum_{k \in \mathbb{Z}^*} \sum_{l=0}^{N-1} k,0 \alpha_{c,l,m} a_{k,0,l} \right] + \left[\sum_{h \in \mathbb{Z}^*} \sum_{l=0}^{N-1} 0,h \alpha_{c,l,m} a_{0,h,l} \right] \\ &\quad + \left[\sum_{(k,h) \in \mathbb{Z}^* \times \mathbb{Z}^*} \sum_{l=0}^{N-1} k,h \alpha_{c,l,m} a_{k,h,l} \right] + \sqrt{\lambda_{c,m}} n_m. \end{aligned}$$

Upper bounds on the second moments of the first two interference terms have been derived in Sections 5.2.2 and 5.2.2, respectively, and it remains to derive one for the third term. By Equation (5.5),

$$\begin{aligned} \mathbb{E} \left[\left| \sum_{k \in \mathbb{Z}^*} \sum_{h \in \mathbb{Z}^*} \sum_{l=0}^{N-1} k,h \alpha_{c,l,m} a_{k,h,l} \right|^2 \right] &\leq N \sum_{l=0}^{N-1} \mathbb{E} \left[\left| \sum_{h \in \mathbb{Z}^*} \sum_{k \in \mathbb{Z}^*} a_{k,h,l} k,h \alpha_{c,l,m} \right|^2 \right] \\ &= N \sum_{l=0}^{N-1} \sum_{h \in \mathbb{Z}^*} \sum_{k \in \mathbb{Z}^*} \mathbb{E} [|a_{k,h,l}|^2] |k,h \alpha_{c,l,m}|^2 \leq N \sum_{l=0}^{N-1} 2P_l T \sum_{k \in \mathbb{Z}^*} \sum_{h \in \mathbb{Z}^*} |k,h \alpha_{c,l,m}|^2. \end{aligned}$$

Using bound (A.13) and Equations (5.6) and (5.8),

$$\mathbb{E} [|w_m|^2] \leq I_{CSB}[m] + I_{STMB}[m] + N \sum_{l=0}^{N-1} (1 - \lambda_{c,l}) 2P_l T + \lambda_{c,m} N_0$$

Alternatively, using bound (A.14),

$$\mathbb{E} [|w_m|^2] \leq I_{CSB}[m] + I_{STMB}[m] + N \sum_{l=0}^{N-1} \frac{\lambda_{c,m}}{\lambda_{c,l}} (1 - \lambda_{c,m}) 2P_l T + \lambda_{c,m} N_0,$$

Therefore, the second moment on the interference terms is upper-bounded by

$$I_{CMB}[m] \triangleq I_{CSB}[m] + I_{STMB}[m] + \min \left(N \sum_{l=0}^{N-1} (1 - \lambda_{c,l}) 2P_l T, \quad N \sum_{l=0}^{N-1} \frac{\lambda_{c,m}}{\lambda_{c,l}} (1 - \lambda_{c,m}) 2P_l T \right).$$

These derivations yield the following lemma:

Lemma 4. *The corresponding lower bound on the data rates is given by*

$$R_{LB}(\epsilon, P, c) = \max_{N, \{P_m\}_m} \frac{1}{T} \left[\sum_{m=0}^{N-1} \log_2 [1 + S_{CMB}[m]] - \sqrt{2 \sum_{m=0}^{N-1} V_1[S_{CMB}[m]] Q^{-1}(\epsilon) + O(\log_2(L))} \right], \quad (5.10)$$

where L is the number of non-zero $\{P_m\}$ s and

$$S_{CMB}[m] \triangleq \frac{\lambda_{c,m}^2 2P_m T}{I_{CMB}[m] + \lambda_{c,m} N_0}.$$

5.3 Numerical Results

In what follows, I use $W = 1$ KHz and $\epsilon = 10^{-3}$ in my computations. Since I consider the complex base-band channel, an equivalent real pass-band channel will have $2W$ -Hz bandwidth (i.e., 2 KHz in my case), and L complex degrees of freedom in base-band is equivalent to $2L$ real degrees of freedom in passband.

In the following I evaluate the degrees of freedom and the data rates for different SNR levels (in dB) as L depends on the SNR; for example, since the water-filling algorithm is used for Equation (5.4), L can be increased by increasing the “water level” in the water-filling algorithm, which means that the SNR level must be increased. Note that for a fixed c , the ratio of the derived bounds to Shannon’s capacity (Equation (2.17)) depends only on P , N_o and W through the ratio $\frac{P}{N_o W}$, and the results for $W \neq 1$ KHz can be extracted from my presented results for an appropriate range of $\text{SNR} = 2P/N_o$.

To evaluate the bounds in Equations (5.4), (5.7), (5.9) and (5.10), I used the optimization toolbox in MATLAB to search for the optimal solution in $\{P_m\}_m$.

When it comes to the upper bound (5.4), the water-filling choice used by Polyanskiy only maximizes the term $\sum_{m=0}^{\infty} \log_2 \left[1 + \frac{2\lambda_{c,m} P_m T}{N_0} \right]$ and not the whole expression. However, the water-filling choice almost achieves the same performance as the optimization routine with negligible differences. When it comes to the term $O(\log_2(L))$, I specialize it to $\frac{1}{2} \log_2(2L)$ for the sake of numerical computations; I used the constant $\frac{1}{2}$ as Polyanskiy conjectured in Equation (4.218) in his thesis [10], and I used the term $2L$ inside $\log_2(\cdot)$ since the number of real independent variables is $2L$ (as explained in Section 5.2.1). Interestingly, the term $\frac{1}{2} \log_2(n)$ (where n is the block-length) was first conjectured by Dobrushin [8] in 1961 for some discrete symmetric channels.

5.3.1 Upper Bound

To solve Equation (5.4), the following quantity is maximized using the optimization tool in MATLAB

$$R_{UB}^N = \max_{P_m} \frac{1}{T} \sum_{m=0}^{N-1} \log_2 \left[1 + \frac{2\lambda_{c,m} P_m T}{N_0} \right] - \frac{1}{T} \sqrt{2 \sum_{m=0}^{N-1} V_1 \left[\frac{2\lambda_{c,m} P_m T}{N_0} \right]} Q^{-1}(\epsilon),$$

over $\{P_m\}_m^{N-1}$ for different values of N . The solution being decreasing with m , L is the value of N where R_{UB}^N saturates (which is the same as the number of non-zero P_m 's after saturation). I adopt similar method and notations for all the bounds presented hereafter.

Figure 5.5 shows the obtained R_{UB}^N for $c = 1000$ and at $\text{SNR} = 50$ dB, and in this example the obtained degrees of freedom are $L = 1004$. The upper bound on the rates is

$$R_{UB}(\epsilon, P, c) = R_{UB}^L + \frac{1}{2T} \log_2(2L) \quad \text{bits/sec.}$$

I compute L and the corresponding upper bound R_{UB} for different values of c and for different levels of SNR. Figure 5.6 shows the difference $(L - c)$ between the obtained degrees of freedom and $c = 2WT$, and as expected, L increases as the SNR increases (in a manner akin to the water-filling solution: as the water level increases, it is possible to use additional PSWFs). However, the additional degrees of freedom (beyond c) increase slowly with c and $L/c \rightarrow 1$ as c increases towards infinity.

In Figure 5.7 I plot the obtained upper bound on the rates together with the Shannon capacity. Notice that the gap between the upper bounds and the Shannon capacity decreases as c increases.

The ratio of the bounds to the Shannon capacity can be seen in Figure 5.9 below.

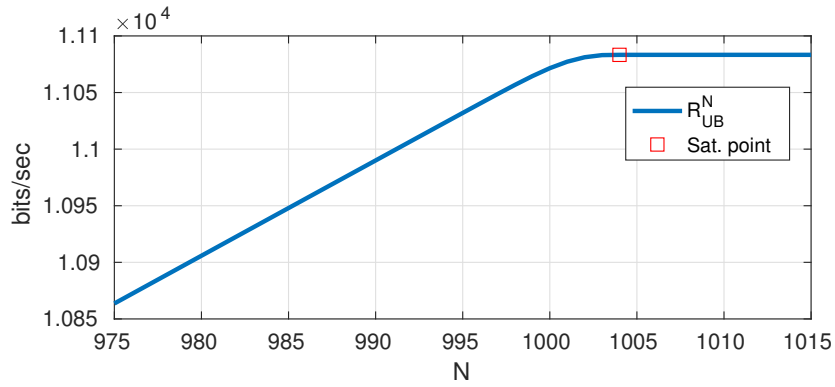


Figure 5.5: Saturation of R_{UB}^N for $c = 1000$ and $\text{SNR} = 50$ dB.

