Title
Large wireless networks : fundamental limits and design issues

Permalink
https://escholarship.org/uc/item/6722937q

Author
Minero, Paolo

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Large wireless networks: fundamental limits and design issues

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Communication Theory and Systems)

by

Paolo Minero

Committee in charge:

    Professor Massimo Franceschetti, Chair
    Professor Young-Han Kim, Co-Chair
    Professor Jorge Cortes
    Professor Bruce Driver
    Professor Ramesh Rao
    Professor Jack Wolf

2010

The dissertation of Paolo Minero is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2010

DEDICATION

To my mother Mariella and the memory of my father Giorgio

कर्मण्ये वाधिकारस्ते म फलेषु कदाचना

कर्मफलेह्तुर भुरमा ते संगोस्त्वकर्मानी॥

You have control over action alone, never over its fruits

Live not for the fruits of action, nor attach yourself to inaction

—Bhagavad Gita

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Foremost, I would like to thank my advisor Professor Massimo Franceschetti for his friendship, mentorship, and support. I will always be indebted to Massimo for my present and future academic achievements. From him, I have learnt the invaluable skills of keeping a broad research vision, of blending multi-disciplinary ideas for solving engineering problems and of presenting interesting results in an accessible manner. Massimo has always treated me like an equal partner in all our research endeavors, and it is largely because of him I find myself well prepared for starting my career in academia.

It has also been a great fortune to have Professor Young-Han Kim as a mentor and collaborator. His passion for excellence, simple proofs, and attention to detail have been addictive. I thank him for teaching me all this, and for his constant encouragement.

I owe a debt of gratitude to Professor David Tse for serving a mentor during my M.S. thesis, which inspired the work in Chapter 3 of this thesis. It is David who was the first to teach me how to do research, starting from my first year at Berkeley. I would also like to thank Professor Tara Javidi and Professor Jack Wolf for their mentorship and support. I am indebted to Professor Marco Donald Migliore for his tutelage and collaboration, and to Professor Subhrakanti Dey for his collaboration during his sabbatical year at UCSD.

I have had good fortune to have wonderful colleagues, with whom I have also shared a warm friendship. In particular, I am indebted to my colleagues Rathinakumar Appuswamy, Ehsan Ardestani, Salman Avestimehr, Krish Eswaran, Leonard Grokop, Nikhil Karamchandani, Bobak Nazer, and Anand Sarwate for the time they dedicated to discussing various research problems with me.

Finally, for all these years that were spent towards this thesis, I thank my family and friends for their love and support, and for always believing in me.

Chapter 2, in part, is a reprint of the material as it appears in M. Franceschetti, M. D. Migliore, P. Minero, "The capacity of wireless networks: information-theoretic and physical limits," *IEEE Trans. on Information Theory*, vol. 55, no. 8, August 2009. The dissertation author was the primary investigator and author

of this paper. Chapter 4, in part, is a reprint of the material as it appears in P. Minero, M. Franceschetti, S. Dey and G. N. Nair, "Data Rate Theorem for Stabilization Over Time-Varying Feedback Channels," *IEEE Trans. on Automatic Control*, vol 54, no. 2, pp. 243-255, February 2009. The dissertation author was the primary investigator and author of this paper.

VITA

2003	Laurea in Electrical Engineering, with highest honors, Polytechnic of Turin, Turin, Italy
2006	M. S. in Electrical Engineering and Computer Science, University of California, Berkeley
2010	Ph. D. in Electrical and Computer Engineering, University of California, San Diego

PUBLICATIONS

P. Minero, M. Franceschetti, S. Dey and G. N. Nair, "Data Rate Theorem for Stabilization Over Time-Varying Feedback Channels," *IEEE Trans. on Automatic Control*, vol 54, no. 2, pp. 243-255, February 2009

M. Franceschetti, M.D. Migliore, P. Minero, "The capacity of wireless networks: information-theoretic and physical limits," *IEEE Trans. on Information Theory*, vol. 55, no. 8, August 2009

ABSTRACT OF THE DISSERTATION

Large wireless networks: fundamental limits and design issues

by

Paolo Minero

Doctor of Philosophy in Electrical Engineering (Communication Theory and
Systems)

University of California, San Diego, 2010

Professor Massimo Franceschetti, Chair
Professor Young-Han Kim, Co-Chair

As information networks grow in magnitude and complexity, new models
and frameworks are necessary to understand the nature of information transmis-
sion. In this thesis we demonstrate how fundamental questions arising in the design
of large wireless networks can be addressed by applying methods from information
theory, physics, networking and control. We focus on three examples of emerging
systems architecture. First, we investigate the maximum achievable throughput
in a wireless *ad-hoc* network. By combining Maxwell's physics of wave propaga-
tion and Shannon's theory of information, and departing from idealistic stochastic
channel models for signal propagation, we derive an upper bound to the law that
determines the scaling of throughput with the population size of the network, and
conclude that the scaling achieved by multi-hop communication is optimal in any
constant density wireless network. Second, we study how to aggregate information
from uncoordinated nodes by considering a random-access system with multiple
nodes transmitting information to a common receiver. We characterize the maxi-

mum achievable throughput of channels of practical interest and demonstrate how the performance of current systems can be improved by allowing encoding rate adaptation at the transmitters and joint decoding at the receiver. Finally, we explore the fundamental limits of control over wireless channels and demonstrate the relationship between the degree of instability of a system and the time varying rate of communication in the feedback link.

Chapter 1

Introduction

As hardware technology advancements lead to dramatically decreasing dimension and cost of embedded sensor devices, deployment of large scale distributed and wireless sensing systems are fast becoming a reality in the near future. Such networks will encompass monitoring and control of all global distributed infrastructures, such as transportation systems, power grids, pipelines, water distribution networks, and the internet. With the growth of information networks' *magnitude* and *complexity*, the realization of this vision of pervasive networking requires us to revolutionize the way we design and manage large networks, and to solve a wide spectrum of engineering and mathematical challenges. How to efficiently and reliably transmit information in an *ad-hoc* network, collect information from a multitude of sensors, and control dynamical systems over wireless channels are some of these common challenges faced by the scientific community and object of study of this dissertation.

New models and frameworks are necessary to tackle these new engineering challenges. Traditionally, different disciplines have focused on specific aspects of large networks independently and in isolation. The structural properties of networks have been the territory of physics, that has developed an arsenal of tools to study the macroscopic behavior of a system derived from the microscopic properties of its constituents. Communication theory has focused on designing coding schemes for reliable information transmission in point-to-point, many-to-one, and one-to-many systems. Control theory has dealt with behavior of dynamical sys-

tems in presence of feedback. What is needed for the design and deployment of emerging large distributed networks of tomorrow is a *unified theory*, based on a cross disciplinary blending of ideas from the aforementioned fields. In this thesis we demonstrate how fundamental questions arising in the design of large wireless networks can be addressed by applying methods from information theory, physics, networking and control.

This thesis contains three self-contained articles focused on distinct systems architecture. In Chapter 2, we address the following fundamental questions in the field of wireless networks: how much information can be carried by a wireless *ad-hoc* network composed of many nodes, and how should the nodes cooperate to transfer such information? The approach taken is to depart from traditional stochastic fading and path loss channel models commonly used in the related literature, and to address these questions using first physical principles. This lead to theoretical results of fundamental flavor, which are not tied to specific fading and path loss models. Our main contribution is to use tools from electromagnetic theory to derive an upper bound to the law that determines the scaling of the throughput with the population size of the network, and to conclude that the scaling achieved by a simple multi-hop communication protocol is optimal in any constant density wireless network.

In Chapter 3, we consider a large wireless random access system where a random set of transmitters communicate to a single receiver, in an impulsive and uncoordinated fashion. In this setting, the amount of information which flows from transmitters to receiver is limited by the random level of interference at the receiver. Despite decades of active research, the theory to study random access communications is far from complete. On the one hand, information theory provides accurate models for the interference caused by simultaneous transmissions, but it ignores random information arrivals at the transmitters; on the other hand, network oriented studies focus on the impulsive nature of information transmission, but do not accurately describe the underlying physical channel model. In a quest to bridge the divide between these two approaches, we develop a model for studying random access systems which is information-theoretic in nature, but which also

accounts for the random activity of the users, as in models arising in the networking literature. We then apply this model to characterize the maximum amount of information which can be sent in several interesting random access systems.

Finally, Chapter 4 considers engineering applications where one or more dynamical systems are controlled using sensing devices and actuators communicating over wireless channels. In this setting, the amount of information which flows from sensors to actuator changes dynamically according to the channels condition. We use information theoretic techniques to characterize the fundamental constraints posed on the on control performance due to random fluctuations of the communication channel. Our main result consists in characterizing tight necessary and sufficient conditions to say when it is possible to design a communication scheme which changes dynamically following the fluctuations of the channels condition and, at the same time, is guaranteed to stabilize the system. Our result also create an important connection between earlier works in the literature, by showing a fundamental relationship between the degree of instability of the plant and the rate of the communication in the feedback control channel.

Chapter 2

Capacity scaling of ad-hoc networks

It is shown that the capacity scaling of wireless networks is subject to a fundamental limitation which is independent of power attenuation and fading models. It is a degrees of freedom limitation which is due to the laws of physics. By distributing uniformly an order of n users wishing to establish pairwise independent communications at fixed wavelength inside a two-dimensional domain of size of the order of n, there are an order of n communication requests originating from the central half of the domain to its outer half. Physics dictates that the number of independent information channels across these two regions is only of the order of $\sqrt{n}$, so the per-user information capacity must follow an inverse square-root of n law. This result shows that information-theoretic limits of wireless communication problems can be rigorously obtained without relying on stochastic fading channel models, but studying their physical geometric structure.

2.1   Introduction

A natural question that arises is whether information theory can provide fundamental bounds on the capacity of wireless *ad-hoc networks*, which are not tied to *ad-hoc physical channel models*. One aim of this work is to show that this is indeed the case, if the information-theoretic approach is appropriately combined with the study of the physics of wave propagation. The main contribution, however, should be seen in a broader context. Relying on functional analysis to study the vector space of the propagating field, rather than assuming stochastic fading channel models, could be a rigorous way to tackle other wireless communication problems.

The information theoretic characterization of the capacity region of wireless networks is one of the holy grails in information theory. It is a problem of great mathematical depth and engineering interest. One way to approach the problem is due to Gupta and Kumar [14], who proposed to study the simpler case in which all the nodes in the network are required to transmit at the same bit-rate, and to look at the *scaling limit* of the achievable rate, as the number of nodes in the network grows. In this way, the capacity region collapses to a single point and order results on its behavior are obtained. Gupta and Kumar's bounds were also derived under some additional assumptions on the physics of propagation, and on some restrictions on the communication strategy employed by the nodes (i.e. multi-hop operation and pairwise coding and decoding). Later, starting with the work of Xie and Kumar [38], *information-theoretic* scaling laws, independent of any strategy used for communication, have been established by many authors. These results, however, heavily depend on the assumptions made on the electromagnetic propagation process. Presence or absence of fading, choice of fading models, and choice of path loss models, lead to different lower and upper bounds on the scaling limit of the information rate. As a consequence, a plethora of articles appeared in the information-theoretic literature [2], [3], [9], [12], [17], [19], [20], [24], [25], [26], [27], [28], [39], [40], presenting bounds ranging from a per-node rate that rapidly decays to zero as the number of nodes in the network tends to infinity, to bounds predicting a slower decay, to bounds that are practically constant. In these works, while

the lower bounds rely on different cooperative schemes employed by the nodes, the upper bounds follow from the application of the same mathematical tool: the information-theoretic cut-set bound [6, Chapter 15]. This single strategy of attack, and the resulting dependence on ad-hoc physical propagation models, are somehow undesirable for a theory that seeks the fundamental limits of communication.

In the same two-dimensional geometric setting of the works above, this work shows that there exists a single scaling law, which is essentially an inverse-square-root of n law, and is dictated by Maxwell's physics of wave propagation, in conjunction to a Shannon-type cut-set bound. The result is then generalized to a three-dimensional setting at the end of this chapter. The main contribution in 2D is expressed as follows.

Claim: *In a wireless network composed of n uniformly distributed nodes subject to an individual (or total) power constraint, operating at a fixed wavelength inside a two-dimensional domain of area n (normalized to the wavelength), and which are uniformly paired into sources and destinations, each source can communicate to its intended destination at most at rate $O((\log n)^2/\sqrt{n})$ bits per second. This scaling law is a consequence of a limitation in the spatial degrees of freedom of the network that is independent of empirical path-loss models and stochastic fading models, but depends only on the geometrical configuration of the network.*

By looking closely at the claim above, we see a reflection of what Shannon has showed us, namely that the information capacity is limited by the power available for communication, *and* by the diversity available in the physical channel. In classical information theory, this diversity is expressed in terms of available frequency bandwidth. In the case of spatially distributed systems, such as wireless networks, this diversity constraint also appears in space. The usual approach of postulating stochastic fading channel models hides the explicit computation of the spatial diversity, while our analysis reveals it.

Being aware of such a fundamental limitation is certainly desirable, but what conclusions can be drawn from it on the optimal design and operation of wireless networks? Unfortunately not many. As it is often the case with fundamental limits, their generality can also be the curse of their practical applicability.

But we are not left completely empty handed of engineering guidelines either. One important implication is that any cooperative communication scheme cannot achieve a rate higher than what is stated in the claim above, at least in the scaling limit sense. Physics simply forbids it. Mathematical proofs of higher capacity scaling [2], [12], [24], [25], [26], [28], achieved using sophisticated cooperative communication schemes, rely on stochastic channel models and in a strict scaling limit sense are artifacts of such models. This highlights the importance of using appropriate mathematical models of reality to derive information-theoretic results. But does this also lead to the irrefutable conclusion that sophisticated cooperative strategies such as network coding, space-time coding, hierarchical cooperation, do not lead to any gain? The general answer is no. Scaling results are only up to order and pre-constants can make a huge difference in practice. Sophisticated cooperative communication schemes could in principle be extremely beneficial in networks of any fixed size. A rigorous proof of this latter statement is, however, difficult to obtain in a non-limiting scenario, and should take into account many practical issues related to protocol overhead, like decentralized medium access synchronization, and availability of channel state information.

Finally, we wish to spend some additional words on the mathematical techniques we use in this work. Resolving the amount of information that can be communicated through wave propagation is a venerable subject that has been treated by a great number of authors in different fields. Papers in optics often refer to the early works of Toraldo di Francia [36], [37]. In the mid nineteen-eighties the problem has been considered again in a more general context by Bucci and Franceschetti [4] [5], who introduced the important concepts of spatial bandwidth and degrees of freedom of scattered fields, and placed them into a rigorous functional analysis framework. More recently, the problem has been treated by the works of Miller [23], Piestum and Miller [32], Poon, Brodersen and Tse [33], and Migliore [22]. Our mathematical framework follows the approach of Bucci and Franceschetti, which we find to be the most rigorous, and does not require far-field approximations. There are some important differences, however. Bucci and Franceschetti first establish the spatial bandlimitation property of the field in their

first paper [4], and then they consider the problem of field reconstruction from a bounded observation set in their second paper [5], using prolate spheroidal functions, which are known to be optimal in the Landau-Pollack-Slepian sense. Given the specifics of our problem, we do not need this full machinery, but only inherit its main philosophy. We follow a singular value decomposition approach, which is standard in communication theory, and use simpler basis functions for the field expansion, which are good enough for our purposes. We then look directly at the behavior of the singular values of this decomposition, without performing a space-band transform. This leads to simpler computations and shortens the treatment considerably.

The next section formally defines the problem and outlines its solution. Section 2.3 completes the solution by studying the dimension of the Hilbert space spanned by the electromagnetic vector field. Section 2.4 presents the extension to a three-dimensional geometry and a final discussion of the results is presented in Section2.5.

2.2 Information-theoretic approach

Throughout this chapter, we consider distance lengths normalized to the carrier wavelength λ. Consider a Poisson point process $\mathscr{P}$ of unit density inside a disc $\mathcal{D}_n$ of radius $\sqrt{2n}$, and partition $\mathcal{D}_n$ into two equal parts by drawing a circular cut of radius $\sqrt{n}$ at the origin, which divides $\mathcal{D}_n$ into the inner disc D and the outer annulus A, where for convenience of notation we do not explicitly indicate the dependence on n. The points of the process represent the nodes of the network and we assume a uniform traffic pattern: nodes are paired independently and uniformly, so that there are an order of n communication requests that need to cross the boundary of the partition, see Figure 2.1.

Assuming that each node in $\mathcal{D}_n$ generates at most P watts [1], we want to find an upper bound on the per-node communication rate $R(n)$ that all nodes can achieve simultaneously to their intended destinations. To do so, we consider the

---

[1]Assuming a total power constraint rather than a per-node one does not change the results.
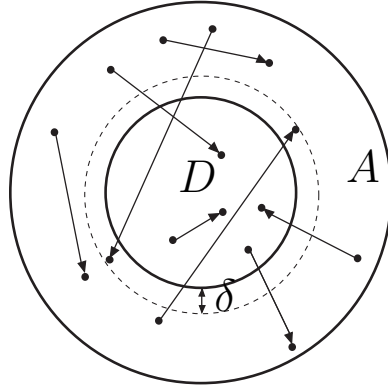
Figure 2.1: The partition considered in the analysis.

sum C_n of the rates that can be sent from the transmitters in D to the receivers in A. We have, with high probability (w.h.p.),

$$R(n) = O\left(\frac{C_n}{n}\right). \tag{2.1}$$

Next, to upper bound C_n we assume that the nodes on one side of the cut can share information instantaneously among themselves, and can also distribute the power among themselves in order to establish communication in the most efficient way with the nodes on the other side; which in turn are able to distribute the received information instantaneously among themselves. In this way, C_n is upper bounded by the capacity of a single user multiple-input multiple-output (MIMO) antenna array communicating across the partition.

The MIMO channel model across the cut is the space-vectorial version of the additive white Gaussian noise channel. In discrete time steps, it has the following representation:

$$Y_d[i] = \sum_{s \in \mathscr{P} \cap D} h_{sd}[i] X_s[i] + Z_d[i], \quad \text{for all } d \in \mathscr{P} \cap A, \tag{2.2}$$

where $X_s[i]$ are the symbols sent by node s at time i, $Y_d[i]$ are the symbols received by node d at time i, and $Z_d[i]$ are (independent space-time) Gaussian variables with unit variance. The coefficients $h_{sd}[i]$ model the strength of the propagation channel

between node s and node d and, given the realization of $\mathscr{P}$, are deterministically dictated by the physics through Maxwell equations. Throughout this chapter we assume a fixed environment, i.e., $h_{sd}[i] = h_{sd}$ for all i, but it will be clear that our results do not change in a dynamic environment where the coefficients h_{sd} vary over time. In matrix form, (2.2) is rewritten as

$$\mathbf{Y}_A[i] = \mathbf{H}\ \mathbf{X}_D[i] + \mathbf{Z}_A[i]. \tag{2.3}$$

Considering coding across time using blocks of m symbols and denoting the mutual information between space-time codewords $\mathbf{X}_D^m$ and $\mathbf{Y}_A^m$ as $\mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_A^m)$, the information flow through the cut can be upper bounded as follows:

$$mC_n \leq \max_{\substack{p(\mathbf{X}_D^m) \\ \sum_{i=1}^m X_s^2[i] \leq mP, \forall s \in \mathscr{P} \cap D}} \mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_A^m). \tag{2.4}$$

We now divide the information flow across the cut into two contributions. Let V be the annulus of constant width $\delta > 0$ around D. The first contribution is the information flow from the nodes in D to the nodes in V. The second contribution is the information flow from the nodes in D to the nodes in $A \setminus V$. Formally:

$$\begin{aligned}
\mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_A^m) &= \mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_V^m, \mathbf{Y}_{A \setminus V}^m) \\
&\leq \mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_V^m) + \mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_{A \setminus V}^m),
\end{aligned} \tag{2.5}$$

where the inequality holds as the space components of $\mathbf{Z}_A[i]$ are independent. Combining (2.4) and (2.5), it follows that

$$\begin{aligned}
mC_n \leq\ & \max_{\substack{p(\mathbf{X}_D^m) \\ \sum_{i=1}^m X_s^2[i] \leq mP, \forall s \in \mathscr{P} \cap D}} \mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_V^m) \\
& + \max_{\substack{p(\mathbf{X}_D^m) \\ \sum_{i=1}^m X_s^2[i] \leq mP, \forall s \in \mathscr{P} \cap D}} \mathcal{I}(\mathbf{X}_D^m; \mathbf{Y}_{A \setminus V}^m) \\
=:\ & mC^{(V)} + mC^{(A \setminus V)}.
\end{aligned} \tag{2.6}$$

Next, we consider the two terms in (2.6) separately and derive corresponding up-

per bounds. The first bound is obtained using standard information-theoretic arguments and relies only on the power constraint and on counting the number of transmitters and receivers, while the second bound is obtained by merging the information theory with a more detailed physical analysis of the wave propagation process.

Let us start with the easy part: we bound $C^{(V)}$ by summing the capacities of the individual multiple-input single-output (MISO) channels between all nodes in D and each receiver in V. We have, w.h.p.,

$$
\begin{aligned}
C_V &\leq \sum_{d \in \mathscr{P} \cap V} \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2} \sum_{s \in \mathscr{P} \cap D} |h_{sd}|^2 \right) \\
&\leq K_1 \sqrt{n} \log \left(1 + \frac{P}{\sigma^2} K_2\, n \max_{s \in \mathscr{P} \cap D, d \in \mathscr{P} \cap V} |h_{sd}|^2 \right) \\
&= O(\sqrt{n} \log n),
\end{aligned}
\tag{2.7}
$$

where K_1, K_2 are positive constants. The first inequality is a standard information-theoretic cut-set bound. The second inequality is due to the number of nodes in V being w.h.p. $O(\sqrt{n})$ and the number of nodes in $A \setminus V$ being w.h.p. $O(n)$. The last equality is due to $\max_{s \in \mathscr{P} \cap D, d \in \mathscr{P} \cap V} |h_{sd}|^2 = O(n)$, as one can at most beamform the total transmitted power on a single channel. Physically, the bound in (2.7) shows something very simple: there are at most a constant times $\sqrt{n}$ independent output channels, and the capacity of each of them is at most proportional to $\log n$, since the total transmitted power is of the order of n. Hence, the bound in this case is independent of the number of degrees of freedom that are effectively available in the physical channel and depends only on the total number of transmitting and receiving antennas.

We now focus on $C_{A \setminus V}$. In this case the number of degrees of freedom effectively available in the physical channel, rather than the total number of antennas available for communication, is the bottleneck that provides the required bound. To show this, we study the physics of the wave propagation process. We start by noting that $C_{A \setminus V}$ is independent of the nodes in V, so their presence does not increase the information flow and the upper bound can be computed assuming

V to be empty. Thanks to the empty separation annulus V, the kernel of the propagation operator connecting transmitters and receivers does not have singularities due to receivers being arbitrarily close to transmitters, and we can study the degrees of freedom of such operator using a functional analysis approach. The result, formally derived in the next sections, is the following. Let $\mathbf{H}^{(A\backslash V)}$ be the matrix with entries h_{sd}, $s \in \mathscr{P} \cap D$, $d \in \mathscr{P} \cap (A \backslash V)$. Although $O(n)$ antennas are available in $A \backslash V$,

$$\mathrm{rank}\big(\mathbf{H}^{(A\backslash V)}\big) = O(\sqrt{n}\log n). \tag{2.8}$$

It then follows, by performing the same steps leading to (2.7) but summing only over the effective number of independent MISO channels given by (2.8), rather than over all the receiving nodes, that w.h.p.,

$$C_{A\backslash V} = O\big(\sqrt{n}\,(\log n)^2\big). \tag{2.9}$$

Combining (2.6), (2.7), and (2.9), we have, w.h.p.,

$$C_n = O\big(\sqrt{n}\,(\log n)^2\big).$$

The final result now follows immediately from (2.1): w.h.p.,

$$R(n) = O\left(\frac{(\log n)^2}{\sqrt{n}}\right). \tag{2.10}$$

We make the following remark. The geometric setting considered above is by now standard in the literature, but it is not the most general one for which our result holds. We could have considered any arbitrary distribution of nodes in the disc D and in the annulus A and any matching between the nodes in the two regions. The only constraint on the distribution of the nodes is either that the node closest to the boundary of the partition must be at fixed distance δ from it, or that the number of nodes violating this minimum distance constraint is at most of $O(\sqrt{n})$, so that their contribution to the information flow can be bounded by a power constraint argument as in (2.7) rather than by a degrees of freedom
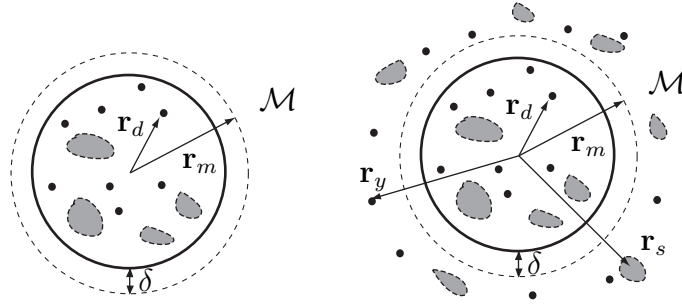
Figure 2.2: Left-hand side: step one, free space propagation. Right-hand side: step two, propagation with scattering elements.

argument.