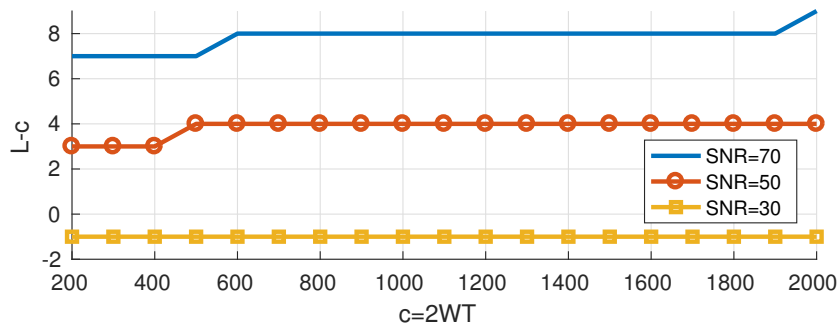


Figure 5.6: Upper bounds on the degrees of freedom.

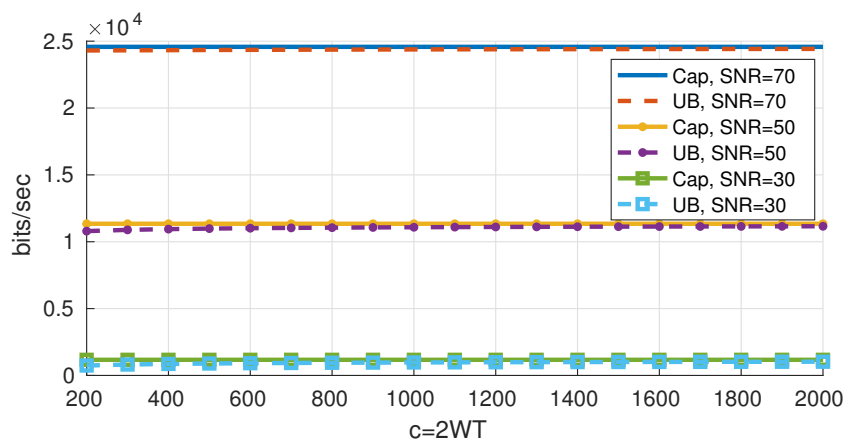


Figure 5.7: Upper bounds on the data rates.

5.3.2 Lower Bounds

In this section, I present the numerical results for Sections 5.2.2–5.2.2 (Equations (5.7), (5.9) and (5.10)), and I apply the same numerical method I used in the previous section. Note that I omitted the results for CSB since the lower bounds when using either CSB or STMB are almost the same with no significant differences.

Figure 5.8 shows the obtained lower bounds on the degrees of freedom. Note that for given scenario (CSB, STMB or the general lower bound), the results are almost the same for different SNR levels; increasing the signal power will only lead to increasing the power of the interference (see Equations (5.7), (5.9) and (5.10)), and the effect on the signal to interference ratio remains negligible. Moreover, the results for the different scenarios are very close (± 1 on average). Although $L - c$ decreases as c increases, it decreases slowly and $L/c \rightarrow 1$ as c increases towards infinity.

I propose approximating the degrees of freedom for the general lower bound by the following equation

$$L_{LB} \hat{=} c - 1.35 \log_2(c) + 3.25, \quad (5.11)$$

and I draw “ $L_{LB} - c$ ” in Figure 5.8. It is expected that “ $L_{LB} - c$ ” is a logarithmic function of c since the transition region of the eigenvalues of PSWFs is a logarithmic function of c (as shown in Section 2.1).

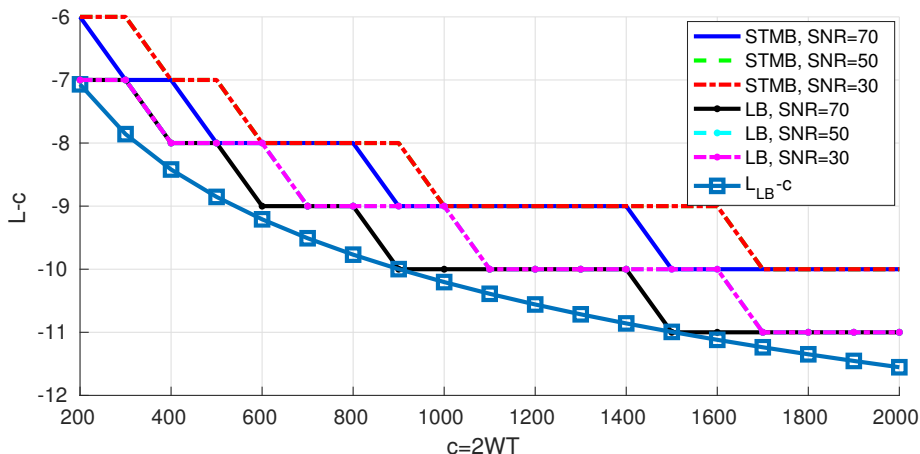


Figure 5.8: Lower bounds on the degrees of freedom.

Figure 5.9 shows the ratio of the obtained bounds to the Shannon capacity. For a fixed c and as SNR increases, the bounds get relatively closer to capacity and the gap between the upper bound and the lower bound increases. In addition, for a fixed SNR, as c increases, the gap between the bounds decreases.

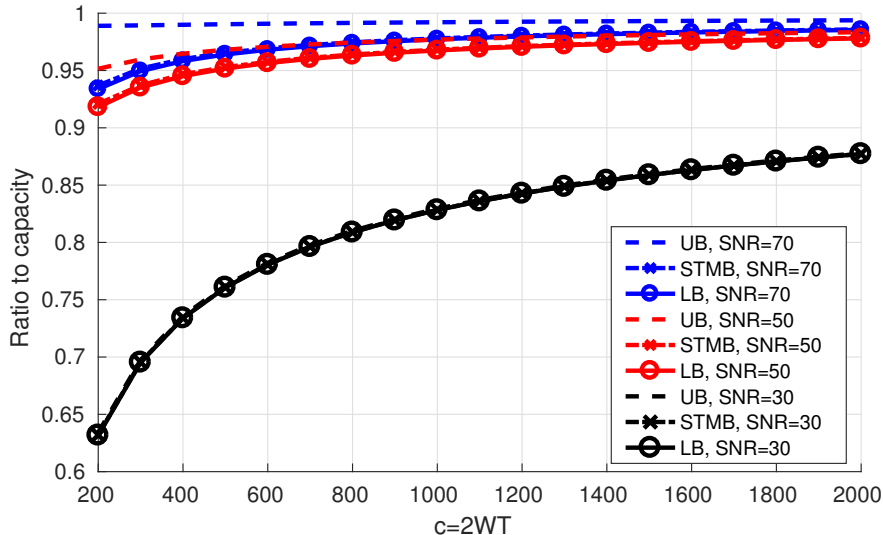


Figure 5.9: Upper and lower bounds on the rates.

5.4 Possible Enhancements on the Bounds

Since the obtained upper and lower bounds on the rates (Figure 5.9) are very close for $\text{SNR} = 30$ dB, the bounds are tight and a good approximation for the optimal data rates is reached. However, the gap between the upper and lower bounds increases as the SNR increases. For instance, the gap between the upper and lower bounds for $c = 200$ and $\text{SNR} = 70$ dB is 5.5% of Shannon's capacity, which means that one or both bounds are loose. In the following I present some possible improvements on the bounds.

5.4.1 A Tighter Upper Bound

To derive the upper bound, I ignored the interference due to other codewords and the obtained degrees of freedom surpassed $2WT$ for $\text{SNR} \in \{50 \text{ dB}, 70 \text{ dB}\}$. However, the results in the literature shows that the asymptotic dimension of the W-T space is $2WT$, and thus adding the constraint that the codewords must be time-limited will not increase the available degrees of freedom. So one can conclude that the obtained upper bound on the degrees of freedom is not tight. One possible way to improve this upper bound is to force the degrees of freedom to be at most $2WT$, and hence, L is constrained to be less or equal to $2WT$ in Equation (5.4). In other words, the normalized time-limited PSWFs $\frac{D\varphi_{c,m}(t)}{\sqrt{\lambda_{c,m}}}$ with $m \geq 2WT$ will not be used to transmit data and their allocated power P_m 's are forced to be zero. The obtained upper bound will decrease and

hence the gap between the upper and lower bounds will decrease.

In Figure 5.10 I present the numerical values for the tighter upper bound (TUB) in addition to those of the upper bound. When $c = 200$, note that the tighter upper bound achieves improvements of 0.8% and 1.5% for SNR = 50 dB and SNR = 70 dB, respectively.

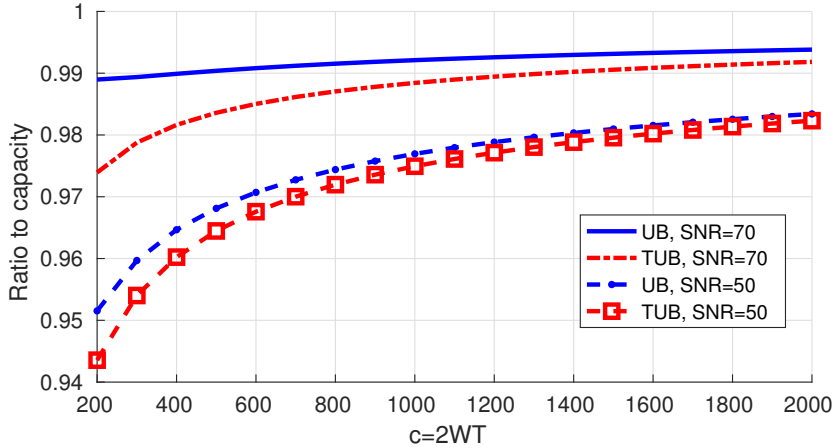


Figure 5.10: Tighter upper bound on the rates vs. the old ones.

5.4.2 Introducing a Guard Time or a Guard Band between the Codewords

The use of a guard time or a guard band is expected to decrease the effect of inter-codeword interference which drives towards increasing the achievable rates. On the other hand a portion of the time resources or frequency resources will not be used which decreases the data rates. Therefore, there must be an optimal guard time and an optimal guard band that maintain a good trade-off between the lost resources and the improvement of the inter-codeword interference.

Using a guard time T_G reduces the upper bound to $R_{UB}(T_G) = R_{UB} \frac{T}{T+T_G}$, but introducing a guard band does not affect it. When it comes to the lower bound, deriving a closed form expression has proven to be difficult.

Nevertheless, if adding a guard time is beneficial, the optimal guard time T_G^* must satisfy $R_{UB}(T_G^*) \geq R_{LB}$ where R_{UB} and R_{LB} are given by Equations (5.4) and (5.10) and hence $\frac{T_G^*}{T} \leq \frac{R_{UB}}{R_{LB}} - 1$. In Figure 5.11, I draw the obtained upper bound on T_G^* . Naturally, it increases as the SNR increases since the relative difference between the bounds increases as the SNR increases.

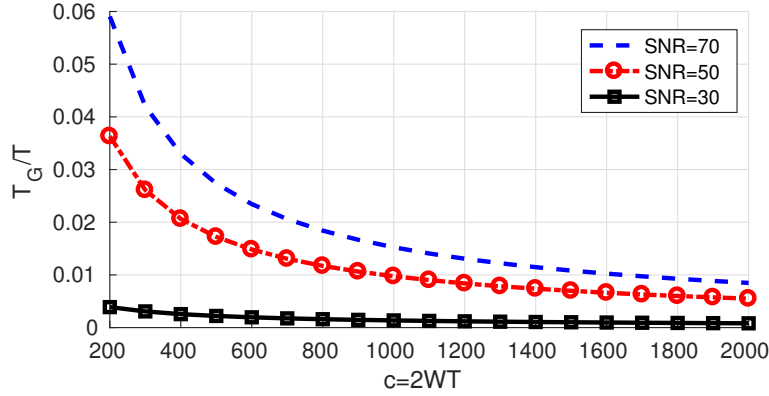


Figure 5.11: Upper bound on $\frac{T_G^*}{T}$ for different SNR levels.

5.5 Conclusion

The derived bounds on the achievable rates were found to be tight for low values of SNR/W (for example $W = 1$ KHz and $\text{SNR} = 30$ dB or equivalently $W = 100$ KHz and $\text{SNR} = 50$ dB). However, the gap between the bounds increases as this SNR/W ratio increases.

When it comes to the available degrees of freedom, the numerical results showed that the potential decrease from $2WT$ is upper bounded by a logarithmic function of $2WT$ and hence the relative reduction is asymptotically negligible.

Based on the results of this chapter, one can approximate the achievable data rates by

$$R(\epsilon, P, c) = \frac{L}{T} \log_2 \left[1 + \frac{PT}{N_0 L} \right] - \frac{\sqrt{2L}}{T} \sqrt{V_1 \left[\frac{PT}{N_0 L} \right]} Q^{-1}(\epsilon) + \frac{1}{2T} \log_2(2L),$$

where $V_1(\theta) = \frac{\theta}{2} \frac{\theta+2}{(\theta+1)^2} \log_2^2 e$ and L is in the range $c - 1.35 \log_2(c) + 3.25 \leq L \leq c$. This approximation is guaranteed to be between the derived bounds, and hence is a “good” one for low values of SNR/W .

Chapter 6

Lossy coding of a time-limited piece of a band-limited white Gaussian source

In this chapter I study the coding rates of a T -seconds finite duration piece from a W -Hz band-limited white Gaussian process with the \mathcal{L}^2 norm of the error as a distortion measure. When it comes to the discrete representation of the CT process, the short duration sub-functions are projected on the set of PSWFs, which produces an infinite number of independent but not identically distributed Gaussian random variables. I show that only a finite set of them can be used for coding, where I refer to the size of this set as the “effective” blocklength. I use similar tools to the ones used to develop rate distortion theory for independent and identically distributed Gaussian random variables in the finite blocklength regime in [13] and [14], and I derive an upper bound and a lower bound and I evaluate them numerically.

6.1 System model

Consider a real continuous-time source that produces a zero mean W -Hz band-limited Gaussian process $X(t)$ with a flat PSD that I label as “white”

$$S_X(f) = \begin{cases} \sigma^2 & f \in [-W, W] \\ 0 & \text{otherwise} \end{cases}, \quad (6.1)$$

and define -as shown in Figure 6.1, the sub-functions $\{X_k(t)\}_{k \in \mathbb{Z}}$ as

$$X_k(t) = \begin{cases} X(t) & t \in [kT - \frac{T}{2}, kT + \frac{T}{2}] \\ 0 & \text{otherwise} \end{cases}.$$

I study the problem of coding independently the sub-functions $\{X_k(t)\}_{k \in \mathbb{Z}}$ and without loss of generality, I focus on the signal $X_0(t)$. With practicality in

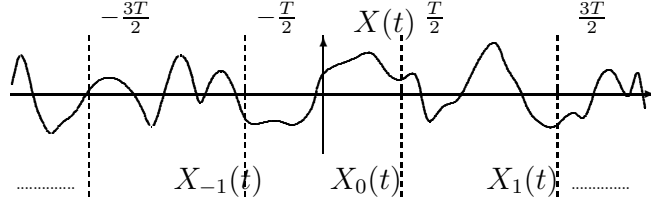


Figure 6.1: $X(t)$ is encoded by independently coding $\{X_k(t)\}_{k \in \mathbb{Z}}$.

mind, $X_0(t)$ must be represented by a finite number of bits, say R bits, and hence only $M = 2^R$ representation signals can be used. Let $\{Y_m(t)\}_{m=1}^M$ denote the set of an optimal representation. To encode $X_0(t)$, the nearest representation signal $Y_{\hat{m}}(t)$ is selected and the binary representation of the integer \hat{m} is used to encode $X_0(t)$. When it comes to the reconstruction of $X_0(t)$, the index \hat{m} is recovered from the generated bits and the reconstruction of $X_0(t)$ is the signal $Y_{\hat{m}}(t)$.

Being different from $X_0(t)$, the loss or distortion of the reconstructed signal is measured using the mean squared error, i.e. the \mathcal{L}^2 -norm of the difference between the original and the reconstructed signals. Since $X_0(t)$ is time-limited, the optimal representation signals are also T -seconds time-limited, and the distortion is defined as

$$d(X_0(t) - Y_{\hat{m}}(t)) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |X_0(t) - Y_{\hat{m}}(t)|^2 dt. \quad (6.2)$$

To evaluate the performance of the coding scheme, I adopt the same criterion described in [13], namely a guaranteed distortion level d with probability $(1 - \epsilon)$,

$$\mathbb{P}[d(X_0(t) - Y_{\hat{m}}(t)) > d] \leq \epsilon. \quad (6.3)$$

The problem at hand is to find the minimum possible rate R^* , for a given fixed d and ϵ . I tackle this problem by finding and solving an equivalent discrete model.