2.3   The physics of the information flow

All that remains to be done is to provide a formal proof of (2.8). We do this in three steps. In a first step, we study the properties of the electromagnetic field that propagates up to distance $\delta > 0$ from the inner disc and is incident on the circumference $\mathcal{M}$, see the left-hand side of Figure 2.2, in which transmitting and receiving antennas are denoted by black dots, while scatterers are denoted in grey. In doing so, we assume to have an arbitrary collection of sources and scatterers placed inside the disc D, while the outside space is empty. Under these assumptions we show that the field incident on $\mathcal{M}$ is completely described by a linear combination of $O(\sqrt{n}\log n)$ basis functions. In other words, the number of degrees of freedom of the incident field is $O(\sqrt{n}\log n)$. In a second step, we consider the presence of scatterers outside the circle $\mathcal{M}$ and show that these do not change the number of degrees of freedom of the field incident on $\mathcal{M}$, see the right-hand side of Figure 2.2. The intuitive justification of this latter fact is that the field backscattered on $\mathcal{M}$ does not provide *new* information, since this has already passed through $\mathcal{M}$. Furthermore, we argue by the uniqueness theorem [15, page 100] that the field at any point outside $\mathcal{M}$, and in particular at the receiving antennas, is given by a linear transformation of the field on $\mathcal{M}$, which does not change the number of degrees of freedom. Finally, in a third step, we
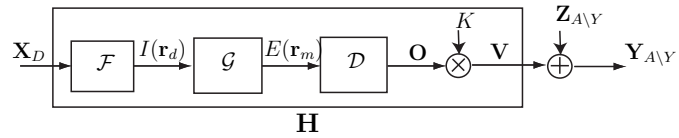
Figure 2.3: The physical channel model.

notice that receiving antennas detect a voltage proportional to the intensity of the field incident on them, plus some thermal noise, and this leads to the desired information-theoretic result.

The physical channel model is summarized in Figure 2.3, which shows the input-output relationship between transmitted and received signals. Such relationship is given by a chain of operators and corresponds to the information-theoretic channel model in (2.3). These operators are formally studied in the next sections according to the outline provided above. The figure shows that arbitrary source symbols represented by the input vector $\mathbf{X}_D$ are mapped into a physical current density $I(\mathbf{r}_d)$ inside the disc D through the operator $\mathcal{F}$. Next, the currents in D are related to the field $E(\mathbf{r}_m)$ on $\mathcal{M}$ though the free-space radiation operator $\mathcal{G}$. The operator $\mathcal{D}$ accounts for the presence of scatterers outside $\mathcal{M}$ and represents the mapping from the field on $\mathcal{M}$ to the vector $\mathbf{O}$ of the intensities of the electric field on the receiving antennas. Finally, the voltage at each receiving antenna is proportional to the intensity of the electric field incident on it, and the output symbol vector $\mathbf{Y}_{A \backslash Y}$ is given by the voltage on the antennas, plus some additive noise.

The proof outline described above can now be revisited in terms of the physical channel model depicted in Figure 2.3. We show that the range space of the operator $\mathcal{G}$ is of dimension $O(\sqrt{n} \log n)$, as n tends to infinity, and that the operator $\mathcal{D}$ is linear and thus does not increase the dimension of the space. Similarly, the linear map $\mathbf{V} = K\mathbf{O}$ does not increase the dimension of the space. The range-space of $\mathcal{F}$ can be assumed of arbitrary dimension, and we conclude that the range-space of $\mathbf{H}^{A \backslash V}$ is of dimension $O(\sqrt{n} \log n)$, as n tends to infinity.

2.3.1 Step one, propagation in free space

In this section, we show that the electric field at any point on $\mathcal{M}$ lies on a Hilbert space of dimension $O(\sqrt{n}\log n)$, as n tends to infinity. We assume sources and scatterers to be present in D, while the outside space is empty.

Being interested in an upper bound on the information flow, we can assume that the sources are arbitrarily located in D and can control the current density inside the disc D. We let such *arbitrary* current density be $I(\mathbf{r}_d)$ [A/m^2], $\mathbf{r}_d \in D$. Notice that singular sources can be thought of as limiting cases of two-dimensional distributions.

Assuming two-dimensional cylindrical propagation, so that the current density is $\hat{z}$ directed, the electric field radiated by currents in D and observed at $\mathbf{r}_m \in \mathcal{M}$ has only the $\hat{z}$ component, and is given by [15, pages 223-232]

$$E(\mathbf{r}_m) = \frac{-\beta^2}{4\omega\epsilon_0} \int_D I(\mathbf{r}_d) H_0^{(2)}\big(\beta|\mathbf{r}_m - \mathbf{r}_d|\big) ds, \quad \mathbf{r}_m \in \mathcal{M}, \tag{2.11}$$

where ds is an element of area perpendicular to $\hat{z}$, $\beta = \frac{2\pi}{\lambda}$ is the wavenumber, ϵ_0 is the permittivity of the vacuum, $H_i^{(2)}(x)$ is the Hankel function of the second kind and order i, and a Fourier transform convention $\exp(j\omega t)$ has been adopted, $\omega = 2\pi/(\sqrt{\epsilon_0\mu_0}\lambda)$ being the angular frequency, and μ_0 being the permeability of the vacuum. Furthermore, we assume the following power constraint:

$$a \int_D |I(\mathbf{r}_d)|^2 ds \leq n\,P, \tag{2.12}$$

wherein a is a normalization constant and P is the individual power constraint of each source. This condition ensures that the power radiated by the sources is finite and linearly proportional to the number of sources in D. Equation (2.11) shows that the currents in D and the electric field on $\mathcal{M}$ are linearly related through the radiation operator $\mathcal{G}$ (whose kernel is the Green's function). It follows that (2.11) can be written as:

$$E(\mathbf{r}_m) = (\mathcal{G}I)(\mathbf{r}_m), \quad \mathbf{r}_m \in \mathcal{M}, \tag{2.13}$$

where

$$(\mathcal{G}I)(\mathbf{r}_m) = \frac{-\beta^2}{4\omega\epsilon_0} \int_D I(\mathbf{r}_d) H_0^{(2)}\big(\beta|\mathbf{r}_m - \mathbf{r}_d|\big) ds \qquad (2.14)$$

represents the radiation operator, which maps a current density in D into the electric field at a point $\mathbf{r}_m \in \mathcal{M}$. We can also see from (2.14) the reason why we have introduced a minimum separation $\delta > 0$ between the sources and the observation domain $\mathcal{M}$, as the kernel of the radiation operator is singular at $\mathbf{r}_m = \mathbf{r}_d$. By introducing a separation $\delta > 0$ between the sources and the observation domain we avoid singularities in the kernel of (2.14), obtaining a compact integral operator with analytic kernel.

Next, we study the analytical properties of the operator $\mathcal{G}$, and show that the range-space of such operator is practically finite when the dimension of the radiating system is large. In order to do so, we represent the electric field on $\mathcal{M}$ in terms of the Hilbert-Schmidt decomposition, that is the equivalent of the singular value decomposition for operators in the L^2 space, and show that the electric field in (2.11) is completely described by $O(\sqrt{n}\log n)$ singular functions, as $n \to \infty$. The Hilbert-Schmidt decomposition of (2.11) is given by

$$E(\mathbf{r}_m) = \sum_{k=-\infty}^{\infty} \sigma_k \, \langle I, v_k \rangle_{L^2} \, u_k(\mathbf{r}_m), \quad \mathbf{r}_m \in \mathcal{M}, \qquad (2.15)$$

where $\{\sigma_k\}$ are the singular values of the operator; u_k and v_k are the k-th left singular function and right singular function respectively, and $\langle a, b \rangle_{L^2} := \int a(\mathbf{r})b^*(\mathbf{r})d\mathbf{r}$ denotes the inner product between functions in L^2. In order to compute the singular values $\{\sigma_k\}$, it is convenient to choose the following set of orthonormal functions:

$$u_k(\mathbf{r}_m) = -\frac{H_k^{(2)}\big(2\pi(\sqrt{n} + \delta)\big) \, e^{jk\angle\mathbf{r}_m}}{\sqrt{2\pi}(\sqrt{n} + \delta)^{1/2}\big|H_k^{(2)}\big(2\pi(\sqrt{n} + \delta)\big)\big|}, \qquad (2.16)$$

where $\mathbf{r}_m \in \mathcal{M}$, and

$$v_k(\mathbf{r}_d) = \frac{J_k(\beta|\mathbf{r}_d|) \, e^{jk\angle\mathbf{r}_d}}{\sqrt{2\pi}\left(\int_0^{\sqrt{n}}\big|J_k(\beta r_d)\big|^2 r_d \, dr_d\right)^{1/2}}, \qquad (2.17)$$

where $\mathbf{r}_d \in D$, $J_k(x)$ is the Bessel function of the first kind and order k, and $|\mathbf{r}|$ and $\angle\mathbf{r}$ are the magnitude and the angular coordinate of the vector $\mathbf{r}$ respectively. Using the addition theorem for Hankel functions [15, page 232], we can write $H_0^{(2)}\left(\beta|\mathbf{r}_m - \mathbf{r}_d|\right)$ in terms of cylindrical wave functions referred to the origin, and (2.11) can be rewritten as

$$E(\mathbf{r}_m) = \frac{-\beta^2}{4\omega\epsilon_0} \int_D I(\mathbf{r}_d) \sum_{k=-\infty}^{\infty} J_k(\beta|\mathbf{r}_d|) \times$$
$$H_k^{(2)}(\beta|\mathbf{r}_m|)e^{jk\angle(\mathbf{r}_m - \mathbf{r}_d)}ds. \tag{2.18}$$

Comparing (2.15) and (2.18), and using (2.16) and (2.17), we immediately obtain that

$$\sigma_k = \frac{\pi\beta^2}{2\omega\epsilon_0} \left|H_k^{(2)}\left(2\pi(\sqrt{n} + \delta)\right)\right| (\sqrt{n} + \delta)^{1/2} \times$$
$$\left(\int_0^{\sqrt{n}} \left|J_k(\beta r_d)\right|^2 r_d \, dr_d\right)^{1/2}. \tag{2.19}$$

The integral in (2.19) can be solved using identity (5.54.2) in [13], yielding

$$\int_0^{\sqrt{n}} \left|J_k(\beta|\mathbf{r}_d|)\right|^2 r_d \, dr_d$$
$$= \frac{x^2}{2}\left[\left(J_k(\beta x)\right)^2 - J_{k-1}(\beta x)J_{k+1}(\beta x)\right]\Big|_0^{\sqrt{n}}. \tag{2.20}$$

Substituting (2.20) into (2.19) we obtain the following expression for the singular values of the operator $\mathcal{G}$:

$$\sigma_k = \sqrt{\frac{\mu_0}{\epsilon_0}}\frac{\sqrt{2}}{2}\pi^2\sqrt{n}\left(\sqrt{n} + \delta\right)^{1/2}\left|H_k^{(2)}\left(2\pi(\sqrt{n} + \delta)\right)\right| \times$$
$$\left(\left(J_k(2\pi\sqrt{n})\right)^2 - J_{k-1}(2\pi\sqrt{n})J_{k+1}(2\pi\sqrt{n})\right)^{1/2}. \tag{2.21}$$

The electric field incident on $\mathcal{M}$ lies on a Hilbert space whose dimension depends on the behavior of the singular values in (2.21) as a function of the index k. It turns out that these are approximately constant up to a *critical value* $k_c \approx$

σ_k for $\delta = 0.01$
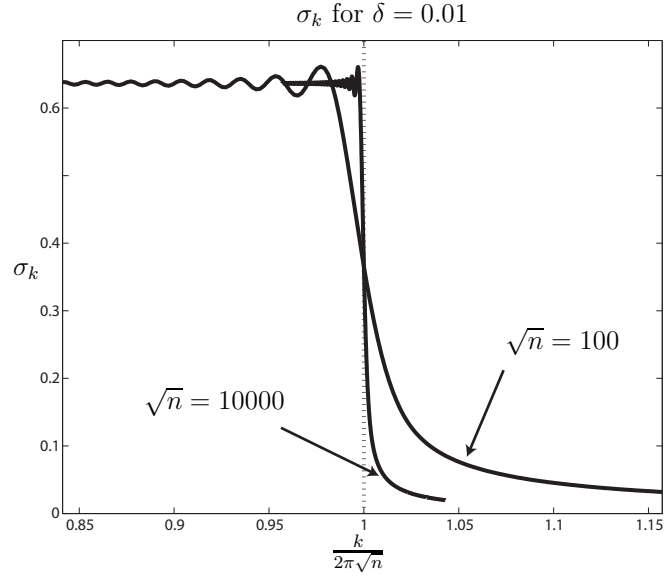
Figure 2.4: Plot showing the phase transition of the singular values σ_k.

$2\pi\sqrt{n}$, after which they undergo a phase transition and rapidly decay to zero. The transition tends to become a step function as $n \to \infty$, as shown in Figure 2.4. This leads to the conclusion that the electric field can be represented with a vanishing error using roughly k_c basis functions. This latter claim is made precise in the next theorem, proven in Appendix 2.6.1.

Theorem 2.3.1. *Let*

$$\widehat{E}_N(\mathbf{r}_m) = \sum_{k=-N}^{N} \sigma_k \, \langle I, v_k \rangle_{L^2} \, u_k(\mathbf{r}_m).$$

There exists an $N_0 = O(\sqrt{n}\log n)$, such that for all $\mathbf{r}_m \in \mathcal{M}$ we have

$$\lim_{n\to\infty} \|E(\mathbf{r}_m) - \widehat{E}_{N_0}(\mathbf{r}_m)\|^2 = 0. \tag{2.22}$$

Some remarks are now in order. The theorem shows that the electric field on $\mathcal{M}$ can be represented using $O(\sqrt{n}\log n)$ functions as n tends to infinity. The $\log n$ factor ensures that we are sufficiently far from the critical value k_c, so that the singular values corresponding to the tail of the field decomposition are essentially zero and the field can be reconstructed with vanishing error.

2.3.2 Step two, the presence of scatterers

In this section we show that the field outside $\mathcal{M}$ can be represented using $O(\sqrt{n}\log n)$ basis functions, even when scattering elements are present in the domain. This result has a simple physical interpretation in terms of an *information conservation principle*, which relies on the electromagnetic uniqueness theorem. The uniqueness theorem ensures that the electric field at any point outside $\mathcal{M}$ is uniquely determined by the field on $\mathcal{M}$. This is composed by the field radiated by D, which by Theorem 2.3.1 we know has a limited number of degrees of freedom, and by the field backscattered from outside $\mathcal{M}$, which does not provide any additional information since $\mathcal{M}$ is a closed curve capturing the whole information flow coming out of D. Next, we place this simple intuition into a more rigorous framework.

The electric field at any point $\mathbf{r}_y \in A \setminus V$, and in particular at the receiving antennas, is given by the superposition of two field vectors, denoted E_D and E_S, representing the field due to the currents inside and outside $\mathcal{M}$, respectively. Formally:

$$E(\mathbf{r}_y) = E_D(\mathbf{r}_y) + E_S(\mathbf{r}_y), \quad \mathbf{r}_y \in A \setminus V. \tag{2.23}$$

We show that both field vectors in (2.23) can be represented using $O(\sqrt{n}\log n)$ basis functions, as $n \to \infty$.

Let us first focus on E_D, i.e. the field vector due to the source currents and to the induced currents inside $\mathcal{M}$. The induced currents are due to the scattered field inside D, and also to the field backscattered from outside D. Since in the analysis of section 2.3.1 the current density in D was assumed arbitrary, the same analysis applies here, by including in $I(\mathbf{r}_d)$ the currents induced by the backscattered field. Thus, by the same steps leading to (2.15) we can now write the field E_D at a point $\mathbf{r}_y$ outside $\mathcal{M}$ as

$$E_D(\mathbf{r}_y) = \sum_{k=-\infty}^{\infty} \sigma_k \frac{H_k^{(2)}(\beta|\mathbf{r}_y|)}{H_k^{(2)}(2\pi(\sqrt{n}+\delta))} \langle I, v_k \rangle_{L^2} \, u_k(\mathbf{r}_y). \tag{2.24}$$

By (2.15) it follows that (2.24) also corresponds to the field due to the currents

inside D at the point $\mathbf{r}_m \in \mathcal{M}$ with $\angle\mathbf{r}_m = \angle\mathbf{r}_y$, having scaled each harmonic by the factor $H_k^{(2)}(\beta|\mathbf{r}_y|)/H_k^{(2)}(2\pi(\sqrt{n}+\delta))$. Then, using (2.13) we conclude that there exists a linear operator $\mathcal{D}_f$ such that

$$E_D(\mathbf{r}_y) = (\mathcal{D}_f \circ \mathcal{G}I)(\mathbf{r}_y), \quad \mathbf{r}_y \in A \setminus V. \tag{2.25}$$

We now focus on E_S, the field vector at the receiving antennas due to the currents induced on the scatterers outside $\mathcal{M}$. We show that E_S is linearly related to the field on the scatterers, and that this field is in turn linearly related to the currents inside D.

Let $\mathcal{S} \subseteq (A \setminus V)$ denote the domain occupied by the scattering elements outside $\mathcal{M}$, and let $I(\mathbf{r}_s)$ denote the induced current density on $\mathcal{S}$. The functional relationship between $I(\mathbf{r}_s)$ and E_S is given by (2.11), where we integrate over $\mathcal{S}$, in lieu of D. Thus,

$$E_S(\mathbf{r}_y) = \frac{-\beta^2}{4\omega\epsilon_0} \int_{\mathcal{S}} I(\mathbf{r}_s) H_0^{(2)}(\beta|(\mathbf{r}_y - \mathbf{r}_s)|)ds, \quad \mathbf{r}_y \in A \setminus V. \tag{2.26}$$

By Maxwell equations, we can write $I(\mathbf{r}_s)$ in terms of the electric field on $\mathcal{S}$ as follows:

$$I(\mathbf{r}_s) = j\omega(\epsilon(\mathbf{r}_s) - \epsilon_0)E(\mathbf{r}_s), \quad \mathbf{r}_s \in \mathcal{S}, \tag{2.27}$$

wherein $\epsilon(\mathbf{r}_s)$ is the permittivity of the dielectric material[2] at $\mathbf{r}_s$. Substituting (2.27) into (2.26) we obtain

$$E_S(\mathbf{r}_y) = \frac{-j\beta^2}{4\epsilon_0} \int_{\mathcal{S}} (\epsilon(\mathbf{r}_s) - \epsilon_0)E(\mathbf{r}_s) \times$$
$$H_0^{(2)}(\beta|(\mathbf{r}_y - \mathbf{r}_s)|)ds, \ \mathbf{r}_y \in A \setminus V, \tag{2.28}$$

which shows that E_S is linearly related to the field on $\mathcal{S}$.

Substituting (2.28) into (2.23) we obtain that the field on $\mathcal{S}$ is given by the

---

[2]The analysis in the case of metallic scatterers is completely equivalent.

solution of the following integral equation:

$$E(\mathbf{r}_s) = E_D(\mathbf{r}_s) + \frac{-j\beta^2}{4\epsilon_0} \int_{\mathcal{S}} (\epsilon(\mathbf{r}_s) - \epsilon_0)E(\mathbf{r}_s) \times$$
$$H_0^{(2)}(\beta|(\mathbf{r}_s - \mathbf{r}')|)ds', \mathbf{r}_s \in \mathcal{S}. \tag{2.29}$$

This is an inhomogeneous Fredholm integral equation of the second kind, whose solution leads to the Liouville-Neumann series. More important for us is that (2.29) shows a linear relationship between E_D and the field on $\mathcal{S}$. Since we have already shown in (2.28) that the field on $\mathcal{S}$ is linearly related to E_S, it now follows that E_S and E_D are also linearly related. Finally, by using (2.25) we conclude that there exists a linear operator $\mathcal{D}_s$, such that

$$E_S(\mathbf{r}_y) = (\mathcal{D}_s \circ \mathcal{G}I)(\mathbf{r}_y), \quad \mathbf{r}_y \in A \setminus V. \tag{2.30}$$

Putting things together, we conclude from (2.23), (2.25) and (2.30) that the electric field at the receiving antennas placed in $A \setminus V$ can be expressed as the superposition of two field vectors. Each of these lies in a Hilbert space whose dimension is limited by the rank of the radiation operator $\mathcal{G}$ and hence can be represented with $O(\sqrt{n}\log n)$ basis functions, as n tends to infinity.

2.3.3   Step three, back to information theory

The input-output relationship between the electromagnetic field at the output of each receiving antenna and the current densities in D can be expressed, in functional form, as:

$$E(\mathbf{r}_y) = (\mathcal{D} \circ \mathcal{G}I)(\mathbf{r}_y), \quad \mathbf{r}_y \in A \setminus V, \tag{2.31}$$

where $\mathcal{D} = \mathcal{D}_f + \mathcal{D}_s$. The values in (2.31) can be stack in a vector $\mathbf{O}$, whose d-th component indicates the intensity of the electric field at receiving node $d \in \mathscr{P} \cap (A \setminus V)$.

The voltage at each receiving antenna is proportional to the intensity of

the field at the antenna and is corrupted by some additive electric noise, which is assumed Gaussian and uncorrelated across antennas. Thus, the voltage values detected by the receiving nodes can be written as

$$\mathbf{Y}_{A\backslash V} = K\,\mathbf{O} + \mathbf{Z}_{A\backslash V}, \tag{2.32}$$

where K is a constant, and $\mathbf{Z}_{A\backslash V}$ is the Gaussian noise vector. Finally, the input-output relationship between symbols sent by nodes in $\mathscr{P}\cap D$ and received by nodes in $\mathscr{P}\cap(A\setminus V)$ at time i is given by

$$\mathbf{Y}_{A\backslash V}[i] = \mathbf{H}^{(A\backslash V)}\,\mathbf{X}_D[i] + \mathbf{Z}_{A\backslash V}[i]. \tag{2.33}$$

where $\mathbf{H}^{(A\backslash V)}$ is given by the composition of the linear operators $\mathcal{D}$, $\mathcal{G}$, $\mathcal{F}$, and the scalar K. It follows from the analysis in the previous section that the rank of $\mathbf{H}^{(A\backslash V)}$ is limited by the rank of $\mathcal{G}$. Thus, we obtain

$$\text{rank}\big(\mathbf{H}^{(A\backslash V)}\big) = O(\sqrt{n}\log n), \tag{2.34}$$

which proves (2.8).

2.4   Extension to three-dimensional networks

In this section we consider networks in which nodes are located according to a Poisson point process of unit density inside a sphere $\mathcal{B}_n$ of radius $(2n)^{1/3}$. As before, points of the Poisson process are paired uniformly at random. Assuming that each node generates at most P watts, we show that w.h.p. all nodes can (simultaneously) communicate to their intended destinations at rate

$$R(n) = O\left(\frac{(\log n)^3}{n^{1/3}}\right). \tag{2.35}$$

The proof follows the same steps as in the two-dimensional case, with some minor differences that we outline below. We partition $\mathcal{B}_n$ into two equal parts by drawing a spherical cut of radius $n^{1/3}$ at the origin, which divides $\mathcal{B}_n$ into the inner

sphere D and the outer spherical annulus A. Since an order of n communication requests have to cross the boundary of the partition, as before we first study the sum C_n of the rates that can be sent from the transmitters in D to the receivers in A, and then divide C_n by n to obtain the per-node rate R_n. We divide A into an inner and an outer part, denoted by V and $A \setminus V$ respectively, by drawing a sphere of radius $n^{1/3} + \delta$. The total information flow from D to A is decomposed into two contributions:

$$C_n \leq C_V + C_{A \setminus V}, \tag{2.36}$$

wherein C_V and $C_{A \setminus V}$ represent the information flow from D to V and from D to $A \setminus V$, respectively.

By summing the capacities of the individual MISO channels between the nodes in D and each receiver in V, we have, w.h.p.,

$$C_V = O(n^{2/3} \log n). \tag{2.37}$$

On the other hand, $C_{A \setminus V}$ is limited by the number of spatial degrees of freedom, which are $O(n^{2/3} (\log n)^2)$. As a consequence, we have that, w.h.p.,

$$C_{A \setminus V} = O(n^{2/3} (\log n)^3). \tag{2.38}$$

Combining (2.36), (2.37) and (2.38), and dividing by n, (2.35) follows.

As before, a proof of (2.38) is obtained by studying the physics of the information flow from D to $A \setminus V$. There are some geometrical differences that we outline below. Assume that the sources are arbitrarily located in D and can generate an arbitrary current density $I(\mathbf{r}_d)$ $[A/m^3]$, $\mathbf{r}_d \in D$, polarized in the $\hat{z}$ direction

The electric field radiated by the currents in D and observed on the surface

$\mathcal{M}$ separating A from $A \setminus V$ is given by

$$\mathbf{E}(\mathbf{r}_m) = -j\omega\mu\mathbf{A}(\mathbf{r}_m) + \frac{1}{j\omega\epsilon_0}\nabla(\nabla \cdot \mathbf{A}(\mathbf{r}_m)), \tag{2.39}$$

$$A_z(\mathbf{r}_m) = \frac{1}{4\pi}\int_D \frac{e^{-j\beta|\mathbf{r}_m-\mathbf{r}_d|}}{|\mathbf{r}_m - \mathbf{r}_d|}I(\mathbf{r}_d)d\mathbf{r}_d, \quad r_m \in \mathcal{M}, \tag{2.40}$$

wherein $\mathbf{A}$ denotes the magnetic vector potential [15], and A_z denotes its z component. The integral kernel in (2.40) can be decomposed into the sum of spherical harmonics [16, page 742], yielding

$$A_z(\mathbf{r}_m) = -j\beta\sum_{k=0}^{\infty}\sum_{i=-k}^{k} h_k^{(2)}(2\pi(n^{1/3} + \delta))\times$$

$$Y_{k,i}(\theta_m, \phi_m)\langle I, j_k Y_{k,i}\rangle_{L_2}, \tag{2.41}$$

wherein $\mathbf{r}_m \in \mathcal{M}$ has spherical coordinates $((n^{1/3} + \delta), \theta_m, \phi_m)$, j_k is the spherical Bessel function of the first kind and order k, $h_k^{(2)}$ is the spherical Hankel function of second kind and order k, and $Y_{k,i}$ is the (k-th, i-th) spherical harmonic function. The Hilbert-Schmidt decomposition of (2.41) can be written as:

$$A_z(\mathbf{r}_m) = \sum_{k=0}^{\infty}\sigma_k\sum_{i=-k}^{k}\langle I, v_{k,i}\rangle_{L^2} u_{k,i}(\mathbf{r}_m), \quad \mathbf{r}_m \in \mathcal{M}, \tag{2.42}$$

wherein, for $\mathbf{r}_m = ((n^{1/3} + \delta), \theta_m, \phi_m) \in \mathcal{M}$,

$$u_{k,i}(\mathbf{r}_m) = \frac{h_k^{(2)}(2\pi(n^{1/3} + \delta))\, Y_{k,i}(\theta_m, \phi_m)}{(n^{1/3} + \delta)\left|h_k^{(2)}(2\pi(n^{1/3} + \delta))\right|} \tag{2.43}$$

while, for $\mathbf{r}_d = (r_d, \theta_d, \phi_d) \in D$,

$$v_{k,i}(\mathbf{r}_d) = -j\,\frac{j_k(\beta r_d)\, Y_{k,i}(\theta_d, \phi_d)}{\left(\int_0^{n^{1/3}} |j_k(\beta r_d)|^2\, r_d^2\, dr_d\right)^{1/2}} \tag{2.44}$$

and

$$\sigma_k = \beta \ |h_k^{(2)}(2\pi(n^{1/3} + \delta))|(n^{1/3} + \delta) \times$$
$$\left(\int_0^{n^{1/3}} |j_k(\beta r_d)|^2 \, r_d^2 dr_d \right)^{1/2}. \tag{2.45}$$

Evaluating the integral in (2.45) using identity (5.54.2) of [13], and writing the spherical Bessel functions in terms of cylindrical Bessel functions of fractional order using identities in [13, par. 10.1.1], we obtain

$$\sigma_k = \frac{\pi\sqrt{\lambda}}{\sqrt{8}} n^{1/3} \sqrt{n^{1/3} + \delta} \ |H_{k+1/2}^{(2)}(2\pi(n^{1/3} + \delta))| \times$$
$$\left(\left(J_{k+1/2}(2\pi n^{1/3}) \right)^2 - J_{k-1/2}(2\pi n^{1/3}) J_{k+3/2}(2\pi n^{1/3}) \right)^{1/2}. \tag{2.46}$$

Let us compare (2.46) and (2.21). The two equations have the same asymptotic behavior, provided that in (2.21) we replace $n^{1/2}$ with $n^{1/3}$, *ceteris paribus*. By following exactly the same steps as in the proof of Theorem 1 and using (2.42), it then follows that there exists an $N_0 = O(n^{1/3} \log n)$, such that A_z can be represented with vanishing error as $n \to \infty$ using

$$\sum_{k=0}^{N_0} (1 + 2k) = O\left(n^{2/3} (\log n)^2 \right),$$

singular functions. We have assumed so far that the current density inside D was arbitrary, but polarized in the $\hat{z}$ direction. By symmetry, the analysis in the cases of polarization in the $\hat{x}$ and $\hat{y}$ directions is equivalent and, by the superposition of the effects, the general case of arbitrary polarization can lead up to a three-fold increase in the degrees of freedom. However, since an arbitrary electromagnetic field in an homogeneous source-free space can be obtained by superposition of Traverse Electric and Transverse Magnetic solutions, and since both of them can be represented in terms of spherical harmonics [15, pp. 129–131, pag. 267], the increase is only two-fold in case of arbitrary polarization. In any case, the order result $O\left(n^{2/3} (\log n)^2 \right)$ does not change in the case of arbitrary polarization. On the other hand, (2.39) shows that the electric field $\mathbf{E}$ and A_z are related through a linear

operator, so $\mathbf{E}$ can also be represented with a vanishing error using $O\left(n^{2/3}(\log n)^2\right)$ basis functions, as $n \to \infty$. Next, proceeding exactly as in section (2.3.2), it follows that the presence of scattering objects in $A \setminus V$ does not increase the number of degrees of freedom of the field at the receiving antennas. Finally, (2.38) is obtained as before, by applying the information-theoretic cut-set bound and assuming to beamform the total transmitted power in each of the $O\left(n^{2/3}(\log n)^2\right)$ spatial channels between transmitters and receivers.

2.5   Linear capacity scaling

The objective of the network engineer is to design wireless systems which fully exploit the number of degrees of freedom available for communication. With a successful design, and if the number of degrees of freedom scales linearly with the number of nodes, then more and more users can be added to the network without sacrificing performance and the engineer fulfills the dream of achieving linear capacity scaling. A recent paper of Özgür, Lévêque and Tse [28] almost fulfilled this dream. The authors assume a stochastic fading channel model in which all emitted signals are received with independent phases, which leads to a number of spatial degrees of freedom that scales linearly with the number of nodes. Then, they propose an ingenious node cooperation protocol which exploits these degrees of freedom, and allows to maintain an almost constant per-node bit rate as the network's size scales, when the path loss is sufficiently low.

However, we have shown that the number of spatial degrees of freedom cannot be assumed to grow linearly with the number of nodes, but in 2D it is limited by the spatial length of the cut that divides the network into two halves, so it can grow at most as $\sqrt{n}$; and in 3D it is limited by the surface of the cut, growing as $n^{2/3}$. Hence, space can be viewed as a *capacity bearing object* which poses a fundamental limit on the achievable information rate, independent of path loss assumptions. An intuitive picture of this is a as follows. Each communication channel can be viewed as occupying a unit of space along the cut through which the information must flow. Sharing this limited spatial resource among all the

nodes leads to our capacity bounds.

Given this limitation, we are led to the following engineering guideline: geometry should play a key role in the design of the network, hand in hand with protocol development. While previously proposed cooperation strategies are not tied to the geometric configuration and dimensionality of the network, with a careful geometric design the spatial resource can be carefully allocated to the users of the network, and then exploited by the communication protocol. For example, one could try to design *sparse networks* in which the number of nodes is small compared to the spatial resource available for communication and investigate whether this spatial resource can be exploited in practice through node cooperation. One such configurations could be a network in which the nodes are confined to a two-dimensional space, while propagation and scattering occurs in all three dimensions. We wish to investigate these issues in a forthcoming work, whose seeds are in [11], and shall not discuss them further here.

Looking in retrospective, we also see that the results reported in this work are of similar flavor as the ones obtained for point-to-point multiple antenna arrays in [18] [21] [33] [34], where physical arguments have been used to challenge the original optimistic results reported in the celebrated works of Foschini [8] and Telatar [35]. This challenge has also been supported by experimental evidence that the mutual coupling between antennas, arising when the spacing between them becomes smaller than the wavelength, does not allow independent signals to be detected at the receivers [7].

To bypass such arguments, it is customary to note that while the above can be of concern in antenna arrays where radiating elements are packed close to each other, in the context of nodes spatially distributed at random on the plane this issue is irrelevant, as nodes are typically in the far field of each other. For example, in a network operating at 3 GHz, the carrier wavelength is 0.1 m, while a reasonable separation distance between nearest neighbor nodes would be of the order of tens of meters, very much beyond the danger of incurring into near field coupling effects! Nevertheless, our results show that this heuristic argument fails in the scaling limit. By the uniqueness theorem, the field on the closed cut considered in our analysis

completely determines the signal measured at all the receivers outside the cut and such field has in 2D only an order of $\sqrt{n}$ degrees of freedom. Therefore, it is not possible to generate an order of n independent signals at the receivers, *even if* all the nodes are in the far field of each other. In other words, the degrees of freedom bottleneck is due to the flow through the cut, rather than to the spacing between the antennas.

As a final remark, we underline that the asymptotic results presented in this work cannot be directly applied to fixed size networks, for which capacity can be limited by numerous other factors. Position and properties of the different objects in the environment that are responsible for reflection, diffraction, scattering, and absorption of the propagating waves play an important role in determining the number of available spatial degrees of freedom, while the results presented here hold *uniformly* over all possible propagation environments, having fixed the dimensionality of the space, and in the limit of large networks. For this reason, the question of when the geometric limitations showed here become of practical relevance does not appear to have a unique answer. For small 2D networks, the number of available degrees of freedom in a rich scattering environment can be much larger than n, before eventually reaching its asymptotic $O(\sqrt{n})$ value as the network grows larger. In contrast, in an environment dominated by absorption the number of available degrees of freedom can be as small as zero, when communication is shaded by large absorbing obstacles.

To conclude, we are still far from reaching the holy grail of information theory for wireless communication, and the mathematical characterization of the capacity region of any fixed-size network remains *"a hope beyond the shadow of a dream"*.

Chapter 2, in part, is a reprint of the material as it appears in M. Franceschetti, M. D. Migliore, P. Minero, "The capacity of wireless networks: information-theoretic and physical limits," *IEEE Trans. on Information Theory*, vol. 55, no. 8, August 2009. The dissertation author was the primary investigator and author of this paper.

2.6   Appendix

2.6.1   Proof of Theorem 2.3.1

From (2.15) we have that, for any $\mathbf{r}_m \in \mathcal{M}$,

$$
\left\| E(\mathbf{r}_m) - \hat{E}_{N_0}(\mathbf{r}_m) \right\|^2
$$

$$
\leq \left\| \sum_{k=-\infty}^{-N_0} \sigma_k \left\langle I, v_k \right\rangle_{L^2} u_k(\mathbf{r}_m) \right\|^2
$$

$$
+ \left\| \sum_{k=N_0}^{\infty} \sigma_k \left\langle I, v_k \right\rangle_{L^2} u_k(\mathbf{r}_m) \right\|^2
$$

$$
\leq \sum_{k=-\infty}^{-N_0} |\sigma_k|^2 \left\langle I, I \right\rangle_{L^2} + \sum_{k=N_0}^{\infty} |\sigma_k|^2 \left\langle I, I \right\rangle_{L^2}
$$

$$
\leq \frac{2nP}{a} \sum_{k=N_0}^{\infty} \left| \sigma_k (\sqrt{n} + \delta) \right|^2, \tag{2.47}
$$

where the first inequality follows from the triangle inequality; the second inequality follows from the fact that u_k and v_k have unit norm and from the Cauchy-Schwarz inequality; the third inequality follows from $\sigma_k = \sigma_{-k}$ (due to the symmetry of Bessel functions of integer order) and the power constrain in (2.12). Thus, in order to prove the theorem, it suffices to show that there exists an $N_0 = O(\sqrt{n} \log n)$, such that

$$
\lim_{n \to \infty} n \sum_{k=N_0}^{\infty} \left| \sigma_k (\sqrt{n} + \delta) \right|^2 = 0. \tag{2.48}
$$

Using the recurrence formulas [1, identity 9.1.27] we can relate the Bessel functions of order $k-1$ and $k+1$ to the corresponding Bessel functions of order k, as follows:

$$
J_{k-1}(2\pi\sqrt{n}) = \frac{k}{2\pi\sqrt{n}} J_k(2\pi\sqrt{n}) + J_k'(2\pi\sqrt{n})
$$

$$
J_{k+1}(2\pi\sqrt{n}) = \frac{k}{2\pi\sqrt{n}} J_k(2\pi\sqrt{n}) - J_k'(2\pi\sqrt{n}),
$$

wherein $J'_k(x)$ denotes the derivative of the Bessel function with respect to the argument x. Thus,

$$J_{k-1}(2\pi\sqrt{n})J_{k+1}(2\pi\sqrt{n})$$
$$= \frac{k^2}{4\pi^2 n}\big(J_k(2\pi\sqrt{n})\big)^2 - \big(J'_k(2\pi\sqrt{n})\big)^2. \tag{2.49}$$

Substituting (2.49) into (2.21), the singular values can be written as:

$$\sigma_k(\sqrt{n}+\delta) = \frac{\sqrt{\mu_0 2n}}{2\sqrt{\epsilon_0}}\pi^2\big(\sqrt{n}+\delta\big)^{1/2}\left|H_k^{(2)}\big(2\pi(\sqrt{n}+\delta)\big)\right| \times$$
$$\left(\big(J'_k(2\pi\sqrt{n})\big)^2 - \big(k^2/(4\pi^2 n)-1\big)\big(J_k(2\pi\sqrt{n})\big)^2\right)^{1/2}, \tag{2.50}$$

where we emphasize the dependence of the singular values on the radius of the circle $\mathcal{M}$. Observe that $(k^2/(4\pi^2 n)-1) \geq 0$ for all $k \geq 2\pi\sqrt{n}$. Thus, from (2.50) it follows that, for $k \geq 2\pi\sqrt{n}$,

$$\left|\sigma_k(\sqrt{n}+\delta)\right|^2 = O\left(n^{3/2}\left|H_k^{(2)}\big(2\pi(\sqrt{n}+\delta)\big)\right|^2 \times \right.$$
$$\left. \left|J'_k(2\pi\sqrt{n})\right|^2\right), \tag{2.51}$$

as $n \to \infty$.

Next, we use Olver's uniform asymptotic expansions for Bessel functions [29] [30] to bound the right-hand side of (2.51). Notice that, while the Hankel function $|H_k^{(2)}(2\pi(\sqrt{n}+\delta))|$ is exponentially increasing in k, the derivative of the Bessel function $|J'_k(2\pi\sqrt{n})|$ is exponentially decreasing in k. In the following, by studying the rate of growth and decay of the two functions, we conclude that the singular values decrease exponentially to zero as k approaches infinity.

Let z denote the ratio between the argument and the order of $J'_k(2\pi\sqrt{n})$, i.e. $z = \frac{2\pi\sqrt{n}}{k}$. Identity (5.10) of [30] and the triangle inequality yield, for $0 < z \leq 1$,

$$|J'_k(kz)| \leq \frac{2}{k^{2/3}\,z}\left(\frac{1-z^2}{4\zeta(z)}\right)^{1/4}\left[\frac{\mathrm{Ai}\big(k^{2/3}\zeta(z)\big)}{k^{2/3}}\right.$$
$$\left. + \left|\mathrm{Ai}'\big(k^{2/3}\zeta(z)\big)\right| + |\eta(k,z)| + \frac{|\epsilon(k,z)|}{k^{2/3}}\right], \tag{2.52}$$

wherein Ai denotes the Airy function, for $0 < z \leq 1$ the function $\zeta(z)$ is defined as,

$$\frac{2}{3}\zeta^{3/2}(z) = \int_z^1 \frac{\sqrt{1-u^2}}{u}du$$
$$= \log\left(\frac{1+\sqrt{1-z^2}}{z}\right) - \sqrt{1-z^2}, \tag{2.53}$$

and $|\epsilon(k,z)|$ and $|\eta(k,z)|$ are subject to the following bounds [30, Section 5]:

$$|\epsilon(k,z)| \leq k^{-1}\text{Ai}\left(k^{2/3}\zeta(z)\right), \tag{2.54}$$

$$|\eta(k,z)| \leq k^{-1}\text{Ai}\left(k^{2/3}\zeta(z)\right). \tag{2.55}$$

Substituting (2.54) and (2.55) into (2.52), and using $\text{Ai}(x)/|\text{Ai}'(x)| \leq 2$, which holds for all $x \geq 0$ [30, page 11], we obtain that, for $0 < z \leq 1$,

$$|J_k'(kz)| \leq \frac{14}{k^{2/3} z}\left(\frac{1-z^2}{4\zeta(z)}\right)^{1/4}\left|\text{Ai}'(k^{2/3}\zeta(z))\right|. \tag{2.56}$$

Equation (2.56) provides a bound (uniform in $0 < z \leq 1$) for $|J_k'(kz)|$ in terms of the derivative of the Airy function. We now want to find a similar bound for the Hankel function. We start by noticing that

$$|H_k^{(2)}(x)| \leq |J_k(x)| + |Y_k(x)|, \tag{2.57}$$

wherein $Y_k(x)$ is the Bessel function of the second kind and order k. Let z_δ denote the ratio between the argument and the order of $H_k^{(2)}(2\pi(\sqrt{n} + \delta))$, i.e. $z_\delta = \frac{2\pi(\sqrt{n}+\delta)}{k} = z + \frac{2\pi\delta}{k}$. By identities (9.3.6) in [1], we have that, for $0 < z_\delta \leq 1$,

$$J_k(kz_\delta) = \left(\frac{4\zeta(z_\delta)}{1-z_\delta^2}\right)^{1/4}\left[\frac{\text{Ai}(k^{2/3}\zeta(z_\delta))}{k^{1/3}}\right.$$
$$\left. + \frac{e^{-2/3k\zeta^{3/2}(z_\delta)}}{1+k^{1/6}\zeta^{1/4}(z_\delta)}O\left(\frac{1}{k^{4/3}}\right)\right] \tag{2.58}$$

and

$$Y_k(kz_\delta) = -\left(\frac{4\zeta(z_\delta)}{1-z_\delta^2}\right)^{1/4}\left[\frac{\text{Bi}\big(k^{2/3}\zeta(z_\delta)\big)}{k^{1/3}}\right.$$

$$\left.+\frac{e^{+2/3k\zeta^{3/2}(z_\delta)}}{1+k^{1/6}\zeta^{1/4}(z_\delta)}O\left(\frac{1}{k^{4/3}}\right)\right]. \tag{2.59}$$

Thus, putting together (2.57), (2.58), and (2.59), we also have a bound (uniform in $0 < z_\delta \leq 1$) for the Hankel function in terms of the Airy functions Ai and Bi. The next step is to provide exponential bounds for the Airy functions.

We have, for $k^{2/3}\zeta \geq 1$ [31, page 394]:

$$\begin{aligned}
\text{Ai}(k^{2/3}\zeta) &\leq \frac{e^{-\frac{2}{3}k\zeta^{3/2}}}{k^{1/6}\zeta^{1/4}}, \\
|\text{Ai}'(k^{2/3}\zeta)| &\leq k^{1/6}\zeta^{1/4}e^{-\frac{2}{3}k\zeta^{3/2}}, \\
\text{Bi}(k^{2/3}\zeta) &\leq \frac{e^{+\frac{2}{3}k\zeta^{3/2}}}{k^{1/6}\zeta^{1/4}}.
\end{aligned} \tag{2.60}$$

By (2.53), we notice that $\zeta(z)$ is a decreasing function of z, which tends to infinity as $z \to 0^+$ and is 0 when $z = 1$. Hence, the condition $k^{2/3}\zeta\left(\frac{2\pi\sqrt{n}}{k}\right) \geq 1$, which is required for (2.60) to hold, is not satisfied when k is close to the critical value $2\pi\sqrt{n}$. However, by choosing $k \geq 2\pi\sqrt{n}\log n$ the desired condition holds for n large.

Thus, substituting (2.60) into (2.56), (2.58), and (2.59), it follows that, for $k \geq 2\pi\sqrt{n}\log n$,

$$\left|H_k^{(2)}\big(2\pi(\sqrt{n}+\delta)\big)\right| = O\left(\frac{1}{k^{1/2}(1-z_\delta^2)^{1/4}}e^{+\frac{2}{3}k\zeta^{3/2}(z_\delta)}\right), \tag{2.61}$$

$$\left|J_k'\big(2\pi(\sqrt{n})\big)\right| = O\left(\frac{(1-z^2)^{1/4}}{k^{1/2}\,z}e^{-\frac{2}{3}k\zeta^{3/2}(z)}\right), \tag{2.62}$$

as $n \to \infty$.

Notice that $\zeta(z_\delta) < \zeta(z)$, since $|z_\delta| > |z|$ for any $\delta > 0$. As a consequence, the rate of growth of the exponential in (2.61) is smaller than the rate of decay of the exponential in (2.62). Substituting (2.61) and (2.62) into (2.51), and using the

fact that $(1 - z^2)/(1 - z_\delta^2) = O(1)$ as $n \to \infty$, we obtain that for $k \geq 2\pi\sqrt{n}\log n$,

$$\left|\sigma_k(\sqrt{n} + \delta)\right|^2 = O\left(\sqrt{n}\exp\left\{-\frac{4}{3}\,k\,\left[\zeta^{3/2}\left(\frac{2\pi\sqrt{n}}{k}\right)\right.\right.\right.$$
$$\left.\left.\left. -\zeta^{3/2}\left(\frac{2\pi\sqrt{n}}{k} + \frac{2\pi\delta}{k}\right)\right]\right\}\right) \tag{2.63}$$

as $n \to \infty$.

Let us focus on the exponent in the right-hand side of (2.63). By (2.53), we have

$$-2k\left[\frac{2}{3}\zeta^{3/2}\left(\frac{2\pi\sqrt{n}}{k}\right) - \frac{2}{3}\zeta^{3/2}\left(\frac{2\pi\sqrt{n}}{k} + \frac{2\pi\delta}{k}\right)\right]$$
$$= -2k\int_{\frac{2\pi\sqrt{n}}{k}}^{\frac{2\pi\sqrt{n}}{k} + \frac{2\pi\delta}{k}}\frac{\sqrt{1 - u^2}}{u}du$$
$$\leq -2k\int_{\frac{2\pi\sqrt{n}}{k}}^{\frac{2\pi\sqrt{n}}{k} + \frac{2\pi\delta}{k}}\left(\frac{1}{u} - 1\right)du$$
$$= -2k\log\left(1 + \frac{\delta}{\sqrt{n}}\right) + 4\pi\delta, \tag{2.64}$$

where the inequality follows from $\sqrt{1 - u^2} \geq 1 - u$, for all $u \in [0, 1]$. Substituting (2.64) into (2.63) it follows that, for all $\delta > 0$ and for all $k \geq 2\pi\sqrt{n}\log n$,

$$\left|\sigma_k(\sqrt{n} + \delta)\right|^2 = O\left(\sqrt{n}e^{-2\,k\,\log\left(1 + \frac{\delta}{\sqrt{n}}\right)}\right), \tag{2.65}$$

as $n \to \infty$.

Finally, to obtain (2.48) we choose $N_0 = \max\{\frac{2}{\delta}, 2\pi\}\sqrt{n}\log n$ and use the bound (2.65), which is uniform in $k \geq \max\{\frac{2}{\delta}, 2\pi\}\sqrt{n}\log n$. Hence, for the choice

of N_0 above, there exists a uniform constant C, such that as $n \to \infty$, we have

$$n \sum_{k=N_0}^{\infty} \left| \sigma_k(\sqrt{n} + \delta) \right|^2$$

$$= n \sum_{k=\max\{\frac{2}{\delta}, 2\pi\} \sqrt{n} \log n}^{\infty} \left| \sigma_k(\sqrt{n} + \delta) \right|^2$$

$$\leq n^{3/2} \sum_{k=\max\{\frac{2}{\delta}, 2\pi\} \sqrt{n} \log n}^{\infty} C e^{-2k \log\left(1 + \frac{\delta}{\sqrt{n}}\right)}$$

$$= n^{3/2} \frac{C \left(e^{-2\log\left(1 + \frac{\delta}{\sqrt{n}}\right)} \right)^{\max\{\frac{2}{\delta}, 2\pi\}\sqrt{n}\log n}}{1 - e^{-2\log\left(1 + \frac{\delta}{\sqrt{n}}\right)}}$$

$$\leq \left(\frac{n^{5/2}}{2\delta\sqrt{n} + \delta^2} + n^{3/2} \right) C \, e^{-\log n^4 \left[\frac{\sqrt{n}}{\delta} \log\left(1 + \frac{\delta}{\sqrt{n}}\right) \right]}$$

$$= o\left(\frac{1}{n^2} \right)$$

$$\to 0,$$

which concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

2.7 Bibliography

[1] M. Abramowitz and I. A. Stegun, "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", *9th ed. New York: Dover*, 1972.