Being band-limited, by property 2.1.1 of the PSWFs, $X(t)$ can be written as a combination of PSWFs $X(t) = \sum_{i \in \mathbb{N}} \hat{X}_i \varphi_{c,i}(t)$, where $c = 2WT$ and $\{\hat{X}_i\}_{i \in \mathbb{N}}$ are independent and identically distributed Gaussian random variables with mean zero and variance σ^2 [7, section 8.4]. Consequently, $X_0(t)$ is a combination of normalized time-limited PSWFs:

$$X_0(t) = \sum_{i \in \mathbb{N}} \sqrt{\lambda_{c,i}} \hat{X}_i \frac{D\varphi_{c,i}(t)}{\sqrt{\lambda_{c,i}}} = \sum_{i \in \mathbb{N}} X_i \frac{D\varphi_{c,i}(t)}{\sqrt{\lambda_{c,i}}},$$

where $\{X_i\}_{i \in \mathbb{N}}$ are independent Gaussian random variables with mean zero and variances $\lambda_{c,i}\sigma^2$ respectively. Being time-limited, by property 2.1.3 the $\{Y_m(t)\}$'s are also linear combinations of these time-limited PSWFs $Y_m(t) = \sum_{i \in \mathbb{N}} Y_{m,i} \frac{D\varphi_{c,i}(t)}{\sqrt{\lambda_{c,i}}}$,

and $\{Y_m(t)\}_{m=1}^M$ can be designed by selecting the appropriate discrete signals $\mathbf{Y}_m = \{Y_{m,i}\}_{i \in \mathbb{N}}$ for $m = 1, \dots, M$.

The problem at hand now boils down to coding a discrete source that generates independent random variables $\mathbf{X} = \{X_i\}_{i \in \mathbb{N}}$ each Gaussian $X_i \sim \mathcal{N}(0, \lambda_{c,i}\sigma^2)$ through selecting the nearest representation signal \mathbf{Y}_m . By Parseval, the distortion (6.2) is equal to

$$d(\mathbf{X}, \mathbf{Y}_m) = \frac{1}{T} \sum_{i \in \mathbb{N}} |X_i - Y_{m,i}|^2.$$

In what follows, I adopt a methodology that is similar to the one used in [13]; I derive upper and lower bounds for the smallest possible R . I emphasize that the problem at hand is different than the one in [13] since here the smallest set with probability $(1 - \epsilon)$ is an –infinite dimensional– ellipsoid. Moreover, an infinite number of identical –infinite dimensional– spheres is required to cover this ellipsoid, as presented in the next section.

6.2 Bounds

Examining the expression of the Gaussian probability density function, one can see that the smallest-volume set \mathbf{E} such that the generated vector \mathbf{X} falls inside it with probability $(1 - \epsilon)$ is an –infinite dimensional– ellipsoid centered at the origin. Ellipsoid \mathbf{E} has a radius $r_i = a\sqrt{\lambda_{c,i}}\sigma$ in dimension i , where a is a constant determined by ϵ .

Now consider any set \mathbf{A} with probability $(1 - \epsilon)$ under $p_{\mathbf{X}}(\cdot)$. To guarantee that (6.3) is satisfied, for any vector $\mathbf{X} \in \mathbf{A}$ there must be at least one representation signal \mathbf{Y}_m such that

$$\sum_{i \in \mathbb{N}} |X_i - Y_{m,i}|^2 \leq dT. \quad (6.4)$$

The distortion being the usual $\mathcal{L}^2(\mathbb{R})$ distance (6.4), one can consider that each \mathbf{Y}_m covers a sphere \mathbf{B} with radius \sqrt{dT} . Therefore, $\{\mathbf{Y}_m\}_{m=1}^M$ and their corresponding spheres should cover \mathbf{A} . Denote by R^* the smallest number of bits needed to encode $X_0(t)$ while satisfying equation (6.3).

6.2.1 Converse – Lower Bound

For the sake of the technical proof, I define for every positive integer k the positive scalar a_k to be the solution to

$$\mathbb{P}[Z_k < a_k^2] = (1 - \epsilon), \quad (6.5)$$

where Z_k is chi-square distributed with k degrees of freedom. Note that a_k^2 is increasing and using the Berry-Esseen theorem [10, theorem 13], for some $|\beta| \leq \frac{12\sqrt{2}}{\sqrt{k}}$,

$$a_k^2 = k + \sqrt{2k} Q^{-1}(\epsilon + \beta) \leq k + \sqrt{2k} Q^{-1}(2\epsilon), \quad k \gg 1. \quad (6.6)$$

Theorem 5. *Under the guarantee that the distortion does not exceed d with probability $(1 - \epsilon)$, coding a T -seconds piece from a W -Hz band-limited white Gaussian process of spectrum σ^2 requires more than*

$$R_L = \max_{k \geq 1} \frac{1}{2} \sum_{i=0}^{k-1} \log_2 \left[\frac{\lambda_{c,i} a_k^2 \sigma^2}{dT} \right] \quad \text{bits,}$$

where the $\{a_k\}$'s are as in (6.5).

Proof. A lower bound on the number of bits needed to encode the subset $\mathbf{X}^k = \{\mathbf{X}_i\}_{i=0}^{k-1}$ while satisfying

$$\mathbb{P} \left[\sum_{i=0}^{k-1} |X_i - Y_{m,i}|^2 > dT \right] \leq \epsilon$$

is also a lower bound on R^* . Therefore I derive a lower bound for encoding this subset -denoted by R_L^k , and a lower bound $R_L = \max_{k \geq 1} R_L^k$ is obtained. The existence of the maximum is presented right after the proof, where I show that $\arg \max_k R_L^k$ is finite and is expected to be close to c .

Whenever $X_i \sim \mathcal{N}(0, \lambda_{c,i} \sigma^2)$, the shape \mathbf{E}_k that contains the smallest volume in which \mathbf{X}_k falls with probability $(1 - \epsilon)$ is a k -dimensional ellipsoid centered at the origin with radius $r_i = a_k \sqrt{\lambda_{c,i} \sigma^2}$ in the i^{th} dimension. The interior of \mathbf{E}_k is represented by the equation $\sum_{i=0}^{k-1} \frac{x_i^2}{\lambda_{c,i}} \leq a_k^2 \sigma^2$. Let \mathbf{A}_k be an arbitrary k -dimensional shape with probability $(1 - \epsilon)$, the volume of which is lower-bounded by the volume of \mathbf{E}_k . Denote by \mathbf{B}_k the k -dimensional sphere with radius \sqrt{dT} . Finally, dividing the lower bound on the volume of \mathbf{A}_k by the volume of \mathbf{B}_k yields $R_L^k = \frac{1}{2} \sum_{i=0}^{k-1} \log_2 \left[\frac{\lambda_{c,i} a_k^2 \sigma^2}{dT} \right]$. \square

Note that the maximization is well defined and one can expect the “effective blocklength”,

$$n_L = \arg \max_k \frac{1}{2} \sum_{i=0}^{k-1} \log_2 \left[\frac{\lambda_{c,i} a_k^2 \sigma^2}{dT} \right], \quad (6.7)$$

to be close to c for the following reasons:

- For $0 \leq i \leq k \ll c$, $\lambda_{c,i} \approx 1$ and a_k^2 increases with k . Therefore, the right hand side of (6.7) increases with k whenever it is sufficiently smaller than c .

- For $c \ll i \leq k$, $\lambda_{c,i}$ decreases exponentially towards 0 (as shown in Figure 2.1) while a_k^2 increases almost linearly with k . Consequently, the right hand side of (6.7) decreases with k when it is sufficiently larger than c .

6.2.2 Achievability – Upper Bound

First note that an infinite number of –infinite dimensional– spheres of shape \mathbf{B} is required to cover the ellipsoid \mathbf{E} : Indeed, let \hat{M}_k to be the number of spheres \mathbf{B}_k required to cover \mathbf{E}_k (where \mathbf{B}_k and \mathbf{E}_k are as defined in Section 6.2.1). Using the result of Dumer et al. [35, Theorem 1], $\log_2 \hat{M}_k$ is lower bounded by

$$\frac{1}{2} \sum_{i \in \Lambda_k} \log_2 \left[\frac{\lambda_{c,i} a_k^2 \sigma^2}{dT} \right],$$

where $\Lambda_k = \{0 \leq i < k \text{ such that } \lambda_{c,i} a_k^2 \sigma^2 > dT\}$. By (6.6) a_k^2 grows to infinity as with k and so does \hat{M}_k . Therefore, finding an achievable covering of the ellipsoid \mathbf{E} will not lead to a useful upper bound.

To derive the upper bound, I give up on coding the whole vector \mathbf{X} and I divide it into two sub-vectors, where the first sub-vector is encoded and the second one is treated as pure distortion.