[2] S. Aeron, V. Saligrama, "Wireless Ad hoc Networks: Strategies and Scaling Laws for the fixed SNR Regime", *IEEE Trans. Inform. Theory*, vol. IT-53, no. 6, pp. 2044-2059, June 2007.

[3] S. Ahmad, A. Jovičić and P. Viswanath, "Outer Bounds to the Capacity Region of Wireless Networks", *IEEE Trans. Inform. Theory*, vol. IT-52, no. 6, pp. 2770-2776, June 2006.

[4] O. M. Bucci and G. Franceschetti, "On the spatial bandwidth of scattered fields", *IEEE Trans. Antennas Propagat.*, vol AP-35, no. 12, pp. 1445-1455, Dec. 1987.

[5] O. M. Bucci and G. Franceschetti, "On the degrees of freedom of scattered

fields", *IEEE Trans. Antennas Propagat.*, vol. AP-37, no. 7,pp. 918-926, July 1989.

[6] T. Cover, J. Thomas. "Elements of information theory." *John Wiley & sons, 2006.*

[7] D. W. Browne, M. Manteghi, M. P. Fitz, and Y. Rahmat-Samii, "Experiments With Compact Antenna Arrays for MIMO Radio Communications", *IEEE Trans. Antennas Propagat.*, vol. AP-54, no. 11, pp. 3239–3250, Nov. 2006.

[8] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas.", *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[9] M. Franceschetti, "A note on Lévêque and Telatar's upper bound on the capacity of wireless ad-hoc networks", *IEEE Trans. Inform. Theory*, vol. IT-53, no. 9, pp. 3207–3211, Sept. 2007.

[10] M. Franceschetti, O. Dousse, D. N. C. Tse and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory", *IEEE Trans. Inform. Theory*, vol. IT-53, no. 3, pp. 1009–1018, Mar. 2007.

[11] M. Franceschetti, P. Minero, M. D. Migliore, "The degrees of freedom of wireless networks: information theoretic and physical limits." *Proceedings of the Allerton Conference on Communications Computing and Control*, Monticello, Illinois, Sept. 2008.

[12] R. Gowaikar, B. Hochwald and B. Hassibi, "Communication over a wireless network with random connections", *IEEE Trans. Inform. Theory*, vol. IT-52, no. 7, pp. 2857–2871, July 2006.

[13] I. S. Gradshteyn and I. M. Ryzhik, "Tables of Integrals, Series, and Products", *A. Jeffrey, Editor, Academic Press* (1994).

[14] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks", *IEEE Trans. Inform. Theory*, vol. IT-42, no. 2, pp. 388–404, Mar. 2000.

[15] R. F. Harrington, "Time-Harmonic Electromagnetic Fields", *New York: Mc-Graw Hill*, 1961.

[16] J.D. Jackson, "Classical Electrodynamics", 2nd Edition, (1962) John Wiley, New York.

[17] A. Jovičić, P. Viswanath and S. R. Kulkarni, "Upper Bounds to Transport Capacity of Wireless Networks", *IEEE Trans. Inform. Theory*, vol. IT-50, no. 11, pp. 2555–2565, Nov. 2004.

[18] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2542–2556, June 2007.

[19] S. R. Kulkarni and P. Viswanath, "A Deterministic Approach to Throughput Scaling in Wireless Networks", *IEEE Trans. Inform. Theory*, vol. IT-50, no. 11, pp. 1041-1049, June 2004.

[20] O. Lévêque and E. Telatar, "Information Theoretic Upper Bounds on the Capacity of Large, Extended Ad-Hoc Wireless Networks", *IEEE Trans. Inform. Theory*, vol. IT-51, no. 3, pp. 858-865, Mar. 2005.

[21] K. Liu, V. Raghavan, A. M. Sayeed, "Capacity Scaling and Spectral Efficiency in Wideband Correlated MIMO Channels", *IEEE Trans. Inform. Theory*, vol. IT-49, no. 10, pp. 2504-2526, Oct. 2003.

[22] M. D. Migliore, "On the role of the number of degrees of freedom of the field in MIMO channels", *IEEE Trans. Antennas Propagat.*, vol. AP-54, no. 2, pp. 620-628, Feb. 2006.

[23] D. A. B. Miller, "Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths", *Appl. Optics*, vol. 39, no. 11, pp. 1681-1699, Apr. 2000.

[24] U. Niesen, P. Gupta and D. Shah, "On capacity scaling in arbitrary wireless networks", *Proceedings of the Information Theory and Applications Workshop (ITA),* p. 5, University of California, San Diego, Febuary 2007

[25] U. Niesen, P. Gupta and D. Shah, "Capacity region of large wireless networks", *Proceedings of the Allerton Conference on Communication Computing and Control,* Monticello, Illinois, September 2008.

[26] A. Özgür, R. Johari, D. N. C. Tse and O. Lévêque, "Information Theoretic Operating Regimes of Large Wireless Networks" *Proceedings of the International Symposium on Information Theory (IEEE-ISIT)*, Toronto, Canada, 2008.

[27] A. Özgür, O. Lévêque, and E. Preissmann, "Scaling laws for one and two-dimensional random wireless networks in the low attenuation regime", *IEEE Trans. Inform. Theory*, vol. IT-53, no. 10, pp. 3573-3585, Oct. 2007.

[28] A. Özgür, O. Lévêque and D. N. C. Tse, "Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks", *IEEE Trans. Inform. Theory*, vol. IT-53, no. 10, pp. 3549-3572, Oct. 2007.

[29] F. W. J. Olver, "The asymptotic expansion of Bessel functions of large order", *Philos. Trans. Roy. Soc. London Ser. A*, 247 (1954), pp. 328-368.

[30] F. W. J. Olver, "Tables for Bessel Functions of Moderate or Large Orders", *National Physical Laboratory Mathematical Tables*, vol. 6, Department of Scientific and Industrial Research, (Her Majesty's Stationery Office), London, 1962.

[31] F. W. J. Olver, "Asymptotics and special functions", *Reprint.* AKP Classics. A K Peters, Ltd., Wellesley, MA, 1997.

[32] R. Piestum, and D. A. B. Miller, "Electromagnetic degrees of freedom of an optical system", *J. Optical Soc. of America*, A, vol. 17, no. 5, pp. 892-902, May 2000.

[33] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, "Degrees of freedom in multiple antenna channels: a signal space approach," *IEEE Trans. Inform. Theory*, vol. IT-51, no. 2, pp. 523-536, Feb. 2005.

[34] A. M. Sayeed, V. Raghavan, J. H. Kotecha, "Capacity of Space-Time Wireless Channels: A Physical Perspective", *in Proc. Information Theory Workshop (ITW '04)*, pp. 434-439, Oct. 2004.

[35] E. Telatar, "Capacity of Multi-Antenna Gaussian Channels", *European Trans. on Telecommun.*, vol. 10, no. 6, pp. 585-596, Nov. 1999.

[36] G. Toraldo di Francia, "Resolving power and information," *J. Optical Soc. America*, vol. 45, no. 7, pp. 497-501, July 1955.

[37] G. Toraldo di Francia. "Directivity, super-gain and information", *IRE Trans. Antennas Propagat.*, vol. AP-4, no. 3, pp. 473–478, July 1956.

[38] L.-L. Xie and P. R. Kumar, "A Network Information Theory for Wireless Communications: Scaling Laws and Optimal Operation", *IEEE Trans. Inform. Theory*, vol. IT-50, no. 5, pp. 748-767, May 2004.

[39] L.-L. Xie and P. R. Kumar, "On the Path-Loss Attenuation Regime for Positive Cost and Linear Scaling of Transport Capacity in Wireless Networks", *IEEE Trans. Inform. Theory*, vol. IT-52, no. 6, pp. 2313-2328, June 2006.

[40] F. Xue, L.-L. Xie and P. R. Kumar, "The Transport Capacity of Wireless Networks over Fading Channels", *IEEE Trans. Inform. Theory*, vol. IT-51, no. 3, pp. 834-847, Mar. 2005.

Chapter 3

An information-theoretic perspective to random access

In this chapter, we consider a random access system where each sender can be in two modes of operation, active or not active, and where the set of active users is available to a common receiver only. Active transmitters encode data into independent streams of information, a subset of which are decoded by the receiver, depending on the value of the collective interference. The main contribution is to present an information-theoretic formulation of the problem which allows us to characterize, with a guaranteed gap to optimality, the rates that can be achieved by different data streams.

Our results are articulated as follows. First, we exactly characterize the capacity region of a two-user system assuming a binary-expansion deterministic channel model. Second, we extend this result to a two-user additive white Gaussian noise channel, providing an approximate characterization within $\sqrt{3}/2$ bit of the actual capacity. Third, we focus on the *symmetric* scenario in which users are active with the same probability and subject to the same received power constraint, and study the maximum achievable expected sum-rate, or throughput, for *any* number of users. In this case, for the symmetric binary expansion deterministic channel (which is related to the packet collision model used in the networking literature), we show that a simple coding scheme which does *not* employ superposition coding achieves the system throughput. This result also shows that the performance of

slotted ALOHA systems can be improved by allowing encoding rate adaptation at the transmitters, achieving constant (rather than zero) throughput as the number of users tends to infinity. For the symmetric additive white Gaussian noise channel, we propose a scheme that is within one bit of the system throughput for any value of the underlying parameters.

3.1   Introduction

Random access is one of the most commonly used medium access control schemes for channel sharing by a number of transmitters. Despite decades of active research in the field, the theory of random access communication is far from complete. What has been notably pointed out by Gallager in his review paper more than two decades ago [12] is still largely true: on the one hand, information theory provides accurate models for the noise and for the interference caused by simultaneous transmissions, but it ignores random message arrivals at the transmitters; on the other hand, network oriented studies focus on the bursty nature of messages, but do not accurately describe the underlying physical channel model. As an example of the first approach, the classic results by Ahlswede [3] and Liao [15] provide a complete characterization of the set of rates that can be simultaneously achieved communicating over a discrete memoryless (DM) multiple access channel (MAC). But the coding scheme they develop assumes a fixed number of transmitters with continuous presence of data to send. As an example of the second approach, Abramson's classic collision model for the ALOHA network [2] assumes that packets are transmitted at random times and are encoded at a *fixed* rate, such that a packet collision occurs whenever two or more transmitters are simultaneously active. The gap between these two lines of research is notorious and well documented by Ephremides and Hajek in their survey article [10].

In this work, we try to bridge the divide between the two approaches described above. We present the analysis of a model which is information-theoretic in nature, but which also accounts for the random activity of users, as in models arising in the networking literature. We consider a crucial aspect of random ac-

cess, namely that the number of simultaneously transmitting users is unknown to the transmitters themselves. This uncertainty can lead to packet collisions, which occur whenever the underlying physical channel cannot support the transmission rates of all active users simultaneously. However, our viewpoint is that the random level of the interference created by the random set of transmitters can also be exploited opportunistically by allowing transmission of different data streams, each of which might be decoded or not, depending on the interference level at the receiver.

To be fair, the idea of transmitting information in layers in random access communication is not new; however an information-theoretic perspective of this layering idea was never exposed. Previously, Medard *et al.* [17] studied the performance of Gaussian superposition coding in a two-user additive white Gaussian noise (AWGN) system, but did not investigate the information-theoretic optimality of such a scheme. In the present work, we present coding schemes with guaranteed gaps to the information-theoretic capacity. We do so under different channel models, and also extending the treatment to networks with a large number of users. Interestingly, it turns out that in the symmetric case in which all users are subject to the same received power constraint and are active with the same probability, superposition coding is not needed to achieve up to one bit from the throughput of an AWGN system.

This chapter is organized in incremental steps, the first ones laying the foundation for the more complex scenarios. Initially, we consider a two-user random access system, in which each sender can be in two modes of operation, active or not active. The set of active users is available to the decoder only, and active users encode data into two streams: one high priority stream ensures that part of the transmitted information is always received reliably, while one low priority stream opportunistically takes advantage of the channel when the other user is not transmitting. Given this set-up, we consider two different channel models. First, we consider a binary-expansion deterministic (BD) channel model in which the input symbols are bits and the output is the binary sum of a shifted version of the codewords sent by the transmitters. This is a first-order approximation

of an AWGN channel in which the shift of each input sequence corresponds to the amount of path loss experienced by the communication link. In this case, we exactly characterize the capacity region and it turns out that senders need to simultaneously transmit both streams to achieve capacity. Second, we consider the AWGN channel and present a coding scheme that combines time-sharing and Gaussian superposition coding. This turns out to be within $\sqrt{3}/2$ bit from capacity. Furthermore, we also show that in the symmetric case in which both users are subject to the same received power constraint, superposition coding is not needed to achieve up to $\sqrt{3}/2$ bit from capacity.

Next, we consider an m-user random access system, in which active transmitters encode data into independent streams of information, a subset of which are decoded by a common receiver, depending on the value of the collective interference. We cast this communication problem into an equivalent information-theoretic network with multiple transmitters and receivers and we focus on the *symmetric* scenario in which users are active with the same probability p, independently of each other, and are subject to the same received power constraint, and we study the maximum achievable expected sum-rate —videlicet throughput. Given this set-up, we again consider the two channel models described above. First, we consider the BD channel model in the symmetric case in which all codewords are shifted by the same amount. In this setting, input and output symbols are bits, so that the receiver observes the binary sum of the codewords sent by the active transmitters. The possibility of decoding different messages in the event of multiple simultaneous transmissions depends on the rate at which the transmitted messages were encoded. Colliding codewords are correctly decoded when the sum of the rates at which they were encoded does not exceed one. This is a natural generalization of the classic packet collision model widely used in the networking literature, where packets are always encoded at rate one, so that transmissions are successful only when there is one active user. We present a simple coding scheme which does *not* employ superposition coding and which achieves the throughput. The coding scheme can be described as follows. When p is close to zero, active transmitters ignore the presence of potential interferers and transmit a stream of

data encoded at rate equal to one. By doing so, decoding at the receiver is successful if there is only one active user, and it fails otherwise. This is what happens in the classic slotted ALOHA protocol, for which a collision occurs whenever two or more users are simultaneously active in a given slot. In contrast, when p is close to one, the communication channel is well approximated by the standard m-user binary sum DM-MAC, for which the number of transmitters is fixed and equal to m. In this regime, active users transmit a stream of data encoded at rate equal to $\frac{1}{m}$, that is, each active user requests an equal fraction of the m-user binary sum DM-MAC sum-rate capacity. Any further increase in the per-user encoding rate would result in a collision. When p is not close to either of the two extreme values, based on the total number of users m and the access probability p, transmitters estimate the number of active users by solving a set of polynomial equations. If k is the inferred number, then transmitters send one stream of data encoded at rate $\frac{1}{k}$, that is, each user requests an equal fraction of the k-user binary sum DM-MAC sum-rate capacity. Interestingly, it turns out that the estimator needed to achieve the throughput is different from the maximum-likelihood estimator $\lfloor mp \rfloor$ for the number of active users. The analysis also shows that the performance of slotted ALOHA systems can be improved by allowing *encoding rate adaptation* at the transmitters. In fact, we show that the expected sum-rate of our proposed scheme tends to one as m tends to infinity. Hence, there is no loss due to packet collisions in the so called scaling limit of large networks. This is in striking contrast with the well known behavior of slotted ALOHA systems in which users cannot adjust the encoding rate, for which the expected sum-rate tends to zero as m tends to infinity. In practice, however, medium access schemes such as 802.11x typically use backoff mechanisms to effectively adapt the rates of the different users to the channel state. It is interesting to note that while these rate control strategies used in practice are similar to the information-theoretic optimum scheme described above for the case of equal received powers, practical receivers typically implement suboptimal decoding strategies, such as decoding one user while treating interference as noise.

Next, we consider the case of the m-user AWGN channel. For this channel,

we present a simple coding scheme which does not employ superposition coding and which achieves the throughput to within one bit — for any value of the underlying parameters. Perhaps not surprisingly, this coding scheme is very similar to the one described above for the case of the BD channel. In fact, the close connection between these two channel models has recently been exploited to solve capacity problems for AWGN networks through their deterministic model counterpart [5].

Finally, we wish to mention some additional related works. Extensions of ALOHA resorting to probabilistic models to explain when multiple packets can be decoded in the presence of other simultaneous transmissions appear in [13] and [19]. An information-theoretic model to study layered coding in a two-user AWGN-MAC with no channel state information (CSI) available to the transmitters was presented in a preliminary incarnation of this work [18]. The two-user BD channel has been studied in the adaptive capacity framework in [14] and in this work we also provide a direct comparison with that model. We also rely on the broadcast approach which has been pursued in [20], and [22] to study multiple access channels with no CSI available. A survey of the broadcast approach and its application to the analysis of multiple antenna systems appeared in [21], and we refer the reader to this work and to [6] for an overview of the method and for additional references. The DM-MAC with partial CSI was studied in [8] assuming two compressed descriptions of the state are available to the encoders.

The rest of the chapter is organized as follows. The next section formally defines the problem in the case of a two-user AWGN random access system. Section 3.5 presents the extension to of the m-user random access system assuming an additive channel model. Section 3.6 consider the case of a BD channel model, while section 3.7 deals with the AWGN channel. A discussion about practical considerations and limitations of our model concludes the chapter.
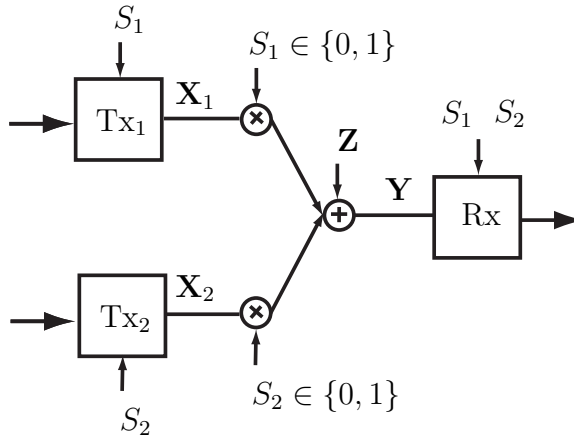
Figure 3.1: The two-user MAC with partial CSI modeling random access communications.

3.2 The two-user Additive Random Access Channel

Consider a two-user synchronous additive DM-MAC where each sender can be in two modes of operation, active or not active, independently of each other. The set of active users is available to the decoder, while encoders only know their own mode of operation. This problem is the compound DM-MAC with distributed state information depicted in Fig. 3.1. Specifically, the state of the channel is determined by two statistically independent binary random variables S_1 and S_2, which indicate whether user one and user two, respectively, are active, and it remains unchanged during the course of a transmission. Each sender knows its own state, while the receiver knows all the senders' states. The presence of side information allows each transmitter to adapt its coding scheme to its state component. We can assume without loss of generality that senders transmit a codeword only when active, otherwise they remain silent.

Each sender transmits several streams of data, which are modeled via independent information messages, a subset of which is decoded by the common receiver, depending on the state of the channel. The notation we use is as follows. We denote by $\mathcal{W}_1 = \{W_{1,1}, \ldots, W_{1,|\mathcal{W}_1|}\}$ and $\mathcal{W}_2 = \{W_{2,1}, \ldots, W_{2,|\mathcal{W}_2|}\}$ the ensem-

ble of independent messages transmitted by user 1 and user 2, respectively. We assume that each message $W_{i,j}$ is a random variable independent of everything else and uniformly distributed over a set with cardinality $2^{nR_{i,j}}$, for some non-negative rate $R_{i,j}$, $j \in \{1, \ldots, |\mathcal{W}_i|\}$, $i \in \{1, 2\}$. We let $\mathcal{W}_i(A) \subseteq \mathcal{W}_i$ denotes the set of messages transmitted by user i, $i \in \{1, 2\}$, that are decoded when the set of senders $A \subseteq \{1, 2\}$ is active. Finally $r_i(A)$ denotes the sum of the rates at which messages in $\mathcal{W}_i(A)$ are encoded.

Therefore, we can distinguish three non-trivial cases: if user 1 is the only active user, then the receiver decodes the messages in $\mathcal{W}_1(\{1\})$ and the transmission rate is equal to $r_1(\{1\})$; similarly, if user 2 is the only active user, then the receiver decodes the messages in $\mathcal{W}_2(\{2\})$, which are encoded at total rate of $r_2(\{2\})$; finally, the receiver decodes messages in $\mathcal{W}_1(\{1, 2\})$ and $\mathcal{W}_2(\{1, 2\})$ when both users are active, so senders communicate at rate $r_1(\{1, 2\})$ and $r_2(\{1, 2\})$, respectively. The resulting information-theoretic network is illustrated in Fig. 3.2, where one auxiliary receiver is introduced for each channel state component. In the illustration, the subscript index in $\mathbf{Y}$ and in $\mathbf{Z}$ denote the set of active users, $\mathcal{W}_1(\{1, 2\})$ and $\mathcal{W}_1(\{1, 2\})$ represent set of messages that are always decoded, while $\mathcal{W}_1(\{1\}) \setminus \mathcal{W}_1(\{1, 2\})$ and $\mathcal{W}_2(\{2\}) \setminus \mathcal{W}_2(\{1, 2\})$ denote messages that are decoded when there is no interference. It is clear from the figure that, upon transmission, each transmitter is connected to the receiver either through a point-to-point link or through an additive DM-MAC, depending on the channel state.

Observe that if the additive noises in Fig. 3.2 have the same marginal distribution, then the channel output sequence observed by the MAC receiver is a *degraded* version of the sequence observed by each of the two point-to-point receivers, because of the mutual interference between the transmitted codewords. As in a degraded broadcast channel, the "better" receiver can always decode the message intended for the "worse" receiver, similarly here each point-to-point receiver can decode what can be decoded at the MAC receiver. Thus, there is no loss of generality in assuming that

$$\mathcal{W}_1(\{1, 2\}) \subseteq \mathcal{W}_1(\{1\}) \tag{3.1}$$
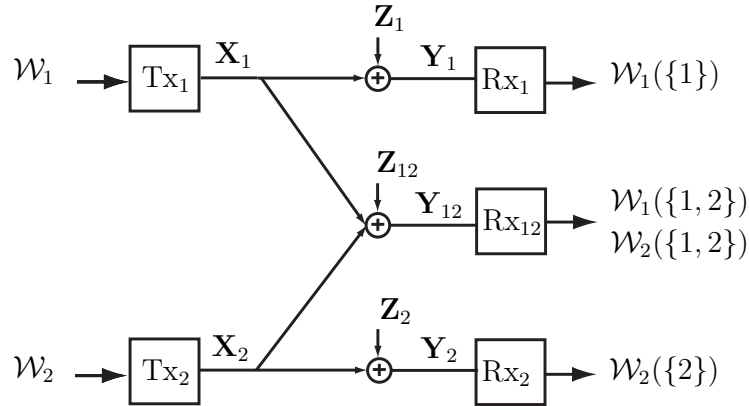
Figure 3.2: Network model for a two-user random access system.

and that

$$\mathcal{W}_2(\{1,2\}) \subseteq \mathcal{W}_2(\{2\}). \tag{3.2}$$

Then, messages in $\mathcal{W}_1(\{1,2\})$ and $\mathcal{W}_1(\{1,2\})$ ensure that some transmitted information is always received reliably, while the remaining messages provide additional information that can be opportunistically decoded when there is no interference. If conditions (3.1) and (3.2) are satisfied, then we say that $\mathscr{W} = (\{\mathcal{W}_1, \mathcal{W}_2\}, \{\mathcal{W}_1(\{1\}), \mathcal{W}_1(\{1,2\}), \mathcal{W}_2(\{2\}), \mathcal{W}_2(\{1,2\})\})$ denotes a *message structure* for the channel in Fig. 3.2.

For a given message structure $\mathscr{W}$, we say that the rate tuple $(r_1(\{1\}), r_2(\{2\}), r_1(\{1,2\}), r_2(\{1,2\}))$ is achievable if there exist a sequence of coding and decoding functions such that each receiver in Fig. 3.2 can decode all intended messages with arbitrarily small error probability as the coding block size tends to infinity. We define the capacity region $\mathscr{C}_{\mathscr{W}}$ as the closure of the set of achievable rate tuples.

Observe that as we vary $|\mathcal{W}_1|$, $|\mathcal{W}_2|$, and the sets of decoded messages, there are infinitely many possible message structures for a given channel. For each one of them we define $\mathscr{C}_{\mathscr{W}}$.

Next, we define the *capacity* of the channel in Fig. 3.2, denoted by $\mathscr{C}$, as the closure of the union of $\mathscr{C}_{\mathscr{W}}$ over all possible message structures $\mathscr{W}$. Note that $\mathscr{C}$ represents the optimal tradeoff among the rates $(r_1(\{1\}), r_2(\{2\}), r_1(\{1,2\}), r_2(\{1,2\}))$ over *all* possible ways of partitioning information into different informa-

tion messages such that conditions (3.1) and (3.2) are satisfied.

In the next section we answer the question of characterizing $\mathscr{C}$ for two additive channels of practical interest. First, we consider the BD channel model, for which we completely characterize the capacity region $\mathscr{C}$. Perhaps not surprisingly, we show that to achieve $\mathscr{C}$ it suffices that each sender transmits *two* independent information messages, one of which carries some reliable information which is always decoded, while the remaining one carries additional information which is decoded when the other user is not transmitting. Second, we consider the AWGN channel, for which we provide a constant gap characterization of $\mathscr{C}$, where the constant is universal and independent of the channel parameters. Finally, we apply this result to the study of the throughput of a two-user random access system under symmetry assumptions.

3.3 Example 1: the two-user BD random access channel

Suppose that channel input and output alphabets are each the set $\{0,1\}^{n_1}$, for some integer number n_1, and that at each time unit $t \in \{1, \ldots, n\}$ inputs and outputs are related as follows:

$$
\begin{aligned}
Y_{1,t} &= X_{1,t}, \\
Y_{12,t} &= X_{1,t} + S^{n_1-n_2} X_{2,t}, \\
Y_{2,t} &= S^{n_1-n_2} X_{2,t},
\end{aligned}
\tag{3.3}
$$

where $n_2 \leq n_1$ denotes an integer number, summation and product are over $\mathrm{GF}(2)$, and $S^{n_1-n_2}$ denotes the $(n_1 - n_2) \times (n_1 - n_2)$ shift matrix having the (i,j)th component equal to 1 if $i = j + (n_1 - n_2)$, and 0 otherwise. By pre-multiplying $X_{2,t}$ by $S^{n_1-n_2}$, the first n_2 components of $X_{2,t}$ are down-shifted by $(n_1 - n_2)$ positions and the remaining elements are set equal to zero. We refer to this model as the two-user BD *random access channel* (RAC), see Fig. 3.3 for a pictorial representation. Physically, this channel represents a first-order approximation of a wireless channel in which continuous signals are represented by their binary expansion, the
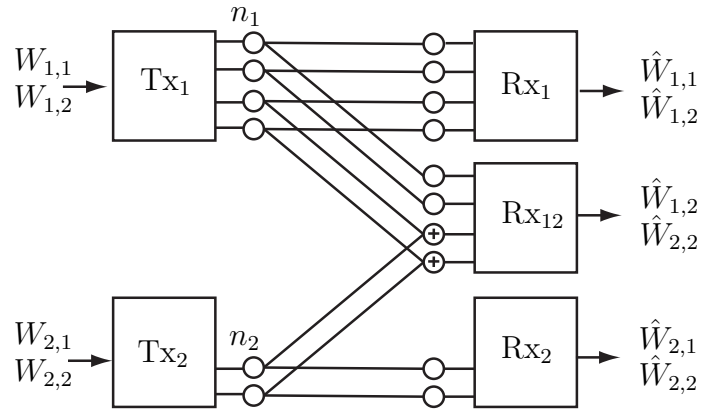
Figure 3.3: The two-user BD-RAC, and the message structure used to prove the achievability of the capacity region.

codeword length n_1 represents the noise cut-off value, and the amount of shift $n_1 - n_2$ corresponds to the path loss of user 2 relative to use 1 [5]. The following theorem characterizes the capacity region of this channel.

Theorem 3.3.1. *The capacity region $\mathscr{C}$ of the two-user BD-RAC is the set of non-negative rate tuples such that*

$$
\begin{aligned}
r_1(\{1\}) &\leq n_1, \\
r_2(\{2\}) &\leq n_2, \\
r_1(\{1\}) + r_2(\{1,2\}) &\leq n_1, \\
r_2(\{2\}) + r_1(\{1,2\}) &\leq n_1, \\
r_1(\{1,2\}) &\leq r_1(\{1\}), \\
r_2(\{1,2\}) &\leq r_2(\{2\}).
\end{aligned}
\tag{3.4}
$$

The proof of the converse part of the above theorem can be sketched as follows. Observe that the common receiver observing $\mathbf{Y}_{12} \triangleq \{Y_{12,1}, \ldots, Y_{12,n}\}$ can decode messages in $\mathcal{W}_2(\{1,2\})$. Let us suppose that this receiver is given messages in $\mathcal{W}_2 \setminus \mathcal{W}_2(\{1,2\})$ as side information. Then, it has full knowledge of $\mathcal{W}_2$, so it can compute the codeword $\mathbf{X}_2$ transmitted by user 2, subtract it from the aggregate received signal $\mathbf{Y}_{12}$, obtaining $\mathbf{X}_1$. Thus, given the side information, the channel output observed by the common receiver becomes statistically equivalent to $\mathbf{Y}_1$.

Since receiver 1 can decode $\mathcal{W}_1(\{1\})$ upon observing $\mathbf{Y}_1$, we conclude that receiver 12 must also be able to decode message $\mathcal{W}_2(\{1,2\})$. Hence, $r_1(\{1\}) + r_2(\{1,2\}) \leq n_1$. By providing side information about message $\mathcal{W}_1 \setminus \mathcal{W}_1(\{1,2\})$ and following the same argument above, we obtain that $r_2(\{2\}) + r_1(\{1,2\}) \leq n_1$. The remaining bounds are trivial.

The proof of the achievability part of the theorem shows that it suffices to partition information into *two* independent messages, such that $\mathcal{W}_1 = \{W_{1,1}, W_{1,2}\}$ and $\mathcal{W}_2 = \{W_{2,1}, W_{2,2}\}$. Messages $W_{1,2}$ and $W_{2,2}$ represent ensure that part of the transmitted information is always received reliably, while $W_{1,1}$ and $W_{2,1}$ are decoded opportunistically when one user is not transmitting. The corresponding message structure is illustrated in Fig. 3.3. In general, the coding scheme which we employ in the proof of the achievability requires that user 1 simultaneously transmits $W_{1,1}$ and $W_{1,2}$. However, in the special symmetric case in which $n_1 = n_2$ all rate tuples in the capacity region can be achieved by means of coding strategies in which each user transmits only one of the two messages.

Proof. First, we prove the converse part of the theorem. The first two inequalities which define $\mathscr{C}$ are standard point-to-point bounds which can be derived via standard techniques. To obtain the third inequality, observe that by Fano's inequality we have that $H(\mathcal{W}_1(\{1,2\})|\mathbf{Y}_{12}) \leq n\epsilon_n$, $H(\mathcal{W}_2(\{1,2\})|\mathbf{Y}_{12}) \leq n\epsilon_n$, as well as $H(\mathcal{W}_i(i)|\mathbf{Y}_i) \leq n\epsilon_n$, where ϵ_n tend to zero as the block length n tends to infinity. From the independence of the source messages, we have that

$$
\begin{aligned}
n(r_1(\{1\}) + r_2(\{1,2\})) &= H(\mathcal{W}_1(\{1\}), \mathcal{W}_2(\{1,2\})), \\
&= H(\mathcal{W}_1(\{1\}), \mathcal{W}_2(\{1,2\})|\mathcal{W}_2 \setminus \mathcal{W}_2(\{1,2\})), \\
&= I(\mathcal{W}_1(\{1\}), \mathcal{W}_2(\{1,2\}); \mathbf{Y}_{12}|\mathcal{W}_2 \setminus \mathcal{W}_2(\{1,2\})), \\
&\quad + H(\mathcal{W}_1(\{1\}), \mathcal{W}_2(\{1,2\})|\mathbf{Y}_{12}, \mathcal{W}_2 \setminus \mathcal{W}_2(\{1,2\})).
\end{aligned}
$$
$$(3.5)$$

Using the memoryless property of the channel and the fact that conditioning reduces the entropy, the first term in the right hand side of (3.5) can be upper

bounded as

$$I(\mathcal{W}_1(\{1\}), \mathcal{W}_2(\{1,2\}); \mathbf{Y}_{12}|\mathcal{W}_2 \setminus \mathcal{W}_2(\{1,2\})) \leq nn_1. \qquad (3.6)$$

On the other hand, from the chain rule, the fact that conditioning reduces the entropy, and Fano's inequality, we have that

$$
\begin{aligned}
H(\mathcal{W}_1(&\{1\}), \mathcal{W}_2(\{1,2\})|\mathbf{Y}_{12}, \mathcal{W}_2 \setminus \mathcal{W}_2(\{1,2\})) \\
&= H(\mathcal{W}_2(\{1,2\})|\mathbf{Y}_{12}, \mathcal{W}_2 \setminus \mathcal{W}_2(\{1,2\})) + H(\mathcal{W}_1(\{1\})|\mathbf{Y}_{12}, \mathcal{W}_2) \\
&\leq H(\mathcal{W}_2(\{1,2\})|\mathbf{Y}_{12}) + H(\mathcal{W}_1(\{1\})|\mathbf{Y}_{12}, \mathcal{W}_2, \mathbf{X}_2) \\
&= H(\mathcal{W}_2(\{1,2\})|\mathbf{Y}_{12}) + H(\mathcal{W}_1(\{1\})|\mathbf{Y}_1) \\
&\leq 2n\epsilon_n \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.7)
\end{aligned}
$$

where the last equality is obtained observing from (3.3) that, if $\mathbf{X}_2$ is given, then $\mathbf{Y}_{12}$ is statistically equivalent to $\mathbf{Y}_1$. Substituting (3.6) and (3.7) into (3.5), we have that

$$n(r_1(\{1\}) + r_2(\{1,2\})) \leq nn_1 + 2n\epsilon_n,$$

and the desired inequality is obtained in the limit of n going to infinity. The fourth inequality in (3.4) is obtained by a similar argument. Finally, the last two inequalities in (3.4) follow from (3.1) and (3.2).

Next, to prove the direct part of the theorem, we establish that $\mathscr{C}$ is equal to the capacity of the two-user BD-RAC for the specific message structure defined by $\mathcal{W}_i = \{W_{i,1}, W_{i,2}\}$, $\mathcal{W}_i(\{i\}) = \mathcal{W}_i$, and $\mathcal{W}_i(\{12\}) = \{W_{i,2}\}$, $i \in \{1,2\}$. For this message structure we have that

$$
\begin{aligned}
r_1(\{1\}) &= R_{1,2} + R_{1,1}, \\
r_2(\{2\}) &= R_{2,2} + R_{2,1}, \\
r_1(\{1,2\}) &= R_{1,2}, \\
r_2(\{1,2\}) &= R_{2,2}.
\end{aligned}
\qquad (3.8)
$$

We have established above that if $(r_1(\{1\}), r_2(\{2\}), r_1(\{1,2\}), r_2(\{1,2\})) \in \mathscr{C}_{\mathscr{W}} \subseteq \mathscr{C}$, then inequalities (3.4) have to be satisfied. Combining the non-negativity of the rates, (3.4), and (3.8), and eliminating $(r_1(\{1\}), r_2(\{2\}), r_1(\{1,2\}), r_2(\{1,2\}))$ from the resulting system of inequalities, we obtain

$$
\begin{aligned}
R_{1,1} + R_{1,2} &\leq n_1, \\
R_{2,1} + R_{2,2} &\leq n_2, \\
R_{1,1} + R_{1,2} + R_{2,2} &\leq n_1, \\
R_{2,1} + R_{1,2} + R_{2,2} &\leq n_1.
\end{aligned}
\tag{3.9}
$$

The above system of inequalities is the image of (3.4) under the linear map (3.8). Since the map is invertible, proving the achievability of all rate tuples $(r_1(\{1\}), r_2(\{2\}), r_1(\{1,2\}), r_2(\{1,2\}))$ satisfying (3.4) is equivalent to proving the achievability of all rate tuples $(R_{1,1}, R_{2,1}, R_{1,2}, R_{2,2})$ satisfying (3.9). It is tedious but simple to verify that the set of non-negative rate tuples satisfying (3.9) is equal to the convex hull of ten extreme points, four of which are dominated by one of the remaining six. Given two vectors $\mathbf{u}$ and $\mathbf{v}$, we say that $\mathbf{u}$ *dominates* $\mathbf{v}$ if each coordinate of $\mathbf{u}$ is greater than or equal to the corresponding coordinate of $\mathbf{v}$. The six dominant extreme points of (3.9) are given by $\mathbf{v}_1 = [n_2, n_2, n_1 - n_2, 0]^T$, $\mathbf{v}_2 = [n_1 - n_2, 0, 0, n_2]^T$, $\mathbf{v}_3 = [0, 0, n_1 - n_2, n_1]^T$, $\mathbf{v}_4 = [n_1, n_2, 0, 0]^T$, $\mathbf{v}_5 = [0, 0, n_1, 0]^T$, $\mathbf{v}_6 = [0, 0, 0, n_2]^T$, where the four coordinates denote $(R_{1,1}, R_{2,1}, R_{1,2}, R_{2,2})$, respectively.

The achievability of $\mathbf{v}_1, \ldots, \mathbf{v}_6$ can be sketched as follows. To achieve $\mathbf{v}_1$ sender 1 transmit simultaneously $W_{1,2}$ and $W_{1,1}$, in the first $n_1 - n_2$ and last n_2 components of X_1, respectively. User 2, instead, transmits $W_{2,1}$ in the first n_2 components of X_2. Because of the downshift in X_2, the multiple access decoder receives the binary sum of $W_{1,1}$ and $W_{2,1}$ in the last n_2 components of Y_{12}, and can successfully decoded $W_{1,2}$ from the first $n_1 - n_2$ interference-free components. Coding is performed so that $W_{1,1}$ and $W_{2,1}$ are received "aligned" at the common receiver, see Fig. 3.4 for a pictorial representation. Observe that in the special case in which $n_1 = n_2$, sender 1 only transmit message $W_{1,2}$. Likewise, $\mathbf{v}_2, \ldots, \mathbf{v}_6$ can be achieved by transmitting one message per user, in such a way that the transmitted codewords do not interfere with each other at the multiple access receiver. For
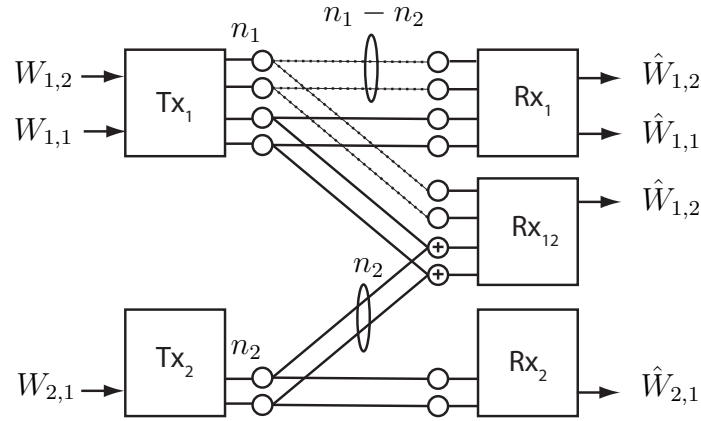
Figure 3.4: The coding scheme achieving the rate tuple $R_{1,1} = n_2$, $R_{1,2} = n_1 - n_2$, $R_{2,1} = n_2$, $R_{1,2} = 0$.

example, to achieve $\mathbf{v}_1$ user 2 transmits $W_{2,2}$ in the first n_2 components of X_2, while user 1 transmits $W_{1,2}$ in the first $n_1 - n_2$ components of X_1.

Next, observe that if an extreme point $\mathbf{v}$ is achievable, then all extreme points dominated by $\mathbf{v}$ are also achievable by simply decreasing the rate of some of the messages. Finally, any point in (3.4) can be written as convex combination of the extreme points, hence it can be achieved by time-sharing among the basic coding strategies which achieve $\mathbf{v}_1, \ldots, \mathbf{v}_6$. This shows that all rate tuples satisfying (3.9) are achievable. $\qquad\square$

3.3.1 The throughput in a symmetric scenario.

Having an exact characterization of the capacity region at hands, it is now possible to formulate and solve optimization problems of practical interests. As an example, we consider the problem of maximizing the throughput in the symmetric scenario where $n_1 = n_2 = 1$, and where each user is independently active with probability p.

This model represents a first-order approximation of a wireless channel in which data arrivals follow the same law, and where transmitted signals are received at the same power level. The codeword length is normalized to 1 so that the maximum amount of information which can be conveyed across the channel is

one bit per channel use, regardless the number of active users. The possibility of decoding different messages in the event of multiple simultaneous transmissions depends on the rate at which the messages were encoded. Colliding codewords are correctly decoded when the sum of the rates at which they were encoded does not exceed one. This is a natural generalization of the classic packet collision model widely used in the networking literature, where packets are always encoded at rate one, so that transmissions are successful only when there is one active user. The parameter p represents the burstiness of data arrivals, and determines the law of the variables S_1 and S_2 in Fig. 3.1, hence the channel law. Based on the knowledge of p, each sender can "guess" the state of operation of the other user, and optimize the choice of the encoding rates so that the expected sum-rate, or throughput, is maximized.

Formally, we look for the solution of the following optimization problem:

$$\max p(1 - p)\left[r_1(\{1\}) + r_2(\{2\})\right] + p^2\left[r_1(\{1, 2\}) + r_2(\{1, 2\})\right]$$

subject to the constraint that the rates should be in $\mathscr{C}$. Observe that the weight assigned to each rate component $r_i(A)$ is uniquely determined by p, and is equal to the probability that users in the set A are active. By means of Theorem 3.3.1, it is easy to show that the solution to the above problem is equal to

$$\begin{cases} 2p(1 - p), & \text{if } p \in (0, 1/2]; \\ p, & \text{if } p \in (1/2, 1]. \end{cases}$$

The coding strategy used to achieve the throughput can be described as follows. If the transmission probability p lies in the interval $(0, 1/2]$, then user i transmits message $W_{i,1}$ encoded at rate 1. A collision occurs in the event that both senders are simultaneously transmitting, which occurs with probability p^2, in which case the common receiver cannot decode the transmitted codewords. Decoding is successful if only one of the two users is active, so the expected sum-rate achieved by this scheme is equal to $2p(1 - p)$. If, instead, the transmission probability p lies in the interval $(1/2, 1]$, then user i transmits message $W_{i,2}$ encoded at rate $1/2$, i.e., at

half the sum-rate capacity of the two-user binary additive MAC. By doing so, the transmitted codewords are never affected by collisions, and can be decoded in any channel state. This yields an expected sum-rate of $2p(1-p)1/2 + p^2$. It should be highlighted that in this symmetric scenario each user transmits only one of the two messages for any value of p.

We show later that this optimization problem can be solved in the general case of a network with more than two users.

3.4   Example 2: the two-user AWGN-RAC

We now turn to another example of additive channels. Assume that at each discrete time step inputs and outputs are related as follows:

$$
\begin{aligned}
Y_{1,t} &= X_{1,t} + Z_{1,t}, \\
Y_{12,t} &= X_{1,t} + X_{2,t} + Z_{12,t}, \\
Y_{2,t} &= X_{2,t} + Z_{2,t},
\end{aligned}
\tag{3.10}
$$

where $Z_{1,t}$, $Z_{2,t}$, and $Z_{12,t}$ are independent standard Gaussian random variables, and the sum is over the field of real numbers. Assume that the realizations of $\{X_{i,t}\}$ satisfy the following average power constraint

$$
\sum_{t=1}^{n} x_{i,t}^2 \le n\mathrm{P}_i
$$

for some positive constant P_i, $i = 1, 2$, and that $\mathrm{P}_1 \ge \mathrm{P}_2$. We refer to the model in (3.10) as the two-user AWGN-RAC. In the rest of this chapter, we use the notation $\mathcal{C}(x) \triangleq 1/2 \log(1 + x)$.

An outer bound to the capacity region $\mathscr{C}$ of the two-user AWGN-RAC in (3.10) is given by the following Theorem.

Theorem 3.4.1. *Let $\overline{\mathscr{C}}$ denote the set of non-negative rates such that*

$$
\begin{aligned}
r_1(\{1\}) &\leq \mathcal{C}(\mathrm{P}_1), \\
r_2(\{2\}) &\leq \mathcal{C}(\mathrm{P}_2), \\
r_1(\{1\}) + r_2(\{1,2\}) &\leq \mathcal{C}(\mathrm{P}_1 + \mathrm{P}_2), \\
r_2(\{2\}) + r_1(\{1,2\}) &\leq \mathcal{C}(\mathrm{P}_1 + \mathrm{P}_2), \\
r_1(\{1,2\}) &\leq r_1(\{1\}), \\
r_2(\{1,2\}) &\leq r_2(\{2\}).
\end{aligned}
\tag{3.11}
$$

Then, $\mathscr{C} \subseteq \overline{\mathscr{C}}$.

The proof of the above theorem is similar to the converse part of Theorem 3.3.1 and it is hence omitted.

Next, we prove an achievability result by computing an inner bound to the capacity region $\mathscr{C}_{\mathscr{W}}$ of the two-user AWGN-RAC for a specific message structure $\mathscr{W}$. As for the BD-RAC, we let $\mathcal{W}_i = \{W_{i,1}, W_{i,2}\}$, $\mathcal{W}_i(i) = \mathcal{W}_i$, and $\mathcal{W}_i(12) = \{W_{i,2}\}$, $i \in \{1,2\}$. The encoding scheme we use is Gaussian superposition coding. Each sender encodes the messages using independent Gaussian codewords having sum-power less or equal to the power constraint. Decoding is performed using successive interference cancelation: messages are decoded in a prescribed decoding order, treating interference of messages which follow in the order as noise. Then, each decoded codeword is subtracted from the aggregate received signal.

Proposition 3.4.2. *Let $\underline{\mathscr{C}}'_{\mathscr{W}}$ denote the set of non-negative rates such that*

$$
\begin{aligned}
r_1(\{1\}) &\leq \mathcal{C}(\mathrm{P}_1), \\
r_2(\{2\}) &\leq \mathcal{C}(\mathrm{P}_2), \\
r_1(\{1\}) + r_2(\{2\}) &\leq \mathcal{C}(\mathrm{P}_1 + \mathrm{P}_2), \\
r_1(\{1,2\}) &\leq r_1(\{1\}), \\
r_2(\{1,2\}) &= r_2(\{2\}).
\end{aligned}
\tag{3.12}
$$

Similarly, let $\underline{\mathscr{C}}''_{\mathscr{W}}$ denote the set of non-negative rates satisfying (3.12) after after swapping the indices 1 and 2. Finally, let $\underline{\mathscr{C}}'''_{\mathscr{W}}$ denote the set of non-negative rates

satisfying the following inequalities

$$
\begin{aligned}
r_1(\{1,2\}) &\le \mathcal{C}\left(\frac{(1-\beta_1)\mathrm{P}_1}{\beta_1\mathrm{P}_1+\beta_2\mathrm{P}_2+1}\right), \\
r_2(\{1,2\}) &\le \mathcal{C}\left(\frac{(1-\beta_2)\mathrm{P}_2}{\beta_2\mathrm{P}_2+\beta_2\mathrm{P}_2+1}\right), \\
r_1(\{1,2\}) + r_2(\{1,2\}) &\le \mathcal{C}\left(\frac{(1-\beta_1)\mathrm{P}_1+(1-\beta_2)\mathrm{P}_2}{\beta_2\mathrm{P}_2+\beta_2\mathrm{P}_2+1}\right), \\
r_1(\{1\}) &\le r_1(\{1,2\}) + \mathcal{C}(\beta_1\mathrm{P}_1), \\
r_2(\{2\}) &\le r_2(\{1,2\}) + \mathcal{C}(\beta_2\mathrm{P}_2).
\end{aligned}
\tag{3.13}
$$

for some $(\beta_1, \beta_2) \in [0,1] \times [0,1]$. *Let* $\underline{\mathscr{C}}_{\mathscr{W}} = closure(\underline{\mathscr{C}}'_{\mathscr{W}} \cup \underline{\mathscr{C}}''_{\mathscr{W}} \cup \underline{\mathscr{C}}'''_{\mathscr{W}})$. *Then,* $\underline{\mathscr{C}}_{\mathscr{W}} \subseteq \mathscr{C}_{\mathscr{W}} \subseteq \mathscr{C}$.