Theorem 6. *It is possible to encode a T -seconds portion of a W -Hz band-limited white Gaussian process with PSD defined in (6.1) using*

$$R_U = \min_k \log_2 \mathbb{M} \left[\frac{\hat{a}_k \sigma}{\sqrt{dT - \hat{b}_k^2 \sigma^2}}, k \right] \quad \text{bits}, \quad (6.8)$$

with the guarantee that the distortion does not exceed d with probability $(1 - \epsilon)$, where

$$\mathbb{M}(r, n) = \begin{cases} e(n \ln n + n \ln \ln n + 5n) r^n & n \leq r \\ n(n \ln n + n \ln \ln n + 5n) r^n & \frac{n}{\ln n} \leq r < n \\ \frac{7^{(4/7) \ln 7}}{4} \sqrt{2\pi} \frac{n\sqrt{n} \left[(n-1) \ln rn + (n-1) \ln \ln n + \frac{1}{2} \ln n + \ln \frac{\pi\sqrt{2n}}{\sqrt{\pi n - 2}} \right]}{r \left(1 - \frac{2}{\ln n}\right) \left(1 - \frac{2}{\sqrt{\pi n}}\right) \ln^2 n} r^n & 2 < r < \frac{n}{\ln n} \\ \sqrt{2\pi} \frac{\sqrt{n} \left[(n-1) \ln rn + (n-1) \ln \ln n + \frac{1}{2} \ln n + \ln \frac{\pi\sqrt{2n}}{\sqrt{\pi n - 2}} \right]}{r \left(1 - \frac{2}{\ln n}\right) \left(1 - \frac{2}{\sqrt{\pi n}}\right)} r^n & 1 < r \leq 2 \end{cases} \quad (6.9)$$

and where \hat{a}_k and \hat{b}_k are defined as

$$\mathbb{P} \left[\sum_{i=0}^{k-1} X_i^2 < \hat{a}_k^2 \sigma^2 \right] = \mathbb{P} \left[\sum_{i=k}^{\infty} X_i^2 < \hat{b}_k^2 \sigma^2 \right] = \sqrt{1 - \epsilon}, \quad (6.10)$$

where X_i are independent $\sim \mathcal{N}(0, \lambda_{c,i} \sigma^2)$.

Proof. First divide \mathbf{X} into two sub-vectors $\mathbf{X}^k = \{\mathbf{X}_i\}_{i=0}^{k-1}$ and $\hat{\mathbf{X}}^k = \{\mathbf{X}_i\}_{i=k}^{\infty}$. Consider the coding scheme where only \mathbf{X}^k is encoded into bits while taking into account the distortion associated with ignoring $\hat{\mathbf{X}}^k$. Consider the k -dimensional sphere \mathbf{S}_k centered at the origin with a radius $\hat{a}_k\sigma$ that guarantees that \mathbf{X}^k falls inside \mathbf{S}_k with probability $\sqrt{1-\epsilon}$. Being independent¹,

$$\mathbb{P} \left[\{\mathbf{X}^k \in \mathbf{S}_k\} \cap \left\{ \sum_{i=k}^{\infty} X_i^2 < \hat{b}_k^2 \sigma^2 \right\} \right] = (1-\epsilon).$$

Note that whenever $\sum_{i=k}^{\infty} X_i^2 < \hat{b}_k^2 \sigma^2$, the maximum allowable distortion for encoding \mathbf{X}^k is $dT - \hat{b}_k^2 \sigma^2$.

The problem at hand becomes the problem of finding the smallest achievable number of spheres with radius $\sqrt{dT - \hat{b}_k^2 \sigma^2}$ that can cover \mathbf{S}_k . The proof proceeds as in that of [13, Theorem 2] and equation (6.9) is identical to [13, equation (7)]—making use of results in [36, 37]. \square

I emphasize that I didn't consider encoding the shape having the smallest volume since the number of equal k -dimensional balls that cover a k -dimensional ellipsoid is less tractable than that of those that cover a sphere, and the available results in the literature are not satisfactory as shown in [38, Section V.A].

I refer to $n_U = \arg \min_k \log_2 \mathbb{M} \left[\frac{\hat{a}_k \sigma}{\sqrt{dT - \hat{b}_k^2 \sigma^2}}, k \right]$ as the effective blocklength for the upper bound. Using the Berry-Esseen theorem and table 2.1, one can show that $\hat{a}_k^2 = \sum_{i=0}^{k-1} \lambda_{c,i} + \sqrt{2 \sum_{i=0}^{k-1} \lambda_{c,i}^2} Q^{-1}(1 - \sqrt{1-\epsilon}) + O(1)$ and $\hat{b}_k^2 = \sum_{i=k}^{\infty} \lambda_{c,i} + \sqrt{2 \sum_{i=k}^{\infty} \lambda_{c,i}^2} Q^{-1}(1 - \sqrt{1-\epsilon}) + O\left(\frac{\sum_{i=k}^{\infty} \lambda_{c,i}^3}{\sum_{i=k}^{\infty} \lambda_{c,i}^2}\right)$. As k increases to infinity, \hat{b}_k^2 and \hat{a}_k^2 converge to zero and a constant respectively by virtue of (2.13) and table 2.1. For $k \gg c$, $\frac{\hat{a}_k \sigma}{\sqrt{dT - \hat{b}_k^2 \sigma^2}}$ remains essentially constant and the $\log_2 \mathbb{M}[\dots]$ term is increasing for sufficiently large k implying that n_U is finite and expected to be close to c .

6.3 Asymptotic analysis as $T \rightarrow \infty$

Theorem 7. *As T goes to infinity, the rate per second converges to the Shannon rate $W \log_2 \left[\frac{2W\sigma^2}{d} \right]$ bits/sec.*

Proof. First note that as T goes to infinity, so does c . I established above that for all $k \in \mathbb{N}$,

$$\frac{1}{2T} \sum_{i=0}^{k-1} \log_2 \left[\frac{\lambda_{c,i} a_k^2 \sigma^2}{dT} \right] \leq \frac{R^*}{T} \leq \frac{1}{T} \log_2 \mathbb{M} \left[\frac{\hat{a}_k \sigma}{\sqrt{dT - \hat{b}_k^2 \sigma^2}}, k \right].$$

¹One can look for another trade-off between the probabilities of the two events.

For any $\gamma > 0$, one can choose $k = c(1 - \gamma)$ in the lower bound. Using equation (2.12) and the fact that $a_k^2 \geq k + \sqrt{k/2}$ for c –and hence k – large enough

$$\frac{R^*}{T} \geq (1 - \gamma)W \log_2 \left[\frac{2W(1 - \gamma)\sigma^2}{d} \right] + (1 - \gamma)W \log_2 \left[1 + \sqrt{\frac{1}{2c(1 - \gamma)}} \right].$$

By continuity of the logarithm, the second term is arbitrarily small as $c \rightarrow \infty$. Similarly, choose $k = c(1 + \gamma)$ in the upper bound and use equation (2.11) to show that as c –and hence k – goes to infinity,

$$\frac{R^*}{T} \leq (1 + \gamma)W \log_2 \left[\frac{2W\sigma^2}{d} \right].$$

□

6.4 Numerical results

I compare the derived bounds with Shannon’s expression (1.1) by computing the ratio $\frac{R}{T \times R_{\text{Shannon}}} = \frac{R}{0.5 c \log_2 \left[\frac{2W\sigma^2}{d} \right]}$.

Using MATLAB, I numerically estimate the solution of equation (6.10), which is needed to compute the upper bound. In Figures 6.2, and 6.3 I present the numerical results using various sample values for d , ϵ , W , and σ^2 .

The results show that the relative gap between the bounds decreases as T increases (or equivalently as c increases).

For $\epsilon = 10^{-2}$ and over the selected range of c , I found:

- $1 \leq n_U - c \leq 2$ and $n_L = c + 2$ for $d = \frac{1}{4}2W$
- $4 \leq n_U - c \leq 6$ and $6 \leq n_L - c \leq 8$ for $d = \frac{1}{1000}2W$.

Although I used relatively low level of distortion in Figure 6.3, the relative gap between the bounds remains small, and the bounds remain tight in this case.

6.5 Remarks

On the discrete-time representation of the time-limited piece: My analysis is done through projecting the time-limited piece on the set of PSWFs. While there are alternative ways to represent the time-limited piece by a discrete-time vector such as a Fourier series decomposition or through sampling, the use of PSWFs is necessary to get independent random variables as shown in [7, section 8.4].