Proof. Suppose that sender two does not transmit message $W_{2,1}$, i.e., $R_{2,1} = 0$. The achievability of $\underline{\mathscr{C}}'_2$ can then be shown by using a standard random coding argument as for the AWGN-MAC. To send $(W_{1,2}, W_{1,1})$, encoder one sends the sum of two independent Gaussian codewords having sum-power equal to P_1. On the other hand, sender two encodes $W_{2,2}$ into a Gaussian codeword having power P_2. A key observation is that the common receiver observing Y_{12} can decode all transmitted messages: $W_{1,2}$, $W_{2,2}$ can be decoded by assumption, while $W_{1,1}$ can be decoded after having subtracted $\mathbf{X}_2$ from the received channel output. Thus, by joint typical decoding, decoding is successful with arbitrarily small error probability if $R_{1,1}+R_{1,2}+R_{2,2} < \mathcal{C}(\mathrm{P}_1+\mathrm{P}_2)$, i.e., $r_1(\{1\})+r_2(\{2\}) < \mathcal{C}(\mathrm{P}_1+\mathrm{P}_2)$. Similarly, the receiver observing Y_1 can decode messages $W_{1,2}$, $W_{1,1}$ as long as $R_{1,1}+R_{1,2} < \mathcal{C}(\mathrm{P}_1)$, i.e., $r_1(\{1\}) < \mathcal{C}(\mathrm{P}_1)$ while the receiver observing Y_2 can decode messages $W_{2,2}$ if $r_2(\{2\}) \le \mathcal{C}(\mathrm{P}_2)$. We conclude that $\underline{\mathscr{C}}'_2$ is an inner bound to the capacity region. By swapping the role of user 1 and user 2 it is easy to see that $\underline{\mathscr{C}}''_2$ is also an inner bound to the capacity region. We claim that $\underline{\mathscr{C}}'''_2$ can be achieved by a coding scheme which combines Gaussian superposition coding and multiple access decoding. As in the Gaussian broadcast channel, to send the message pair $(W_{i,1}, W_{i,2})$, encoder i sends the codeword $\mathbf{X}_i(W_{i,1}, W_{i,2}) = \mathbf{U}_i(W_{i,2}) + \mathbf{V}_i(W_{i,1})$, where the sequences $\mathbf{U}_i$ and $\mathbf{V}_i$ are independent Gaussian codewords having power $(1-\beta_i)\mathrm{P}_i$ and $\beta_i\mathrm{P}_i$ respectively, $i = 1, 2$. Upon receiving Y_{12}, decoder 12 first decodes $W_{1,2}$ and $W_{2,2}$ using a MAC decoder and treating $\mathbf{V}_1(W_{1,1}) + \mathbf{V}_2(W_{2,1})$ as noise. Decoding is

successful with arbitrarily small error probability if

$$
\begin{aligned}
R_{2,2} &< \mathcal{C}\left(\frac{(1-\beta_1)\mathrm{P}_1}{\beta_1\mathrm{P}_1+\beta_2\mathrm{P}_2+1}\right), \\
R_{1,2} &< \mathcal{C}\left(\frac{(1-\beta_2)\mathrm{P}_2}{\beta_1\mathrm{P}_1+\beta_2\mathrm{P}_2+1}\right), \\
R_{1,2}+R_{2,2} &< \mathcal{C}\left(\frac{(1-\beta_1)\mathrm{P}_1+(1-\beta_2)\mathrm{P}_2}{\beta_1\mathrm{P}_1+\beta_2\mathrm{P}_2+1}\right).
\end{aligned}
\tag{3.14}
$$

Upon receiving $Y_i = \mathbf{U}_i(W_{i,2}) + \mathbf{V}(W_{i,1}) + \mathbf{Z}_i$, decoder i performs decoding via successive interference cancelation: it first decodes $W_{i,2}$ treating $\mathbf{V}_i(W_{i,1}) + \mathbf{Z}_i$ as noise, then it subtracts $\mathbf{U}_i(W_{i,2})$ from Y_i and decodes $W_{i,1}$ from $\mathbf{V}_i(W_{i,1}) + \mathbf{Z}_i$. Thus, decoding of $W_{i,2}$ is successful if $R_{i,2} < \mathcal{C}\left(\frac{(1-\beta_i)\mathrm{P}_i}{\beta_i\mathrm{P}_i+1}\right)$, while decoding of $W_{i,1}$ is successful if $R_{i,1} < \mathcal{C}(\beta_1\mathrm{P})$. After combining these conditions to the equalities which relate $(r_1(\{1\}), r_2(\{2\}), r_1(\{1,2\}), r_2(\{1,2\})$ to $(R_{1,1}, R_{2,1}, R_{1,2}, R_{2,2})$, and eliminating $(R_{1,1}, R_{2,1}, R_{1,2}, R_{2,2})$ from the resulting system of inequalities, we obtain that (3.14) have to be satisfied for the above coding scheme to work. Finally, a standard time-sharing argument can be used to show that the closure$(\underline{\mathscr{C}}_2' \cup \underline{\mathscr{C}}_2'' \cup \underline{\mathscr{C}}_2''') \subseteq \mathscr{C}_2$ $\qquad\square$

The following theorem explicitly characterizes the gap between the above inner and outer bounds on $\mathscr{C}$.

Theorem 3.4.3. *Let* $\mathbf{R} \in \overline{\mathscr{C}}$. *Then, there exists* $\mathbf{R}' \in \underline{\mathscr{C}}_{\mathscr{W}}$ *such that* $\| \mathbf{R} - \mathbf{R}' \| \leq \frac{\sqrt{3}}{2}$.

Proof. See Appendix 3.9.1. $\qquad\square$

Observe Gaussian superposition coding is *not* the optimal coding strategy for the AWGN channel under consideration. However, the above theorem ensures that Gaussian superposition coding achieves to within $\sqrt{3}/2$ bit from the capacity $\mathscr{C}$. It is important to note that this bound holds independently of the power constraints P_1 and P_2. The proof of the above theorem is established by showing that for any extreme point $\mathbf{v}$ of $\overline{\mathscr{C}}_2$, there exists an $\mathbf{r} \in \underline{\mathscr{C}}_2$ at distance less that $\sqrt{3}/2$ from $\mathbf{v}$. Since any point $\mathbf{R}$ in $\overline{\mathscr{C}}_2$ is a convex combination of extreme points of $\overline{\mathscr{C}}_2$, we can employ a time-sharing protocol among the various achievable rate points $\{\mathbf{r}\}$ and achieve a rate point at distance less that $\sqrt{3}/2$ from $\mathbf{R}$.

3.4.1 An approximate expression for the throughput.

As an application of the above result, consider the symmetric scenario where $P_1 = P_2 = P$, and where each user is active with probability p. Based on the knowledge of p, transmitters optimize the choice of the encoding rates so that the throughput is maximized. Formally, we look for the solution of the following optimization problem:

$$\max p(1-p)\left[r_1(\{1\}) + r_2(\{2\})\right] + p^2\left[r_1(\{1,2\}) + r_2(\{1,2\})\right]$$

subject to the constraint that the rates should be in $\mathscr{C}$. Combining Theorem 3.4.1 and Theorem 3.4.2, it is possible to show that the above maximum is equal to $T(p, P) + \varepsilon(p, P)$, where

$$T(p, P) = \begin{cases} 2p(1-p)\mathcal{C}(P), & \text{if } p \in (0, p_1(P)]; \\ p\mathcal{C}(2P), & \text{if } p \in (p_1(P), 1], \end{cases}$$

$p_1(P) = 1 - \mathcal{C}(2P)/(2\mathcal{C}(P)) \in (0, 1/2]$, and $0 \leq \varepsilon(p, P) \leq 1$. Observe that the bound on the error term holds for any choice of the parameters p and P.

The coding strategy used to achieve $T(p, P)$ is similar to the one described for the case of the symmetric BD channel. If the transmission probability p lies in the interval $(0, p_1(P)]$, then user i transmits message $W_{i,1}$ encoded at the maximum point-to-point coding rate, i.e., $\mathcal{C}(P)$. If, instead, the transmission probability p lies in the interval $(p_1(P), 1]$, then each active user transmits message $W_{i,2}$ encoded at rate $1/2\mathcal{C}(2P)$, i.e., at half the sum-rate capacity of the two-user AWGN-MAC. The parameter $p_1(P)$ represents a threshold value below which it is worth taking the risk of incurring in a packet collision. Observe that $p_1(P) \rightarrow 1/2$ as $P \rightarrow \infty$.

Fig. 3.5 compares $T(p, P)$ to the expected sum-rate achieved under the adaptive-rate framework [14], and to its counterpart assuming that full CSI is available to the transmitters. In the adaptive-rate framework, each sender transmits at a rate of $\mathcal{C}(2P)/2$, so that users can always be decoded. The figure illustrates how our approach allows us to improve upon the expected adaptive sum-rate for small values of p, for which the collision probability is small. In this regime, our inner
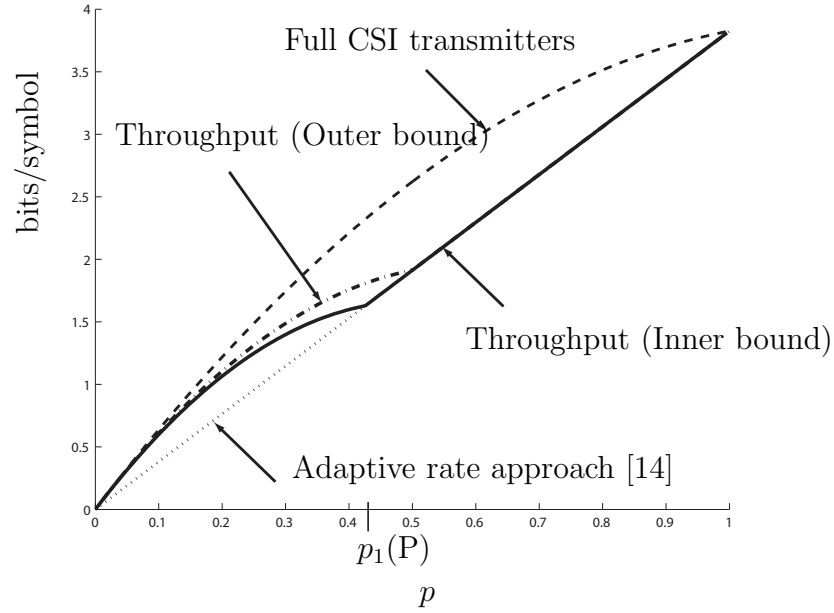
Figure 3.5: Throughput of the two-user symmetric AWGN-RAC (P = 20dB).

bound is in fact close to the curve obtained giving full CSI to the transmitters. Later in this chapter, we shall see that the gain provided by our approach becomes more significant when the population size of the network increases.

3.5   The m-user additive RAC

In this section we extend the analysis previously developed for a two-user system to the case of an m-user MAC, where m denotes an integer ≥ 2, and in which each transmitter can be in two modes of operation, active or not active. The set of active users, denoted in the sequel by A, determines the *state* of the channel. That is, the channel is said to be in state A if all users in the set A are active. As in the two-user case, transmitters only know their own state component, and encode data into independent streams of information. The common receiver knows the set of active users, and decodes subsets of the transmitted data streams depending on the state of the channel.

By introducing one auxiliary receiver per each channel state, we can map this problem to a broadcast network with m transmitters and $2^m - 1$ receivers. A

one-to-one correspondence exists between the set of receivers and the set of non-empty subsets of $\{1, \ldots, m\}$, so that for each set of active users A, there exists a unique corresponding receiver, which with abuse of notation we refer to as receiver A. Receiver A observes the sum of the codewords transmitted by users in A plus noise, and decodes a subset of the data streams sent by the active users. Observe that for a given channel state, only one among these auxiliary broadcast receivers corresponds to the actual physical receiver.

The formal description of the problem is as follows.

3.5.1 Problem formulation

Definition 3.5.1. *An m-user DM-RAC ($\{\mathcal{X}_1, \ldots, \mathcal{X}_m\}$, $\{\mathcal{Y}_A : A \subsetneq \{1, \cdots, m\}\}$, $(p(\{y_A : A \subsetneq \{1, \cdots, m\}\}|x_1, \ldots, x_m))$ consists of m input sets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, $2^m - 1$ output sets $\{\mathcal{Y}_A\}$, and a collection of conditional probabilities on the output sets.*

The channel is *additive* if at any discrete unit of time $t \in \{1, \ldots, n\}$, the input symbols $(X_{1,t}, \ldots, X_{m,t})$ are mapped into $2^m - 1$ channel output symbols $\{Y_{A,t}\}$ via the additive map

$$Y_{A,t} = \sum_{a \in A} X_{a,t} + Z_{A,t}, \tag{3.15}$$

where the $\{Z_{A,t}\}$ are mutually independent random variables with values in a set $\mathcal{Z}$, and the sum is over a field F such that there exists m embeddings $F_i : \mathcal{X}_i \to F$, and one embedding $F_{m+1} : \mathcal{Z} \to F$. In the next section we consider two classes of additive random access channels: the symmetric BD-RAC, for which the channel inputs are strings of bits, and the sum is binary; and the symmetric AWGN-RAC, for which $\mathcal{X} = \mathcal{Z} = \mathbb{R}$, the channel inputs are subject to an average power constraint, and the sum is over the reals.

Definition 3.5.2. *A* message structure *$\mathscr{W} = (\{\mathcal{W}_1, \ldots, \mathcal{W}_m\}, \{\mathcal{W}_i(A) : i \in A \subseteq \{1, \cdots, m\}\})$ for an m-user RAC consists of m input message sets $\mathcal{W}_i$, $\mathcal{W}_i = \{W_{i,1}, \cdots, W_{i,|\mathcal{W}_i|}\}$, and $m2^{m-1}$ output sets $\mathcal{W}_i(A)$, $\mathcal{W}_i(A) \subseteq \mathcal{W}_i$, such that the following condition is satisfied:*

A1. $\mathcal{W}_i(B) \subseteq \mathcal{W}_1(A)$ *for all* $i \in A \subseteq B \subseteq \{1, \ldots, m\}$.

For each i and $j \in \{1, \ldots, |\mathcal{W}_i|\}$, message $W_{i,j}$ is a random variable independent of everything else and uniformly distributed over a set with cardinality $2^{nR_{i,j}}$, for some non-negative rate $R_{i,j}$, $j \in \{1, \ldots, |\mathcal{W}_i|\}$.

The reason for imposing condition **A1.** is as follows. Observe from (3.15) that if $A \subseteq B$ and the marginal distributions of the noises Z_B and Z_A are equal, then Y_B is a (stochastically) degraded version of Y_A. Then, condition **A1.** says that the "better" receiver A *must* decode what can be decoded at the "worse" receiver B.

For a given message structure $\mathcal{W}$, let

$$r_i(A) = \sum_{j : W_{i,j} \in \mathcal{W}_i(A)} R_{i,j} \tag{3.16}$$

denote the sum of the rates of the messages in $\mathcal{W}_i(A)$. Observe that (3.16) defines a linear mapping from $\mathbb{R}_+^{|\mathcal{W}_1| \times \ldots \times |\mathcal{W}_m|}$ into $\mathbb{R}_+^{m 2^{m-1}}$ that shows how a *macroscopic* quantity, the rate at which user i communicates to receiver A, is related to various *microscopic* quantities, the coding rates of the individual transmitted messages.

Definition 3.5.3. *An n-code for the RAC* $(\{\mathcal{X}_1, \ldots, \mathcal{X}_m\}, \{\mathcal{Y}_A : A \subsetneq \{1, \cdots, m\}\},$ $(p(\{y_A : A \subsetneq \{1, \cdots, m\}\} | x_1, \ldots, x_m))$ *and for the message structure* $\mathcal{W}$ *consists of* m *encoding functions (encoders) and* $2^m - 1$ *decoding functions (decoders). Encoder* i *maps each* $\{W_{i,1}, \cdots, W_{i,|\mathcal{W}_i|}\}$ *into a random codeword* $\mathbf{X}_i \triangleq \{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\}$ *of n random variables with values in the set* $\mathcal{X}_i$. *Decoder* A *maps each channel output sequence* $\mathbf{Y}_A \in \mathcal{Y}_A^n$ *into a set of indexes* $\cup_{j : W_{i,j} \in \mathcal{W}_i(A)} \{\hat{W}_{i,j}\}$, *where each index* $\hat{W}_{i,j} \in \{1, \ldots, 2^{2nR_{i,j}}\}$ *is an estimate of the corresponding transmitted message* $W_{i,j} \in \mathcal{W}_i(A)$.

Definition 3.5.4. *For a given n-code, the average* probability *of decoding error at the decoder A is defined as*

$$Pr\left\{\hat{W}_{i,j} \neq W_{i,j} : W_{i,j} \in \mathcal{W}_i(A), j \in \{1, \ldots, |\mathcal{W}_i(A)|\}, i \in A\right\}. \tag{3.17}$$

Definition 3.5.5. *A rate tuple* $\{r_i(A)\}$ *is said to be* achievable *if there exists a sequence of n codes such that the average probability of a decoding error (3.17) for each decoder vanishes to zero as the block size n tends to infinity.*

Definition 3.5.6. *The* capacity region $\mathscr{C}_{\mathscr{W}}$ *of the m-user RAC* $(\{\mathcal{X}_1, \ldots, \mathcal{X}_m\}, \{\mathcal{Y}_A : A \subsetneq \{1, \cdots, m\}\}, (p(\{y_A : A \subsetneq \{1, \cdots, m\}\}|x_1, \ldots, x_m))$ *for the message structure $\mathscr{W}$ is closure of the set of achievable rate vectors* $\{r_i(A)\}$.

Finally,

Definition 3.5.7. *The capacity region $\mathscr{C}$ of the m-user RAC* $(\{\mathcal{X}_1, \ldots, \mathcal{X}_m\}, \{\mathcal{Y}_A : A \subsetneq \{1, \cdots, m\}\}, (p(\{y_A : A \subsetneq \{1, \cdots, m\}\}|x_1, \ldots, x_m))$ *is defined as*

$$\mathscr{C} = closure(\cup_{\mathscr{W}} \mathscr{C}_{\mathscr{W}}).$$

3.5.2 An outer bound to the capacity $\mathscr{C}$

Theorem 3.5.1. *The capacity region $\mathscr{C}$ of the additive m-user additive RAC in (3.15) is contained inside the set of non-negative rate tuples satisfying*

$$r_i(B) \leq r_i(A) \quad \textit{for all } i \in B \subseteq A, \tag{3.18}$$

and

$$\sum_{k=1}^{K} r_{i_k}(\{i_1 \ldots i_k\}) \leq I(X_{i_1}, \ldots, X_{i_K}; Y_{i_1 \ldots i_K}), \tag{3.19}$$

for all $K \in \{1, \ldots, m\}$ and $i_1 \neq \ldots \neq i_m \in \{1, \ldots, m\}$, and some joint distribution $p(q)p(x_1|q) \cdots p(x_m|q)$. Here the auxiliary random variable Q has the cardinality bound $|\mathcal{Q}| \leq e^{\Gamma(m+1,1)} - 1$, where $\Gamma(;)$ denote the incomplete Gamma function.

Proof. See Appendix 3.9.2. □

Remark 1: In the special case of a network with two users, it is immediate to verify that the outer bound given by the above theorem reduces to the region

given by Theorem 3.3.1 and Theorem 3.4.1 for the two-user BD-RAC and the two-user AWGN-RAC, respectively.

Remark 2: An inspection of the proof of the above theorem shows that the additive channel model assumed in the theorem can be replaced with a more general family of maps, namely with those channels with the property that, if $\mathbf{X}_{A'}$ is given, then $\mathbf{Y}_A$ is statistically equivalent to $\mathbf{Y}_{A \setminus A'}$, $A' \subseteq A$.

Remark 3: Observe that (3.19) gives $\binom{m}{K}K!$ inequalities for any value of $K \in \{1, \ldots, m\}$, so it defines $\sum_{K=1}^{m} \binom{m}{K}K! = e^{\Gamma(m+1,1)} - 1$ inequalities. It can be shown that $e^{\Gamma(m+1,1)} \to e^{m!}$ as $m \to \infty$.

Equation (3.19) can be obtained as follows. Suppose that we fix a set of active users $i_1, \ldots, i_K$, for some $K \in \{1, \ldots, m\}$, and we provide the receiver observing $\mathbf{Y}_{i_1 \ldots i_K}$ with messages in the set $\cup_{r=1}^{K} \mathcal{W}_{i_{K-r+1}} \setminus \mathcal{W}_{i_{K-r+1}}(\{i_1 \ldots i_{K-r+1}\})$ as side information. Suppose that this receiver decodes one user at the time, starting with user i_K and progressing down to user i_1. Let us consider the first decoding step. By assumption, receiver $\{i_1 \ldots i_K\}$ can decode information in $\mathcal{W}_{i_K}(\{i_1 \ldots i_K\})$ so, given the side information $\mathcal{W}_{i_K} \setminus \mathcal{W}_{i_K}(\{i_1 \ldots i_K\})$ it has full knowledge of $\mathcal{W}_{i_K}$, it can compute the codeword $\mathbf{X}_{i_K}$ transmitted by user i_K and subtract it from the aggregate received signal, obtaining $\mathbf{Y}_{i_1 \ldots i_K} - \mathbf{X}_{i_K} = \mathbf{Y}_{i_1 \ldots i_{K-1}}$. Thus, at the end of the first decoding step the channel output observed by receiver $\{i_1 \ldots i_K\}$ is statistically equivalent to $\mathbf{Y}_{i_1 \ldots i_{K-1}}$. It follows that at the next decoding step it can decode information in $\mathcal{W}_{i_{K-1}}(\{i_1 \ldots i_{K-1}\})$. By proceeding this way, at the rth iteration we obtain a sequence which is statistically equivalent to $\mathbf{Y}_{i_1 \ldots i_{K-r+1}}$. Hence, receiver $\{i_1 \ldots i_K\}$ can decode information in $\mathcal{W}_{i_{K-r+1}}(\{i_1 \ldots i_{K-r+1}\})$, then make use of the side information $\mathcal{W}_{K-r+1} \setminus \mathcal{W}_{i_{K-r+1}}(\{i_1 \ldots i_{K-r+1}\})$ to compute $\mathbf{X}_{i_{K-r+1}}$ and subtract it from the aggregate received signal before turning to decoding the next user. In other words, at the rth step of the iteration user i_{K-r+1}'s signal is only subject to interference from users $i_1, \ldots, i_{k-r}$, as the signal of the remaining users has already been canceled. Therefore, user i_{k-r+1} communicates to the receiver at a rate equal to $r_{i_{k-r+1}}(\{i_1 \ldots i_{k-r+1}\})$.

In summary, equation (3.19) says that the sum of the communication rates across the K iterations cannot exceed the mutual information between the chan-

nel inputs on the transmitters side and the channel output on the receiver side, regardless of the permutation on the set of users originally chosen.

3.5.3 The throughput of a RAC

Assume that each user is active with probability p, independently of other users, and that p is available to the encoders. In light of these assumptions,

Definition 3.5.8. *The maximum expected sum-rate, or* throughput*, of a RAC is defined as*

$$T(p,m) \triangleq \max \sum_{A \subseteq \{1,\ldots,m\}} p^{|A|}(1-p)^{m-|A|} \sum_{i \in A} r_i(A). \tag{3.20}$$

where the maximization is subject to the constraint that the rates should be in the capacity region $\mathscr{C}$ of that channel.

The fact that each user is active with the same probability p has one important consequence. By re-writing the objective function in (3.20) as

$$\sum_{k=1}^{m} p^k(1-p)^{m-k} \sum_{\substack{A \subseteq \{1,\ldots,m\} \\ |A|=k}} \sum_{i \in A} r_i(A)$$

and defining

$$\rho_k = \sum_{\substack{A \subseteq \{1,\ldots,m\} \\ |A|=k}} \sum_{i \in A} r_i(A), \quad k \in \{1,\ldots,m\}, \tag{3.21}$$

it is clear that the objective function in (3.20) depends only on $\rho_1,\ldots,\rho_m$. It follows that in order to compute $T(p,m)$ it is sufficient to characterize the optimal tradeoff among these m variables. This motivates the following definition

Definition 3.5.9. *Let $\mathscr{C}_{\boldsymbol{\rho}}$ denote the image of the capacity $\mathscr{C}$ of an m-user additive RAC under the linear transformation given by (3.21).*

It should be emphasized that the symmetry of the problem allow us to greatly reduce the complexity of the problem: instead of characterizing $\mathscr{C}$, which is a convex subset of $\mathbb{R}_+^{m2^{m-1}}$, it suffices to study the set $\mathscr{C}_{\boldsymbol{\rho}}$, which is a convex

subset of $\mathbb{R}_+^m$. Thus, we have that

$$T(p, m) = \max_{\rho_1, \ldots, \rho_m \in \mathscr{C}_{\boldsymbol{\rho}}} \sum_{k=1}^m p^k (1 - p)^{m-k} \rho_k. \qquad (3.22)$$

In the sequel, outer and inner bounds on $\mathscr{C}_{\boldsymbol{\rho}}$ are denoted by $\overline{\mathscr{C}}_{\boldsymbol{\rho}}$ and $\underline{\mathscr{C}}_{\boldsymbol{\rho}}$ respectively. In what follows, we denote by

$$f_{m,k}(p) \triangleq \binom{m}{k} p^k (1 - p)^{m-k}$$

the probability of getting exactly k successes in m independent trials with success probability p, and we denote by

$$F_{m,k}(p) \triangleq \sum_{i=0}^k f_{m,i}(p)$$

the probability of getting at most k successes.

3.6   Example 1: the m-user symmetric BD-RAC

In this section, we consider the m-user generalization of the symmetric BD-RAC considered in Section 3.3, where all transmitted codewords are shifted by the *same* amount. This model represents an approximation of a wireless channel in which signals are received at the same power level.

Suppose the $\mathcal{X}$ and $\mathcal{Y}$ alphabets are each the set $\{0, 1\}$, the additive channel (3.15) is noise-free, so $Z_A \equiv 0$, and the sum is over GF(2). Observe that this is the m-user version of the channel model in (3.3) in the special case where $n_1 = \ldots = n_m = 1$. The codeword length is normalized to 1. As mentioned above, this channel model can be thought of as a natural generalization of the packet collision model widely used in the networking literature, where packets are always encoded at rate one, so that transmissions are successful only when there is one active user. Theorem 3.5.1 yields the following proposition.

Proposition 3.6.1. *The capacity region $\mathscr{C}$ of the m-user symmetric BD-RAC is*

contained inside the set of $\{r_i(A)\}$ tuples satisfying

$$r_i(B) \leq r_i(A) \quad \text{for all } i \in B \subseteq A, \tag{3.23}$$

and

$$\sum_{k=1}^{m} r_{i_k}(\{i_1 \ldots i_k\}) \leq 1, \tag{3.24}$$

for all $i_1 \neq \ldots \neq i_m \in \{1, \ldots, m\}$.

3.6.1   The throughput of the symmetric BD-RAC

Next, we turn to the problem of characterizing the throughput $T(p, m)$ for the symmetric BD-RAC. The following theorem provides the exact characterization of $\mathscr{C}_{\boldsymbol{\rho}}$ for this channel.

Theorem 3.6.2. *$\mathscr{C}_{\boldsymbol{\rho}}$ for the m-user symmetric BD-RAC is equal to the $(\rho_1, \ldots, \rho_m)$ tuples satisfying*

$$\frac{\rho_k}{\binom{m}{k}} \geq \frac{\rho_{k+1}}{(k+1)\binom{m}{k+1}} \geq \ldots \geq \frac{\rho_m}{m\binom{m}{m}} \geq 0, \tag{3.25a}$$

and

$$\sum_{k=1}^{m} \frac{\rho_k}{k\binom{m}{k}} \leq 1. \tag{3.25b}$$

Proof. See Appendix 3.9.3. □

We outline the proof of the theorem as follows. The outer bound in the above theorem makes use of Proposition 3.6.1. To prove the achievability, we show that $\mathscr{C}_{\boldsymbol{\rho}}$ is equal to the image under the linear transformation given by (3.21) of the capacity region $\mathscr{C}_{\mathscr{W}}$ of the m-user symmetric BD-RAC for the message structure $\mathscr{W}$ defined by

$$\mathcal{W}_i = \{W_{i,1}, \ldots, W_{i,m}\}, \quad i \in \{1, \ldots, m\} \tag{3.26}$$

and

$$\mathcal{W}_i(A) = \cup_{j \geq |A|} W_{i,j}, \tag{3.27}$$

for $i \in A \subseteq \{1, \ldots, m\}$. This message structure is the natural generalization of the message structure used for the two-user BD-RAC. Each sender transmits m independent messages, which are ordered according to the amount of interference which they can tolerate, so that message $W_{i,j}$ is decoded when there are less than j interfering, regardless the identity of the interferers.

To prove the achievability of $\mathscr{C}_{\boldsymbol{\rho}}$ using this message structure, we observe that $\mathscr{C}_{\boldsymbol{\rho}}$ is the convex hull of m extreme points, and that to achieve the kth extreme points it suffices that user i transmits a *single* information message, namely $W_{i,k}$, encoded at rate $\frac{1}{k}$. Thus, a simple single-layer coding strategy can achieve all extreme points of $\mathscr{C}_{\boldsymbol{\rho}}$, and the proof of the achievability is completed by means of a time-sharing argument.

Having an exact characterization of $\mathscr{C}_{\boldsymbol{\rho}}$ at hands, we can explicitly solve the throughput optimization problem. The main result of this section is given by the following theorem.

Theorem 3.6.3. *Let* Π_m *represent the partition of the unit interval into the set of* m *intervals*

$$(p_0, p_1], (p_1, p_2], \ldots, (p_{m-1}, p_m],$$

where $p_0 \triangleq 0$, $p_m \triangleq 1$ *and, for* $0 < k < m$, p_k *is defined as the unique solution in* $(0, 1)$ *to the following polynomial equation in* p

$$\frac{1}{k+1} F_{m-1,k}(p) = \frac{1}{k} F_{m-1,k-1}(p). \tag{3.28}$$

Then, the following facts hold

1. $p_1 = \frac{1}{m}$, $p_{m-1} = \frac{1}{m^{1/(m-1)}}$, *and* $p \in (0, \frac{k}{m})$ *for* $k \in \{2, \ldots, m-2\}$.

2. *The throughput of the m-user symmetric BD-RAC is given by*

$$T(p, m) = \frac{mp}{k} F_{m-1,k}(p), \ \text{if } p \in (p_{k-1}, p_k], \tag{3.29}$$

for $k \in \{1, \dots, m\}$.

3. *$T(p, m)$ is achieved when all active senders transmit a* single *message encoded at rate*

$$r(p) = \frac{1}{k}, \;\; if \, p \in (p_{k-1}, p_k],$$ (3.30)

for $k \in \{1, \dots, m\}$.

4. *$T(p, m)$ is a continuous function of p; it is concave and strictly increasing in each interval of the partition Π_m.*

Proof. See Appendix 3.9.4. □

Remarks: The above theorem says that $T(p, m)$ can be achieved by a coding strategy which does not require simultaneous transmission of multiple messages. Instead, each active user transmits a single message encoded at rate $r(p)$. Inspection of (3.33) reveals that $r(p)$ is a piecewise constant function of p, whose value depends on the transmission probability p. If p is in the kth interval of the partition Π_m, then $r(p)$ is equal to $\frac{1}{k}$. Similarly, the corresponding achievable throughput $T(p, m)$ is a piecewise polynomial function of p. The boundary values of the partition, denoted by the sequence $\{p_k\}$, are given in semi-analytic form as solutions of (3.28), and closed form expressions are available only for some special values of m and k. Nevertheless, Theorem 3.7.3 provides the upper bound $p_k < \frac{k}{m}$.

The structure of the solution is amenable to the following intuitive interpretation. Based on the knowledge of m and p, transmitters estimate the number of active users. More precisely, if p is in the kth interval of the partition Π_m, i.e., $p_{k-1} < p \leq p_k$, then transmitters estimate that there are k active users. Since $p_k < \frac{k}{m}$, it is interesting to observe that the computed estimator is in general different from the maximum-likelihood estimator $\lfloor mp \rfloor$. Then, they encode their data at rate $\frac{1}{k}$, that is, each user requests an equal fraction of the k-user binary MAC sum-rate capacity. Clearly, there is a chance that the actual number of active users exceed k, in which case a collision occurs. Vice-versa, the scheme results in an inefficient use of the channel when the number of active users is less than

k. However, this strategy represents the right balance between the risk of packet collisions and inefficiency.

It is interesting to note that when $p \leq p_{k-1}$ the optimal strategy consists of encoding at rate 1, i.e., at the maximum rate supported by the channel. As already remarked, this is the coding strategy used in the classic ALOHA protocol. Notice that since $p_1 = \frac{1}{m}$, this strategy is optimal when the probability of being active is less that the inverse of the population size in the network. In this case, there is no advantage in exploiting the multi-user capability at the receiver. On the other hand, for $p > \frac{1}{m}$, the throughput of an ALOHA system is limited by packet collisions, which become more and more frequent as p increases. In this regime, the encoding rate has to decrease in order to accommodate the presence, which become more and more likely as p increases, of other potential active users.

3.6.2 Throughput scaling for increasing values of m

If we let the population size m grow while keeping p constant, the law of large number implies that the number of active users concentrates around mp, so one would expect that the uncertainty about the number of active users decreases as m increases. This intuition is confirmed by the following corollary, which states that the probability of collision tends to zero as m grows to infinity.

Corollary 3.6.4. *Let $p \in (0,1)$. Then, $\lim_{m \to \infty} T(p,m) = 1$.*

So far, we have been assuming that p does not depend on m. Assume now that the total packet arrival rate in the system is λ, and let $p = \frac{\lambda}{m}$ be the arrival probability at each transmitting node. Let $T(\lambda)$ denote the throughput in the limit $m \to \infty$. Then, by applying the law of rare events to (3.28) and (3.29) we obtain the following corollary to Theorem 3.6.3.

Corollary 3.6.5. *Let $\lambda_0 \triangleq 0$, $\lambda_\infty \triangleq \infty$ and, for $0 < k < \infty$, let λ_k be defined as the unique solution in $(0,\infty)$ to the following polynomial equation in λ*

$$\frac{1}{k+1}\Gamma(k+1,\lambda) = \Gamma(k,\lambda)$$
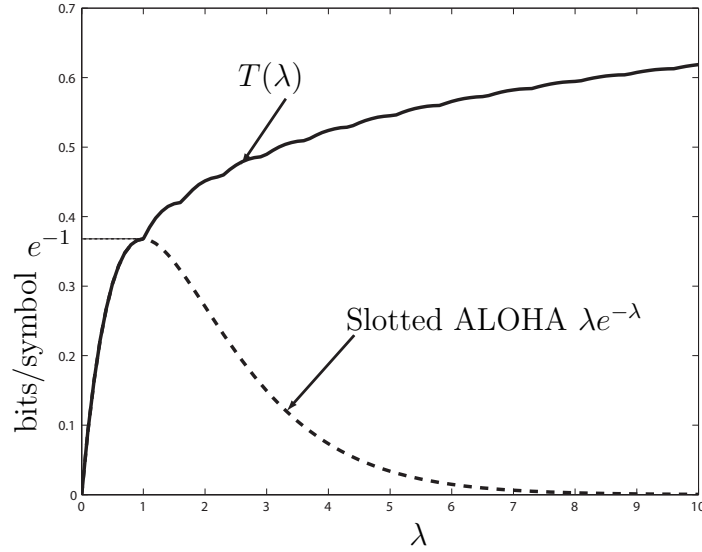
Figure 3.6: Comparison between $T(\lambda)$ and the throughput of the slotted ALOHA protocol.

where $\Gamma(k+1, \lambda)$ is the incomplete gamma function. Then, as m tends to infinity, the throughput is given by

$$T(\lambda) = \frac{\lambda}{k!k}\Gamma(k+1, \lambda), \ \ if \ \lambda \in (\lambda_{k-1}, \lambda_k],$$

for $k \in \mathbb{Z}$. The rate which attains the throughput is given by $r(\lambda) = \frac{1}{k}$, if $\lambda \in (\lambda_{k-1}, \lambda_k]$, $k \in \mathbb{Z}$. Finally, $T(\lambda)$ is a continuous function of λ; it is concave and strictly increasing in each interval $(\lambda_{k-1}, \lambda_k]$, and $\lim_{\lambda \to \infty} T(\lambda) = 1$.

Note that the claim above is in striking contrast with the throughput scaling of the classic slotted ALOHA protocol. The throughput of slotted ALOHA increases for small λ, it reaches a maximum e^{-1} at $\lambda = 1/m$, after which it decreases to zero as λ tends to infinity. See Fig. 3.6 for a comparison between $T(\lambda)$ and the throughput of standard ALOHA as a function of λ.

3.7 Example 2: the m-user symmetric AWGN-RAC

We now turn to another important example of additive channels. Suppose that the codewords generated by the m encoders are composed by n random variables taking values over the reals, and whose realizations satisfy the following average power constraint

$$\sum_{t=1}^{n} x_{i,t}^2 \leq n\mathrm{P}$$

for some positive constant P. Observe that we focus on the *symmetric* case in which all users are subject to the same received power constraint. Furthermore, suppose that $\{Z_A\}$ in (3.15) are independent standard Gaussian random variables, and that the sum in (3.15) is over the field of real numbers. Applying Theorem 3.5.1, we obtain the following proposition.

Proposition 3.7.1. *The capacity region $\mathscr{C}$ of the m-user symmetric AWGN-RAC is contained inside the set of $\{r_i(A)\}$ tuples satisfying*

$$r_i(B) \leq r_i(A) \quad \textit{for all } i \in B \subseteq A,$$

and

$$\sum_{k=1}^{K} r_{i_k}(\{i_1 \ldots i_k\}) \leq \mathcal{C}(K\mathrm{P}),$$

for all $K \in \{1, \ldots, m\}$ and $i_1 \neq \ldots \neq i_m \in \{1, \ldots, m\}$.

3.7.1 An approximate expression to within one bit for the throughput

Next, we turn to the problem of characterizing the throughput $T(p, m, \mathrm{P})$ for the symmetric AWGN-RAC as a function of the transmission probability p, the population size m, and the available power P. First, we provide inner and outer bounds on $\mathscr{C}_{\boldsymbol{\rho}}$ for this channel.

Theorem 3.7.2. *Let $\overline{\mathscr{C}}_{\boldsymbol{\rho}}$ denote the set of rates $\{\rho_k\} \in \mathbb{R}^m$ such that*

$$\frac{\rho_k}{k\binom{m}{k}} \geq \frac{\rho_{k+1}}{(k+1)\binom{m}{k+1}} \geq \ldots \geq \frac{\rho_m}{m\binom{m}{m}} \geq 0,$$

and

$$\sum_{k=1}^{K} \frac{\rho_k}{k\binom{m}{k}} \leq \mathcal{C}(K\mathrm{P}),$$

for all $K \in \{1, \ldots, m\}$. Let $\underline{\mathscr{C}}_{\boldsymbol{\rho}}$ denote the set of rates $\{\rho_k\} \in \mathbb{R}^m$ that satisfy (3.25a) and

$$\frac{1}{\mathcal{C}(\mathrm{P})} \frac{\rho_1}{\binom{m}{1}} + \sum_{k=2}^{m} \left(\frac{k}{\mathcal{C}(k\mathrm{P})} - \frac{k-1}{\mathcal{C}((k-1)\mathrm{P})} \right) \frac{\rho_k}{k\binom{m}{k}} \leq 1.$$

Then, $\underline{\mathscr{C}}_{\boldsymbol{\rho}} \subseteq \mathscr{C}_{\boldsymbol{\rho}} \subseteq \overline{\mathscr{C}}_{\boldsymbol{\rho}}$.

The proof of the above theorem is omitted since it closely follows the proof of Theorem 3.6.2. As for the case of the BD-RAC, the achievable region in the above theorem is obtained by considering the message structure defined by (3.26) and (3.27) and the coding scheme we utilize does *not* require the use of Gaussian superposition coding.

In virtue of Theorem 3.7.2 it is possible to bound $T(p, m)$ as

$$\underline{T}(p, m, \mathrm{P}) \leq T(p, m) \leq \overline{T}(p, m, \mathrm{P}),$$

where lower and upper bounds are given by (3.22) after replacing $\mathscr{C}_{\boldsymbol{\rho}, m}$ with $\underline{\mathscr{C}}_{\boldsymbol{\rho}, m}$ and $\overline{\mathscr{C}}_{\boldsymbol{\rho}, m}$ respectively. The following theorem provides an expression for $\underline{T}(p, m, \mathrm{P})$.

Theorem 3.7.3. *Let $\Pi_m(\mathrm{P})$ represent the partition of the unit interval into the set of m intervals*

$$(p_0(\mathrm{P}), p_1(\mathrm{P})], \ldots, (p_{m-1}(\mathrm{P}), p_m(\mathrm{P})],$$

where $p_0(\mathrm{P}) \triangleq 0$, $p_m(\mathrm{P}) \triangleq 1$ and, for $k \in \{1, \ldots, m-1\}$, $p_k(\mathrm{P})$ is defined as the

unique solution in $\left(0, \frac{k}{m}\right)$ *to the following polynomial equation in p*

$$\frac{\mathcal{C}((k+1)\mathrm{P})}{k+1}F_{m-1,k}(p) = \frac{\mathcal{C}(k\mathrm{P})}{k}F_{m-1,k-1}(p). \tag{3.31}$$

Then, $\underline{T}(p, m, \mathrm{P})$ *is a continuous function of p, concave, strictly increasing in each interval of the partition* $\Pi_m(\mathrm{P})$, *and is given by*

$$\underline{T}(p, m, \mathrm{P}) = \frac{\mathcal{C}(k\mathrm{P})}{k}mpF_{m-1,k-1}(p), \ \ if \ p \in (p_{k-1}(\mathrm{P}), p_k(\mathrm{P})], \tag{3.32}$$

for $k \in \{1, \dots, m\}$. *To achieve* $\underline{T}(p, m, \mathrm{P})$, *it suffices that each active user transmits a unique message encoded at rate*

$$r(p, m, \mathrm{P}) = \frac{\mathcal{C}(k\mathrm{P})}{k} \ \ if \ p \in (p_{k-1}(\mathrm{P}), p_k(\mathrm{P})], \tag{3.33}$$

for $k \in \{1, \dots, m\}$.

The proof of the above theorem is omitted since it closely follows the proof of Theorem 3.6.2. Similarly to what stated by Theorem 3.6.2 for the BD-RAC, the above theorem says that $\underline{T}(p, m, \mathrm{P})$ can be achieved by a coding strategy which does not require superposition coding: each active user transmits a single message encoded at rate $r(p, m, \mathrm{P})$. Both $r(p, m, \mathrm{P})$ and $\underline{T}(p, m, \mathrm{P})$ are piecewise constant function of p, whose value depends on the transmission probability p.

The coding scheme used to achieve $\underline{T}(p, m, \mathrm{P})$ for the symmetric AWGN-RAC is similar to the one used to achieve the throughput of the symmetric BD-RAC: based on the knowledge of m and P and p, transmitters estimate the number of active users. More precisely, if p is in the kth interval of the partition $\Pi_m(\mathrm{P})$, i.e., $p_{k-1}(\mathrm{P}) < p \le p_k(\mathrm{P})$, then transmitters estimate that there are k active users. Then, they encode their data at rate $\frac{1}{k}\mathcal{C}(k\mathrm{P})$, that is, each user requests an equal fraction of the k-user AWGN MAC sum-rate capacity.

A natural question to ask is how close this scheme is to the optimal performance. To answer this question, we first need to provide an expression for $\overline{T}(p, m, \mathrm{P})$. This is done in the next Theorem.

Theorem 3.7.4. *Let* Π_m *represent the partition of the unit interval into the set of* m *intervals*

$$(p_0, p_1], \ldots, (p_{m-1}, p_m],$$

where $p_0 \triangleq 0$, $p_m \triangleq 1$ *and, for every* $k \in \{1, \ldots, m\}$, p_k *is defined as the unique solution in* $\left(0, \frac{k}{m}\right)$ *to the following polynomial equation in* p

$$\frac{1}{k+1}F_{m-1,k}(p) = \frac{1}{k}F_{m-1,k-1}(p). \tag{3.34}$$

Then, $\overline{T}(p, m, \mathrm{P})$ *is a continuous function of* p*, concave and strictly increasing in each interval of the partition* $\Pi_m(\mathrm{P})$*, and is given by*

$$\overline{T}(p, m, \mathrm{P}) = mp \sum_{i=1}^{m} v_{k,i} F_{m-1,i-1}(p) \ \text{if } p \in (p_{k-1}, p_k], \tag{3.35}$$

for $k \in \{1, \ldots, m\}$*, where*

$$v_{1,i} = \begin{cases} 2\mathcal{C}(2\mathrm{P}) - \mathcal{C}(\mathrm{P}), & i = 1, \\ 2\mathcal{C}(i\mathrm{P}) - \mathcal{C}((i+1)\mathrm{P}) - 2\mathcal{C}((i-1)\mathrm{P}), & i \in \{2, \ldots, m\}, \\ \mathcal{C}(m\mathrm{P}) - \mathcal{C}((m-1)\mathrm{P}), & i = m, \end{cases} \tag{3.36}$$

For $k \in \{2, \ldots, m-2\}$

$$v_{k,i} = \begin{cases} 0, & i \in \{1, \ldots, k-1\}, \\ \frac{k+1}{k}\mathcal{C}(k\mathrm{P}) - \mathcal{C}((k+1)\mathrm{P}), & i = k, \\ 2\mathcal{C}(i\mathrm{P}) - \mathcal{C}((i+1)\mathrm{P}) - \mathcal{C}((i-1)\mathrm{P}), & i \in \{k+1, \ldots, m-1\}, \\ \mathcal{C}(m\mathrm{P}) - \mathcal{C}((m-1)\mathrm{P}), & i = m, \end{cases} \tag{3.37}$$

For $k = m-1$

$$v_{m-1,i} = \begin{cases} 0, & i \in \{1, \ldots, m-2\}, \\ \frac{m}{m-1}\mathcal{C}((m-1)\mathrm{P}) - \mathcal{C}(m\mathrm{P}), & i = m-1, \\ \mathcal{C}(m\mathrm{P}) - \mathcal{C}((m-1)\mathrm{P}), & i = m, \end{cases} \tag{3.38}$$

For k = m

$$v_{m,i} = \begin{cases} 0, & i \in \{1, \ldots, m-1\}, \\ \frac{1}{m}\mathcal{C}(m\mathrm{P}), & i = m. \end{cases} \tag{3.39}$$

Proof. See Appendix 3.9.5. □

The proof of the above theorem is conceptually simple but technical, as it requires finding the analytic solution of a linear program. Comparing the statements of Theorems 3.7.3 and 3.7.4, one can observe that the basic structure of $\overline{T}(p, m, \mathrm{P})$ and $\underline{T}(p, m, \mathrm{P})$ is the same. As opposed to the sequence $\{p_k(\mathrm{P})\}$ defined in Theorem 3.7.3, the sequence $\{p_k\}$ in Theorem 3.7.4 does not depend on the power P. It is easy to see that $p_k(\mathrm{P}) \leq p_k \leq k/m$, for every k. Furthermore, the sequence $\{p_k\}$ defined in Theorem 3.7.4 is equal to the sequence defined in Theorem 3.6.2. By directly comparing $\overline{T}(p, m, \mathrm{P})$ and $\underline{T}(p, m, \mathrm{P})$ we obtain the following result.

Theorem 3.7.5. *Let $p \in (0, 1]$, $m \geq 2$ and $\mathrm{P} > 0$. Then,*

$$\overline{T}(p, m, \mathrm{P}) - \underline{T}(p, m, \mathrm{P}) \leq 1.$$

Proof. See Appendix 3.9.6. □

The above theorem says that our suggested coding scheme achieves an expected sum-rate which is only 1 bit away from the optimum, independently of the values of p, P and m. It it remarkable that the gap does not increase with the population size of the system. Thus we conclude that transmitting at rate $\frac{1}{k}\mathcal{C}(k\mathrm{P})$ when p is in the kth interval of the partition $\Pi_m(\mathrm{P})$ represents the right balance between risk of collision and efficiency: encoding rates above $\frac{1}{k}\mathcal{C}(k\mathrm{P})$ would increase the collision probability, yielding a decrease in the expected sum-rate. Viceversa, rates lower than $\frac{1}{k}\mathcal{C}(k\mathrm{P})$ would result in an inefficient use of the channel.

Fig. 3.7 shows plots of $\overline{T}(p, m, \mathrm{P})$, $\underline{T}(p, m, \mathrm{P})$, and $r(p, m, \mathrm{P})$ for the case of networks with four users. Observe that the $\underline{T}(p, m, \mathrm{P})$ is a piecewise concave function of the transmission probability.
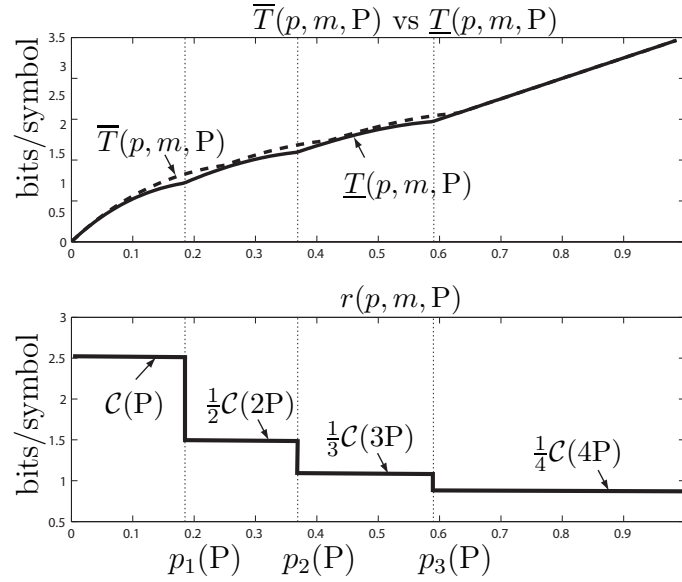
Figure 3.7: Bounds on the throughput of a four-user symmetric AWGN-RAC and encoding rate achieving the lower bound (P = 15 dB).