On non-white band-limited Gaussian sources: Whenever $X(t)$ is a W -Hz band-limited Gaussian process with non-flat PSD, projecting it on the set of

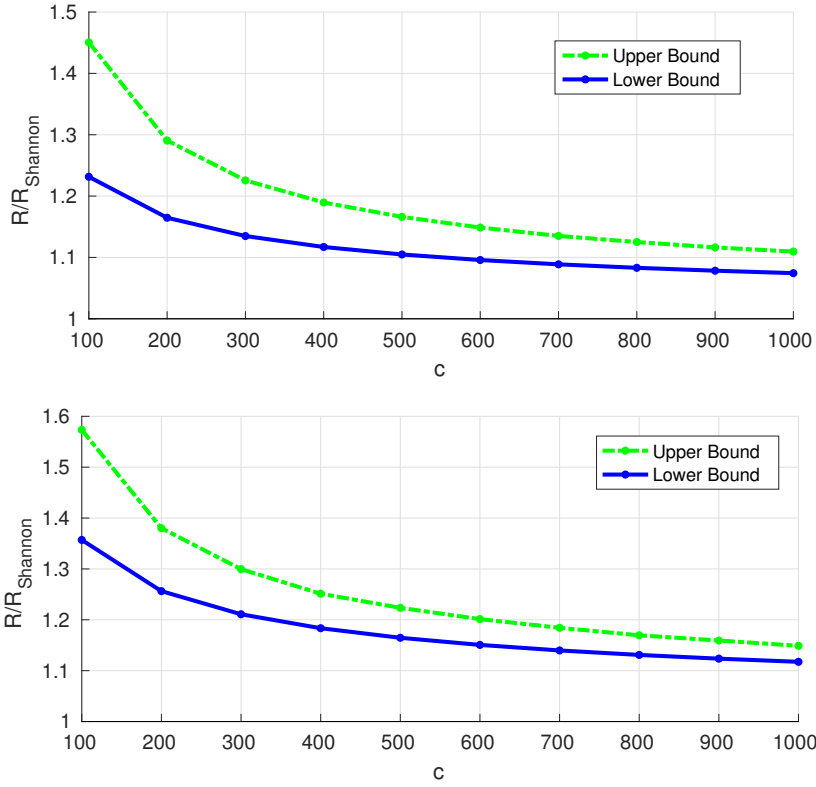


Figure 6.2: Bounds vs. c for $\sigma = 1$, $W = 1000\text{Hz}$, $d = \frac{1}{4}2W$ and $\epsilon = 10^{-2}$ in the first figure and $\epsilon = 10^{-4}$ in the second.

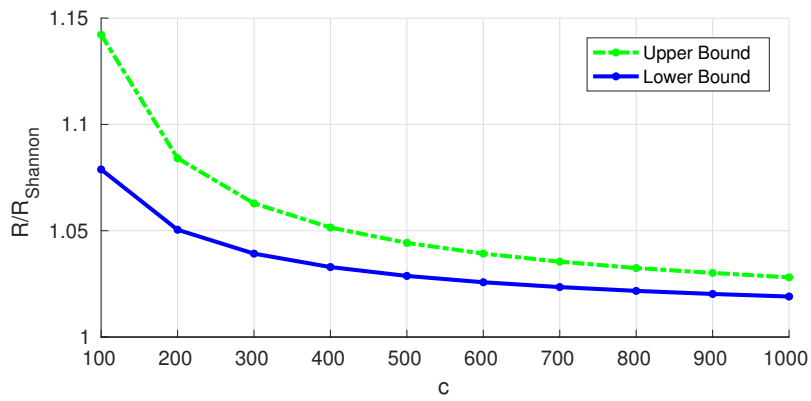


Figure 6.3: Bounds vs. c for $\sigma = 1$, $W = 1000\text{Hz}$, $d = \frac{1}{1000}2W$, $\epsilon = 10^{-2}$

PSWFs does not provide independent discrete-time symbols. One possible way to tackle the problem is to follow the derivations by Gallager in [7, section 8.4] to find a CON set that produces independent discrete-time symbols. However, the resulting Fredholm integral equation is difficult to solve [23].

Summary and conclusions

Wireless communication systems are allocated specific radio bands, where the transmitter should confine its transmissions within the allocated band while maintaining low level of interference on the neighbouring bands, and the receiver should look for the data in this band since the neighbouring bands may be used by someone else. Therefore the capacity of band-limited channels derived by Shannon is very useful since it determines the highest possible data rates for such systems. However, band-limited signals are not practical since they need an infinite time support. It is therefore interesting to study the achievable data rates when sending time-limited-signals over band-limited channels.

A related problem is to determine the optimal source coding rates of a time-limited piece taken from a band-limited source; some “approximately band-limited” signals are treated as band-limited signals and are usually passed through a low pass filter and then sampled, because any information outside the band of interest is useless (for example frequency components of the sound that are outside the human audible range).

In chapter 3 I evaluated the achievable information rates when using finite duration pulses in a combined PAM-OMM system over a band-limited Gaussian channel. I have shown that when the time-frequency factor $c = 2WT$ is greater or equal to 1, Shannon’s channel capacity can be achieved by using the proposed combined PAM-OMM system. Although achieving the exact Shannon’s capacity requires the use of very large number of parallel filters, I numerically showed that the losses can be made negligible by using a finite number of parallel filters. For example, for $c = 1$ using 5 parallel filters incurs less than $1.5 \times 10^{-6}\%$ losses in the data rates over the considered SNR range. Finally, I provided a new and non-asymptotic (in WT) proof for the fact that one can have at most $2WT$ independent symbols given certain time window T and certain bandwidth W .

In chapter 4 I studied the information rates of PAM systems when the channel is a W -Hz band-limited AWGN channel and assessed the performance when using time-limited pulses. I considered three scenarios with different signaling rates; first using Nyquist pulses with $2WT > 1$, second transmitting at the Nyquist rate ($2WT = 1$), and third when $2WT < 1$. I derived the achievable rates for the three cases when using band-limited (hence non-constrained) pulses which serve as upper bounds for practical PAM systems with time-limited pulses. In the first

scenario my numerical computations indicate that the use of time-limited RRC has a negligible degradation in achievable rates at low $\frac{\text{SNR}}{2W}$. In the second scenario I numerically searched for the optimal time-limited filters by performing an expansion on the PSWFs basis. Considering incrementally more basis elements, I approximate the optimal solution by often using only a few PSWFs. In the third scenario, numerical results show that signaling at faster than the Nyquist rate is significantly beneficial and surprisingly, it can achieve rates that are very close to Shannon's capacity.

Based on the results of chapter 3 and chapter 4, one can conclude that: when using infinite blocklength codewords, the effect of using time-limited pulses can be made negligible.

In chapter 5 I studied the maximum achievable rates and the available degrees of freedom when transmitting T -seconds time-limited codewords over a W -Hz band-limited AWGN channel. I made use of the PSWFs to switch to DT and then applied the results by Polyanskiy. I derived the upper bound by ignoring the interference due to other codewords, and I derived lower bounds by deriving upper bounds on the inter-codeword interference. Finally I proposed an approximation of achievable rates and the available degrees of freedom based on the numerical results.

In chapter 6 I studied the rate distortion limits when independently coding T -seconds time-limited signals from a W -Hz band-limited white Gaussian source using finite block lengths. I derived upper and lower bounds and I numerically evaluated them. As the time frequency index $c = 2WT$ increases, the relative gap between the bounds decreases and the derived bounds approach Shannon's formula.

The natural extension of this work is the study of the achievable rates for CT fading channels. The difficulty of the problem is accentuated by the Doppler effect where the bandwidth of the signals changes. Additionally, the signals undergo additional transformations when passed through the channel which renders these problems harder.

Appendix A

Properties of ${}_{k,h}\alpha_{c,l,m}$

The inner product between the l^{th} normalized time-limited PSWF already shifted in time and frequency by kT -seconds and $h2W$ -Hz respectively, and the band-limited version of the m^{th} normalized time-limited PSWF is denoted in this thesis by ${}_{k,h}\alpha_{c,l,m}$. By Equation (2.10) and using Parseval,

$$\begin{aligned} {}_{k,h}\alpha_{c,l,m} &\hat{=} \left\langle \frac{D\varphi_{c,l}(t-kT)}{\sqrt{\lambda_{c,l}}} e^{j2\pi h2W(t-kT)}, \sqrt{\lambda_{c,m}}\varphi_{c,m}(t) \right\rangle \\ &= \left\langle \Phi_{c,l}^D(f-h2W)e^{-j2\pi kTf}, \Phi_{c,m}^D(f) \text{rect}\left(\frac{f}{2W}\right) \right\rangle \\ &= \left\langle \Phi_{c,l}^D(f-h2W)e^{-j2\pi kTf} \text{rect}\left(\frac{f}{2W}\right), \Phi_{c,m}^D(f) \right\rangle. \end{aligned}$$

Property A1. The coefficients ${}_{k,h}\alpha_{c,l,m}$ satisfy

$${}_{k,h}\alpha_{c,l,m} = j^{m-l} \sqrt{\frac{\lambda_{c,m}}{\lambda_{c,l}}} {}_{-h,-k}\alpha_{c,m,l}.$$

Proof. This can be readily obtained by noting that by (2.9)

$$\begin{aligned} {}_{k,h}\alpha_{c,l,m} &= \left\langle \frac{D\varphi_{c,l}(t-kT)}{\sqrt{\lambda_{c,l}}} e^{j2\pi h2W(t-kT)}, \sqrt{\lambda_{c,m}}\varphi_{c,m}(t) \right\rangle \\ &= \left\langle \Phi_{c,l}^D(f-h2W)e^{-j2\pi kTf} \text{rect}\left(\frac{f}{2W}\right), \Phi_{c,m}^D(f) \right\rangle \\ &= \left\langle j^l \sqrt{\frac{T}{2W}} \varphi_{c,l}\left(\frac{T}{2W}(f-h2W)\right) e^{-j2\pi kTf} \text{rect}\left(\frac{f}{2W}\right), j^m \sqrt{\frac{T}{2W}} \varphi_{c,m}\left(\frac{T}{2W}f\right) \right\rangle \\ &= j^{l-m} \frac{T}{2W} \left\langle \varphi_{c,l}\left(\frac{T}{2W}(f-h2W)\right), \varphi_{c,m}\left(\frac{T}{2W}f\right) \text{rect}\left(\frac{f}{2W}\right) e^{j2\pi kTf} \right\rangle \end{aligned}$$

By a change of variable $\nu = Tf/2W$,

$$\begin{aligned}
{}_{k,h}\alpha_{c,l,m} &= j^{l-m} \left\langle \varphi_{c,l}(\nu - hT), \varphi_{c,m}(\nu) \text{rect} \left(\frac{\nu}{T} \right) e^{j2\pi k 2W \nu} \right\rangle \\
&= j^{l-m} \left\langle \varphi_{c,l}(\nu), D\varphi_{c,m}(\nu + hT) e^{j2\pi k 2W(\nu + hT)} \right\rangle \\
&= j^{l-m} \sqrt{\frac{\lambda_{c,m}}{\lambda_{c,l}}} \left\langle \frac{D\varphi_{c,m}(\nu + hT)}{\sqrt{\lambda_{c,m}}} e^{-j2\pi k 2W(\nu + hT)}, \sqrt{\lambda_{c,l}} \varphi_{c,l}(\nu) \right\rangle \\
&= j^{m-l} \sqrt{\frac{\lambda_{c,m}}{\lambda_{c,l}}} {}_{-h,-k}\alpha_{c,m,l}.
\end{aligned}$$

□

Upper Bounds

Various upper bounds on the magnitude of ${}_{k,h}\alpha_{c,l,m}$ may be found using Cauchy Schwarz's inequality.

Property A2. The magnitude of ${}_{k,h}\alpha_{c,l,m}$ is upperbounded by

$$|{}_{k,h}\alpha_{c,l,m}|^2 \leq \lambda_{c,m} \int_{(2k-1)T/2}^{(2k+1)T/2} |\varphi_{c,m}(t)|^2 dt \quad (\text{A.1})$$

$$|{}_{k,h}\alpha_{c,l,m}|^2 \leq \lambda_{c,m} \int_{(-2h-1)T/2}^{(-2h+1)T/2} |\varphi_{c,l}(t)|^2 dt \quad (\text{A.2})$$

$$|{}_{k,h}\alpha_{c,l,m}|^2 \leq \int_{(-2k-1)T/2}^{(-2k+1)T/2} |g_{h,l}(t)|^2 dt \quad (\text{A.3})$$

$$|{}_{k,h}\alpha_{c,l,m}|^2 \leq \frac{\lambda_{c,m}}{\lambda_{c,l}} \int_{(2h-1)T/2}^{(2h+1)T/2} |g_{-k,m}(t)|^2 dt \quad (\text{A.4})$$

where

$$\begin{aligned}
g_{h,l}(t) &\hat{=} \left[\frac{D\varphi_{c,l}(t)}{\sqrt{\lambda_{c,l}}} e^{j2\pi h 2Wt} \right] * 2W \text{sinc}(2Wt) \\
G_{h,l}(f) &\hat{=} \Phi_{c,l}^D(f - h 2W) \text{rect} \left(\frac{f}{2W} \right).
\end{aligned}$$

Before proceeding to the proof, I specialize the bounds (A.3) and (A.4) to the values of $h = 0$ and $k = 0$ respectively. Using the fact that $g_{0,l}(t) = \left[\frac{D\varphi_{c,l}(t)}{\sqrt{\lambda_{c,l}}} \right]^* 2W \operatorname{sinc}(2Wt) = \sqrt{\lambda_{c,l}} \varphi_{c,l}(t)$,

$$|_{k,0} \alpha_{c,l,m}|^2 \leq \int_{(-2k-1)T/2}^{(-2k+1)T/2} |g_{0,l}(t)|^2 dt = \lambda_{c,l} \int_{(-2k-1)T/2}^{(-2k+1)T/2} |\varphi_{c,l}(t)|^2 dt \quad (\text{A.5})$$

$$|_{0,h} \alpha_{c,l,m}|^2 \leq \frac{\lambda_{c,m}}{\lambda_{c,l}} \int_{(2h-1)T/2}^{(2h+1)T/2} |g_{0,m}(t)|^2 dt = \frac{\lambda_{c,m}^2}{\lambda_{c,l}} \int_{(2h-1)T/2}^{(2h+1)T/2} |\varphi_{c,m}(t)|^2 dt. \quad (\text{A.6})$$

Proof. Bound (A.1) is obtained by using Cauchy Schwarz,

$$\begin{aligned} {}_{k,h} \alpha_{c,l,m} &= \left\langle \frac{D\varphi_{c,l}(t-kT)}{\sqrt{\lambda_{c,l}}} e^{j2\pi h 2W(t-kT)}, \sqrt{\lambda_{c,m}} \varphi_{c,m}(t) \right\rangle \\ &= \left\langle \frac{D\varphi_{c,l}(t-kT)}{\sqrt{\lambda_{c,l}}} e^{j2\pi h 2W(t-kT)}, \sqrt{\lambda_{c,m}} \varphi_{c,m}(t) \operatorname{rect}\left(\frac{t-kT}{T}\right) \right\rangle \\ |{}_{k,h} \alpha_{c,l,m}|^2 &\leq \lambda_{c,m} \left\| \varphi_{c,m}(t) \operatorname{rect}\left(\frac{t-kT}{T}\right) \right\|^2 \left\| \frac{D\varphi_{c,l}(t-kT)}{\sqrt{\lambda_{c,l}}} \right\|^2 \\ &= \lambda_{c,m} \int_{(2k-1)T/2}^{(2k+1)T/2} |\varphi_{c,m}(t)|^2 dt. \end{aligned}$$

When it comes to (A.2), using Property A1,

$$|{}_{k,h} \alpha_{c,l,m}|^2 = \frac{\lambda_{c,m}}{\lambda_{c,l}} |_{-h,-k} \alpha_{c,m,l}|^2 \leq \lambda_{c,m} \int_{(-2h-1)T/2}^{(-2h+1)T/2} |\varphi_{c,l}(t)|^2 dt.$$

Bound (A.3) may be derived as follows:

$$\begin{aligned}
{}_{k,h}\alpha_{c,l,m} &= \left\langle \Phi_{c,l}^D(f - h 2W) e^{-j2\pi k T f} \operatorname{rect}\left(\frac{f}{2W}\right), \Phi_{c,m}^D(f) \right\rangle \\
&= \left\langle \Phi_{c,l}^D(f - h 2W) \operatorname{rect}\left(\frac{f}{2W}\right), \Phi_{c,m}^D(f) e^{j2\pi k T f} \right\rangle \\
&= \langle G_{h,l}(f), \Phi_{c,m}^D(f) e^{j2\pi k T f} \rangle \\
&= \left\langle g_{h,l}(t), \frac{D\varphi_{c,m}(t + k T)}{\sqrt{\lambda_{c,m}}} \right\rangle \\
&= \left\langle g_{h,l}(t) \operatorname{rect}\left(\frac{t + k T}{T}\right), \frac{D\varphi_{c,m}(t + k T)}{\sqrt{\lambda_{c,m}}} \right\rangle,
\end{aligned}$$

and hence, by Cauchy Schwarz

$$|{}_{k,h}\alpha_{c,l,m}|^2 \leq \int_{(-2k-1)T/2}^{(-2k+1)T/2} |g_{h,l}(t)|^2 dt.$$

Using Property A1, bound (A.4) is readily obtained

$$|{}_{k,h}\alpha_{c,l,m}|^2 = \frac{\lambda_{c,m}}{\lambda_{c,l}} |{}_{-h,-k}\alpha_{c,m,l}|^2 \leq \frac{\lambda_{c,m}}{\lambda_{c,l}} \int_{(2h-1)T/2}^{(2h+1)T/2} |g_{-k,m}(t)|^2 dt.$$

□

When it comes to sums over one index, the following inequalities can be immediately obtained.

Corollary A1. The following bounds are directly obtained from the respective bounds

derived above.

$$\sum_{k \in \mathbb{Z}^*} |k, h \alpha_{c, l, m}|^2 \leq \lambda_{c, m} (1 - \lambda_{c, m}) \quad (\text{A.7})$$

$$\sum_{h \in \mathbb{Z}^*} |k, h \alpha_{c, l, m}|^2 \leq \lambda_{c, m} (1 - \lambda_{c, l}) \quad (\text{A.8})$$

$$\sum_{k \in \mathbb{Z}^*} |k, h \alpha_{c, l, m}|^2 \leq \left[\|g_{h, l}(t)\|^2 - \int_{-T/2}^{T/2} |g_{h, l}(t)|^2 dt \right] \leq \|g_{h, l}(t)\|^2 \quad (\text{A.9})$$

$$\sum_{h \in \mathbb{Z}^*} |k, h \alpha_{c, l, m}|^2 \leq \frac{\lambda_{c, m}}{\lambda_{c, l}} \left[\|g_{-k, m}(t)\|^2 - \int_{-T/2}^{T/2} |g_{-k, m}(t)|^2 dt \right] \leq \frac{\lambda_{c, m}}{\lambda_{c, l}} \|g_{-k, m}(t)\|^2 \quad (\text{A.10})$$

$$\sum_{k \in \mathbb{Z}^*} |k, 0 \alpha_{c, l, m}|^2 \leq \lambda_{c, l} (1 - \lambda_{c, l}) \quad (\text{A.11})$$

$$\sum_{h \in \mathbb{Z}^*} |0, h \alpha_{c, l, m}|^2 \leq \frac{\lambda_{c, m}^2}{\lambda_{c, l}} (1 - \lambda_{c, m}). \quad (\text{A.12})$$

Finally, considering sums over the two indices k and h , note by Plancherel that

$$\sum_{h \in \mathbb{Z}^*} \|g_{h, l}(t)\|^2 = \sum_{h \in \mathbb{Z}^*} \|G_{h, l}(f)\|^2 = \sum_{h \in \mathbb{Z}^*} \int_{(2h-1)W}^{(2h+1)W} |\Phi_{c, l}^D(f)|^2 df = (1 - \lambda_{c, l}),$$

and hence bounds (A.9) and (A.10) imply

$$\sum_{h \in \mathbb{Z}^*} \sum_{k \in \mathbb{Z}^*} |k, h \alpha_{c, l, m}|^2 \leq (1 - \lambda_{c, l}) \quad (\text{A.13})$$

$$\sum_{h \in \mathbb{Z}^*} \sum_{k \in \mathbb{Z}^*} |k, h \alpha_{c, l, m}|^2 \leq \frac{\lambda_{c, m}}{\lambda_{c, l}} (1 - \lambda_{c, m}). \quad (\text{A.14})$$

Appendix B

Abbreviations

AWGN	Additive White Gaussian Noise
CON	Complete Orthonormal
CMB	Consecutive Multi-Band
CSB	Consecutive Single Band
CT	Continuous Time
DT	Discrete Time
FTN	Faster Than Nyquist
ISI	Inter-Symbol Interference
JWSS	Jointly Wide Sense Stationary
OMM	Orthogonal Multi-pulse Modulation
PAM	Pulse Amplitude Modulation
PDF	Probability Density Function
PSD	Power Spectral Density
PSWF	Prolate Spheroidal Wave Function
RC	Raised Cosine
RF	Radio Frequency
RRC	Root-Raised Cosine
SNR	Signal to Noise Ratio
STMB	Single Time-slot Multi-Band
TUB	Tighter Upper Bound
WSS	Wide Sense Stationary

Bibliography

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379-423; 623-656, July 1948.
- [2] A. D. Wyner, “Fundamental limits in information theory,” *Proc. IEEE*, pp. 239-251, Frb. 1981.
- [3] C. E. Shannon, “Communication in the Presence of Noise,” *Proc. IRE*, vol. 37, pp. 10-21, Jan. 1949.
- [4] “Scenarios, requirements and KPIs for 5G mobile and wireless system,” , ICT-317669-METIS/D1.1, 2013.
- [5] A. D. Wyner, “The Capacity of the band-Limited Gaussian Channel,” *Bell Syst. Tech. J.*, 45, pp. 359-395, Dec. 1965.
- [6] A. D. Wyner, “A Note on The Capacity of the Band-Limited Gaussian Channel,” *Bell Syst. Tech. J.*, vol. 55, no. 3, pp. 343-346, 1976.
- [7] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Nov. 1968.
- [8] R. Dobrushin, “Mathematical problems in the shannon theory of optimal coding of information,” in *Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability*, vol. 1, pp. 211–252, 1961.
- [9] V. Strassen, “Asymptotische abschatzugen in shannon’s informationstheorie,” in *Transactions of the Third Prague Conference on Information Theory etc, 1962. Czechoslovak Academy of Sciences, Prague*, pp. 689–723, 1962.
- [10] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits*. PhD thesis, Princeton University, 2010.
- [11] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Dispersion of gaussian channels,” in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pp. 2204–2208, IEEE, 2009.

- [12] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall electrical engineering series, Prentice-Hall, 1971.
- [13] V. Kostina and S. Verdú, “Fixed-length lossy compression in the finite block-length regime: Gaussian source,” in *2011 IEEE Information Theory Workshop*, pp. 457–461, IEEE, 2011.
- [14] A. Ingber and Y. Kochman, “The dispersion of lossy source coding,” in *2011 Data Compression Conference*, pp. 53–62, IEEE, 2011.
- [15] H. Nyquist, “Certain Topics in Telegraph Transmission Theory,” *AIEE Trans.*, vol. 47, pp. 617-644, April 1928.
- [16] D. Gabor, “Theory of Communication,” *J. Inst. Elect. Eng. (London)*, vol. 93, no. 26, pp. 429-457, 1946.
- [17] H. J. Landau and H. O. Pollak, “Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-III: The Dimension of the Space Essentially Time- and Band-Limited Signals,” *Bell Syst. Tech. J.*, vol 41, pp. 1295-1336, July 1962.
- [18] D. Slepian, “On Bandwidth,” *Proc. IEEE*, vol. 64, No. 3, pp. 292-300, 1976.
- [19] D. Slepian and H. Pollak, “Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-I,” *Bell Syst. Tech. J.*, vol 40, pp. 43-64, 1961.
- [20] F. Smithies, *Integral Equations*. No. 49, Cambridge Tracts in Mathematics and Mathematical Physics, 1958.
- [21] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. 1. Interscience Publishers, New York, 1953.
- [22] D. Slepian, “Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-IV: Extensions to Many Dimensions; Generalized Prolate Spheroidal Functions,” *Bell Syst. Tech. J.*, vol 43, no. 6, pp. 3009-3057, 1964.
- [23] F. A. Grunbaum, “Differential operators commuting with convolution integral operators,” *J. Math. Anal. Appl.*, 91, pp. 80-93, 1983.
- [24] J. Mercer, “Functions of positive and negative type, and their connection with the theory of integral equations,” *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, vol. 209, pp. 415–446, 1909.

- [25] R. Adelman, N. A. Gumerov, and R. Duraiswami, “Software for Computing the Spheroidal Wave Functions Using Arbitrary Precision Arithmetic,” <http://arxiv.org/abs/1408.0074>, (Software at <http://www.github.com/radelman/scattering>), 2014.
- [26] B. S. Tsybakov, “Capacity of a discrete-time gaussian channel with a filter,” *Problemy Peredachi Informatsii*, vol. 6, no. 3, pp. 78–82, 1970.
- [27] Emre Telatar, “Capacity of Multi-antenna Gaussian Channels,” *Europ. Trans. on Telecommun.*, vol. 10, pp. 585–595, 1999.
- [28] H. Wolkowicz and G. P. Styan, “Bounds for eigenvalues using traces,” *Linear algebra and its applications*, vol. 29, pp. 471–506, 1980.
- [29] D. J. Hartfiel, *Matrix theory and applications with MATLAB*. Crc Press, 2000.
- [30] J. Mazo, “Faster-than-nyquist signaling,” *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1451–1462, 1975.
- [31] J. E. Mazo and H. J. Landau, “On the minimum distance problem for faster-than-nyquist signaling,” *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1420–1427, 1988.
- [32] F. Rusek and J. B. Anderson, “Constrained capacities for faster-than-nyquist signaling,” *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 764–775, 2009.
- [33] A. D. Liveris, *On distributed coding, quantization of channel measurements and faster-than-Nyquist signaling*. PhD thesis, Texas A&M University, 2006.
- [34] J. B. Anderson, *Bandwidth Efficient Coding*. John Wiley & Sons, 2017.
- [35] I. Dumer, M. S. Pinsker, and V. V. Prelov, “On coverings of ellipsoids in euclidean spaces,” *IEEE transactions on information theory*, vol. 50, no. 10, pp. 2348–2356, 2004.
- [36] C. Rogers, “Covering a sphere with spheres,” *Mathematika*, vol. 10, no. 2, pp. 157–164, 1963.
- [37] J.-L. Verger-Gaugry, “Covering a ball with smaller equal balls in n ,” *Discrete & Computational Geometry*, vol. 33, no. 1, pp. 143–155, 2005.
- [38] P. Tian and V. Kostina, “The dispersion of the gauss-markov source,” *IEEE Transactions on Information Theory*, 2019.