3.7.2 Comparison with other notions of capacity

The expression for the throughput derived in the previous section can be compared to similar expressions obtained assuming other notions of capacity. A natural outer bound is given by the throughput achieved assuming that full CSI is available to the transmitters. In this case, the sum-rate of the k-user AWGN-MAC can be achieved whenever k users are active. Averaging over the message arrival probability, we obtain the following expression for the throughput:

$$T_{CSI}(p, m, \mathrm{P}) \triangleq \sum_{k=1}^{m} f_{m,k}(p)\mathcal{C}(k\mathrm{P}). \tag{3.40}$$

On the other hand, if we study the symmetric AWGN-RAC following the adaptive capacity framework as in [14], then each transmitter designs a code which has to be decoded regardless the number of active users. This is a conservative viewpoint and forces each user to choose a rate of $1/m\mathcal{C}(m\mathrm{P})$ so that users can be decoded
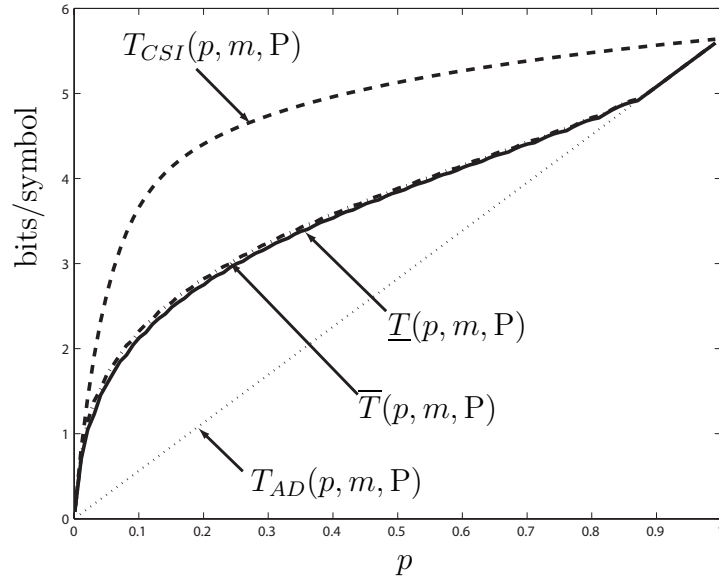
Figure 3.8: Throughput of the symmetric AWGN-RAC with $m = 25$ users ($P = 20$dB).

even when all m transmitters are active. Thus, we obtain

$$T_{AD}(p, m, P) \triangleq p\mathcal{C}(mP). \tag{3.41}$$

 Fig. 3.8 compares the obtained bounds on $T(p, m, P)$ for the case $m = 25$ and $P = 20$dB to the throughput under the adaptive-rate framework (3.41), and assuming full CSI available to the transmitters (3.40).

Finally, observe that in order to achieve $\underline{T}(p, m, P)$ transmitters have to estimate the number of active users by solving the polynomial equations (3.31). A natural question to ask is what is the achievable throughput performance if a maximum-likelihood estimator for the number of active user is used instead. Consider the following strategy. Suppose that, based on the knowledge of m and p, and assuming no prior on the number of active users, transmitters compute k_{ML}, the maximum-likelihood estimator for the number of active users, and encode their data at rate $\mathcal{C}(k_{ML}P)/k_{ML}$. Since the most probable outcome of $(m-1)$ Bernoulli trials[1] with success probability p is the integer number between $mp-1$ and mp, we

---

[1]Each active transmitter estimates the state of the remaining $(m-1)$ users.
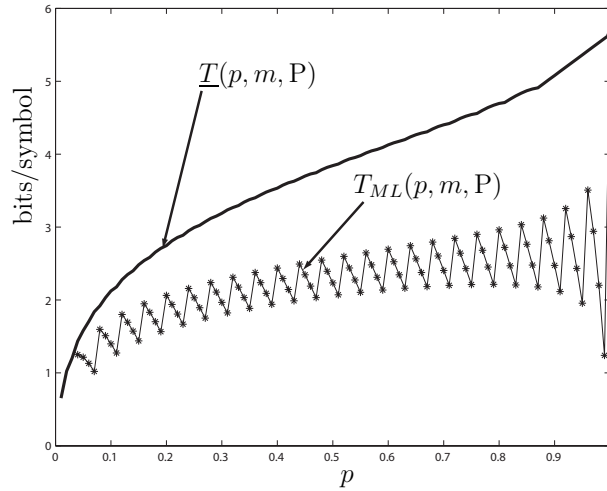
Figure 3.9: Throughput performance of the proposed estimator vs ML estimator for the number of active users ($P = 20$dB, $m = 25$).

have that $k_{ML} = \lfloor mp \rfloor$. Thus, we obtain the following expression for the expected sum-rate capacity:

$$T_{ML}(p, m, P) \triangleq \frac{mp}{k_{ML}} \mathcal{C}(k_{ML}P) F_{m-1, k_{ML}-1}(p). \qquad (3.42)$$

Fig. 3.9 compares $\underline{T}(p, m, P)$ and (3.42) for the case $m = 25$ and $P = 20$dB. We remark is that the ML estimator for the number of active users result in a strictly suboptimal throughput performance.

3.8   Discussion and practical considerations

In networking, much research effort has been put in the design of distributed algorithms where each agent has limited information about the global state of the network. The model we developed in this work allowed us to focus on the rate allocation problem that occurs when multiple nodes attempt to access a common medium, and when the set of active users is not available to the transmitters. Our analysis has lead to a distributed algorithm which is easily implementable in practical systems, and which is optimal in some information-theoretic sense. The rule of thumb which we have developed is that, upon transmission, senders should

estimate the number of active users according to a prescribed algorithm based on the knowledge of the population size and the transmission probability m, and then choose the encoding rate accordingly.

In this work we focused primarily on the problem of characterizing the throughput assuming perfect symmetry in the network, that is, the same transmission probability and received power constraint across users. The reasons for enforcing symmetry are twofold. First, throughput maximization is a meaningful performance metric only in symmetric scenarios. Second, it allows us to focus on random packet arrivals at the transmitters, and not on the different power levels at which transmitted signals are received by the common receiver. This set-up is a realistic model for uplink communications in power-controlled cellular wireless systems. Nevertheless, an interesting open question is how to apply the layering approach to the m-user AWGN-RAC with unequal power levels at the receiver, assuming that each sender only knows its own power level and state.

We made the underlying assumption that users can be synchronized, both at block and symbol level. In light of this assumption, a time-sharing protocol could be employed to prove achievability results. A simple way to achieve this partial form of cooperation among senders is to establish, prior to any transmission, that different coding schemes are used in different fractions of the transmission time. However, in practice achieving such complete synchronization may not be feasible. An interesting open question is to characterize the performance loss due to lack of synchronism. In this case, the resulting capacity region need not be convex, as for the collision model without feedback studied by Massey and Mathys [16].

We also assumed that the receiver has perfect CSI, that is, it knows the set of active users. The question, relevant in practice, of how the receiver can acquire such information is not discussed here, and we refer the reader to the recent studies of Fletcher *et al.* [11], Angelosante *et al.* [4], and Biglieri and Lops [7], which address the issue using sparse signal representation techniques and random set theory.

Finally, in this work the transmission probability p and the number of users m play a pivotal role in setting the encoding rate, and these quantities are

supposed to be known at the transmitters. The probability p is determined by the burstiness of the sources, while m has to be communicated from the receiver to the transmitters. In practice, our model applies to communication scenarios in which the base station grants access to the uplink channel to m users, but where only a subset of these users actually transmit data.

3.9 Appendix

3.9.1 Proof of Theorem 3.4.3

Observe that $\underline{\mathscr{C}}_2$ is a polytope in $\mathbb{R}_+^4$ defined as the intersection of eight hyperplanes, two of which representing non-negativity constraints. By the Weyl-Minkowski theorem, $\underline{\mathscr{C}}_2$ is the convex hull of finitely many rate vectors. It is tedious but simple to verify that

$$\underline{\mathscr{C}}_2 = \operatorname{conv}\{\mathbf{v}_1,\ldots,\mathbf{v}_{14}\}$$

$$= \operatorname{conv}\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \mathcal{C}(\mathrm{P}_2) \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\mathrm{P}_2) \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ 0 \\ \mathcal{C}(\mathrm{P}_1) \\ 0 \end{bmatrix}, \right.$$

$$\begin{bmatrix} 0 \\ \mathcal{C}(\mathrm{P}_2) \\ 0 \\ \mathcal{C}(\mathrm{P}_2) \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\frac{\mathrm{P}_1}{\mathrm{P}_2+1}) \\ \mathcal{C}(\mathrm{P}_2) \\ 0 \\ \mathcal{C}(\mathrm{P}_2) \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\frac{\mathrm{P}_2}{\mathrm{P}_1+1}) \\ 0 \\ \mathcal{C}(\frac{\mathrm{P}_2}{\mathrm{P}_1+1}) \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\frac{\mathrm{P}_2}{\mathrm{P}_1+1}) \\ \mathcal{C}(\mathrm{P}_1) \\ 0 \end{bmatrix},$$

$$\left. \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\frac{\mathrm{P}_2}{\mathrm{P}_1+1}) \\ \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\frac{\mathrm{P}_2}{\mathrm{P}_1+1}) \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\frac{\mathrm{P}_1}{\mathrm{P}_2+1}) \\ \mathcal{C}(\mathrm{P}_2) \\ \mathcal{C}(\frac{\mathrm{P}_1}{\mathrm{P}_2+1}) \\ \mathcal{C}(\mathrm{P}_2) \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\mathrm{P}_2) \\ \mathcal{C}(\frac{\mathrm{P}_1}{\mathrm{P}_2+1}) \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\mathrm{P}_2) \\ 0 \\ \mathcal{C}(\frac{\mathrm{P}_2}{\mathrm{P}_1+1}) \end{bmatrix}, \begin{bmatrix} \mathcal{C}(\mathrm{P}_1) \\ \mathcal{C}(\mathrm{P}_2) \\ \mathcal{C}(\frac{\mathrm{P}_1}{\mathrm{P}_2+1}) \\ \mathcal{C}(\frac{\mathrm{P}_2}{\mathrm{P}_1+1}) \end{bmatrix} \right\}. \quad (3.43)$$

By convexity, it suffices to show that for every $i \in \{1,\ldots,14\}$, there exists an achievable rate vector $\mathbf{r}_i$ such that $d(\mathbf{v}_i,\mathbf{r}_i) \leq 1$. It is straightforward to verify that, for every $i \in \{1,\ldots,11\}$, $\mathbf{v}_i \in \underline{\mathscr{C}}_2' \cup \underline{\mathscr{C}}_2''$. Thus, $d(\mathbf{v}_i,\mathbf{r}_i) = 0$ for all $i \in \{1,\ldots,11\}$.

Consider the rate vector $\mathbf{r}_{12} \in \underline{\mathscr{C}}_2'''$ obtained by setting equality sign in

the inequalities (3.14) with $\beta_1 = \frac{P_2}{P_1}$ and $\beta_2 = 1$, i.e., $\mathbf{r}_{12} = \left[\mathcal{C}(P_2) + \mathcal{C}\left(\frac{P_1 - P_2}{2P_2 + 1}\right), \mathcal{C}(P_2), \mathcal{C}\left(\frac{P_1 - P_2}{2P_2 + 1}\right), 0\right]^{\mathrm{T}}$. We have that

$$
\begin{aligned}
d(\mathbf{v}_{12}, \mathbf{r}_{12}) &\leq \sqrt{\left|\mathcal{C}(P_1) - \mathcal{C}(P_2) - \mathcal{C}\left(\frac{P_1 - P_2}{2P_2 + 1}\right)\right|^2 + \left|\mathcal{C}\left(\frac{P_1}{P_2 + 1}\right) - \mathcal{C}\left(\frac{P_1 - P_2}{2P_2 + 1}\right)\right|^2} \\
&= \sqrt{\left|\frac{1}{2}\log\left(1 + \frac{P_1 P_2 - P_2^2}{P_1 P_2 - P_2^2}\right)\right|^2 + \left|\frac{1}{2}\log\frac{2P_2 + 1}{P_2 + 1}\right|^2} \\
&\leq \sqrt{\left|\frac{1}{2}\log\frac{2P_1 P_2}{P_1 P_2 - P_2^2}\right|^2 + \left|\frac{1}{2}\log\frac{2P_2 + 1}{P_2 + 1}\right|^2} \\
&\leq \frac{1}{\sqrt{2}}.
\end{aligned}
\tag{3.44}
$$

Next, consider the rate vector $\mathbf{r}_{13} = [\mathcal{C}(P_1), \mathcal{C}(P_2), 0, 0]^{\mathrm{T}} \in \mathscr{C}_2'''$, obtained by setting equality sign in the inequalities (3.14) with $\beta_1 = 1$ and $\beta_2 = 1$. We have that

$$
d(\mathbf{v}_{13}, \mathbf{r}_{13}) = \left|\mathcal{C}\left(\frac{P_2}{P_1 + 1}\right)\right| \leq \frac{1}{\sqrt{2}}.
\tag{3.45}
$$

Finally, the distance between $\mathbf{v}_{14}$ and $\mathbf{r}_{12}$ can be bounded as follows

$$
\begin{aligned}
&d(\mathbf{v}_{14}, \mathbf{r}_{12}) \\
&\leq \sqrt{\left|\mathcal{C}(P_1) - \mathcal{C}(P_2) - \mathcal{C}\left(\frac{P_1 - P_2}{2P_2 + 1}\right)\right|^2 + \left|\mathcal{C}\left(\frac{P_1}{P_2 + 1}\right) - \mathcal{C}\left(\frac{P_1 - P_2}{2P_2 + 1}\right)\right|^2 + \left|\mathcal{C}\left(\frac{P_2}{P_1 + 1}\right)\right|^2} \\
&\leq \frac{\sqrt{3}}{2}.
\end{aligned}
\tag{3.46}
$$

Combining (3.44), (3.45), and (3.46) we conclude that $d(\mathbf{v}_i, \mathbf{r}_i) \leq \frac{\sqrt{3}}{2}$, $i \in \{12, 13, 14\}$, which concludes the proof.

3.9.2 Proof of Theorem 3.5.1

Inequalities (3.18) follow immediately from assumption **A1.**. Next, fix $i_1 \neq i_2 \neq \ldots \neq i_m \in \{1, \ldots, m\}$. By Fano's inequality, we have that, for all $r \in$

$\{1, \ldots, m\}$,

$$H\left(\bigcup_{k=1}^{r} \mathcal{W}_{i_k}(\{i_1 \ldots i_r\}) \,|\, \mathbf{Y}_{i_1 \ldots i_r}\right) \leq n\epsilon_n, \tag{3.47}$$

where $\epsilon_n \to 0$ in the limit of n going to infinity. In particular, (3.47) implies that

$$H\left(\mathcal{W}_{i_r}(\{i_1 \ldots i_r\}) \,|\, \mathbf{Y}_{i_1 \ldots i_r}\right) \leq n\epsilon_n. \tag{3.48}$$

Let $K \in \{1, \ldots, m\}$. Then, the following chain of equalities holds:

$$
\begin{aligned}
n \sum_{k=1}^{K} & r_{i_k}(\{i_1 \ldots i_k\}) \\
&= H\left(\bigcup_{k=1}^{K} \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\right) \\
&= H\left(\bigcup_{k=1}^{K} \mathcal{W}_{i_k}(\{i_1 \ldots i_k\}) \,\Big|\, \bigcup_{k=1}^{K} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}\right) \\
&= I\left(\bigcup_{k=1}^{K} \mathcal{W}_{i_k}(\{i_1 \ldots i_k\}); \mathbf{Y}_{i_1 \ldots i_K} \,\Big|\, \bigcup_{k=1}^{K} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}\right) \\
&\quad + H\left(\bigcup_{k=1}^{K} \mathcal{W}_{i_k}(\{i_1 \ldots i_k\}) \,\Big|\, \mathbf{Y}_{i_1 \ldots i_K}, \bigcup_{k=1}^{K} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}\right) \tag{3.49}
\end{aligned}
$$

The first term in the right hand side of (3.49) can be upper bounded as follows

$$
\begin{aligned}
I\left(\bigcup_{k=1}^{K} \mathcal{W}_{i_k}(\{i_1 \ldots i_k\}); \mathbf{Y}_{i_1 \ldots i_K} \,\Big|\, \bigcup_{k=1}^{K} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}\right) & \\
= H\left(\mathbf{Y}_{i_1 \ldots i_K}\right) - H\left(\mathbf{Y}_{i_1 \ldots i_K} \,\Big|\, \bigcup_{k=1}^{K} \mathcal{W}_{i_k}\right) & \\
\leq H\left(\mathbf{Y}_{i_1 \ldots i_K}\right) - H\left(\mathbf{Y}_{i_1 \ldots i_K} \,\Big|\, \bigcup_{k=1}^{K} \mathcal{W}_{i_k}, \mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_K}\right) & \\
= \sum_{t=1}^{n} I\left(X_{i_1,t}, \ldots, X_{i_K,t}; Y_{i_1 \ldots i_K,t}\right) & \tag{3.50}
\end{aligned}
$$

where we use the fact conditioning reduces the entropy and the memoryless property of the channel. On the other hand, application of the chain rule on the second

term at the left hand side of (3.49) yields

$$H\left(\bigcup_{k=1}^{K} \mathcal{W}_{i_k}(\{i_1 \ldots i_k\}) \middle| \mathbf{Y}_{i_1 \ldots i_K}, \bigcup_{k=1}^{K} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}\right)$$

$$= \sum_{r=1}^{K} H\left(\mathcal{W}_{i_r}(\{i_1 \ldots i_r\}) \middle| \mathbf{Y}_{i_1 \ldots i_K}, \bigcup_{k=1}^{K} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}, \bigcup_{k=r+1}^{K} \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\right)$$

$$= \sum_{r=1}^{K} H\left(\mathcal{W}_{i_r}(\{i_1 \ldots i_r\}) \middle| \mathbf{Y}_{i_1 \ldots i_K}, \bigcup_{k=1}^{r} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}, \bigcup_{k=r+1}^{K} \mathcal{W}_{i_k}\right)$$

$$= \sum_{r=1}^{K} H\left(\mathcal{W}_{i_r}(\{i_1 \ldots i_r\}) \middle| \mathbf{Y}_{i_1 \ldots i_K}, \bigcup_{k=1}^{r} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}, \bigcup_{k=r+1}^{K} \mathcal{W}_{i_k}, \bigcup_{k=r+1}^{K} \mathbf{X}_{i_k}\right)$$

$$\tag{3.51}$$

$$= \sum_{r=1}^{K} H\left(\mathcal{W}_{i_r}(\{i_1 \ldots i_r\}) \middle| \mathbf{Y}_{i_1 \ldots i_r}, \bigcup_{k=1}^{r} \{\mathcal{W}_{i_k} \setminus \mathcal{W}_{i_k}(\{i_1 \ldots i_k\})\}\right)$$

$$= \sum_{r=1}^{K} H\left(\mathcal{W}_{i_r}(\{i_1 \ldots i_r\}) \middle| \mathbf{Y}_{i_1 \ldots i_r}\right) \tag{3.52}$$

$$\leq K n \epsilon_n, \tag{3.53}$$

where (3.51) uses the fact that $\mathbf{X}_{i_k}$ is a function of $\mathcal{W}_{i_k}$, (3.52) uses the fact conditioning reduces the entropy, and (3.53) follows from (3.48).

Therefore, substituting (3.50) and (3.53) into (3.49), we obtain that

$$n \sum_{k=1}^{K} r_{i_k}(\{i_1 \ldots i_k\}) \leq \sum_{t=1}^{n} I\left(X_{i_1,t}, \ldots, X_{i_K,t}; Y_{i_1 \ldots i_K,t}\right) + n K \epsilon_n, \tag{3.54}$$

and the claim is completed by introducing a standard timesharing random variable and letting the block size n tend to infinity.

3.9.3   Proof of Theorem 3.6.2

Let $\mathcal{P}$ denote the convex subset of $\mathbb{R}^m$ described described by inequalities (3.25a) and (3.25b). First we prove the converse part, by establishing that $\mathscr{C}_{\boldsymbol{\rho}} \subseteq \mathcal{P}$.

As a first step, we derive a useful identity. Let $k \in \{1, \ldots, m\}$. Then,

$$\sum_{i_1 \neq \cdots \neq i_m \in \{1,\ldots,m\}} r_{i_k}(\{i_1 \ldots i_k\}) = (m-k)! \sum_{i_1 \neq \cdots \neq i_k \in \{1,\ldots,m\}} r_{i_k}(\{i_1 \ldots i_k\})$$

$$= (m-k)!(k-1)! \sum_{\substack{A \subseteq \{1,\ldots,m\} \\ |A|=k}} \sum_{i \in A} r_i(A)$$

$$= (m-k)!(k-1)!\rho_k, \tag{3.55}$$

where the second equality uses the fact that $r_{i_k}(i_1 \ldots i_k) = r_{i_k}(\{i_{\sigma_1}, \ldots, i_{\sigma_{k-1}}, i_k\})$ for any permutation $\boldsymbol{\sigma}$ over the set $\{1, \ldots, k-1\}$. Now we can establish the necessity of (3.25b). It follows from (3.24) that the following inequality has to hold

$$\sum_{k=1}^{m} r_{i_k}(i_1 \ldots i_k) \leq 1, \tag{3.56}$$

for all $i_1 \neq \cdots \neq i_m \in \{1, \ldots, m\}$. By summing both sides of (3.56) over all permutations over the first m integers, we obtain

$$\sum_{i_1 \neq \cdots \neq i_m \in \{1,\ldots,m\}} \sum_{k=1}^{m} r_{i_k}(\{i_1 \ldots i_k\}) \leq m!. \tag{3.57}$$

By means of (3.55), (3.57) can be re-written as

$$\sum_{k=1}^{m} (m-k)!(k-1)!\rho_k \leq m!. \tag{3.58}$$

Dividing both sides of (3.58) by $m!$, we conclude that (3.25b) is a necessary condition for the achievability of a rate vector $\boldsymbol{\rho}$.

Next, note from (3.23) that $r_i(A) \geq r_i(B)$ for all $i \in A \subseteq B \subseteq \{1, \ldots, m\}$ is a necessary condition to the achievability of a rate vector $\{r_i(A)\}$. By summing

these inequalities over all B having cardinality $|A| + 1$, we obtain that

$$r_i(A) \geq \frac{1}{m - |A|} \sum_{\substack{B:i \in A \subseteq B \subseteq \{1,\ldots,m\} \\ |B| = |A| + 1}} r_i(B). \tag{3.59}$$

Next, observe that, for every $k \in \{1, \ldots, m - 1\}$,

$$\rho_k \geq \sum_{\substack{A:A \subseteq \{1,\ldots,m\} \\ |A| = k}} \sum_{i \in A} r_i(A)$$

$$= \sum_{i=1}^{m} \sum_{\substack{A:A \subseteq \{1,\ldots,m\} \\ i \in A, |A| = k}} r_i(A)$$

$$\geq \sum_{i=1}^{m} \sum_{\substack{A:A \subseteq \{1,\ldots,m\} \\ i \in A, |A| = k}} \frac{1}{m - k} \sum_{\substack{B:i \in A \subseteq B \subseteq \{1,\ldots,m\} \\ |B| = k+1}} r_i(B) \tag{3.60}$$

$$= \frac{1}{m - k} \sum_{i=1}^{m} \sum_{\substack{B:B \subseteq \{1,\ldots,m\} \\ i \in B, |B| = k+1}} \sum_{\substack{A:A \subseteq B \\ i \in A, |A| = k}} r_i(B) \tag{3.61}$$

$$= \frac{k}{m - k} \sum_{i=1}^{m} \sum_{\substack{B:B \subseteq \{1,\ldots,m\} \\ i \in B, |B| = k+1}} r_i(B)$$

$$= \frac{k}{m - k} \rho_{k+1} \tag{3.62}$$

where (3.60) follows from (3.59), while (3.61) is obtained observing that there are k subsets of B which have cardinality k and contain the element i. After multiplying right and left hand side of (3.62) by $\frac{((m-1)!)}{(k-1)!}$ and rearranging the terms, we obtained the desired inequality

$$\frac{\rho_k}{k \binom{m}{k}} \geq \frac{\rho_{k+1}}{(k + 1) \binom{m}{k+1}}$$

which proves (3.25a). In summary, we showed that inequalities (3.25a) and (3.25b) are necessary conditions for the achievability of a rate vector $\boldsymbol{\rho}$, i.e., $\mathscr{C}_{\boldsymbol{\rho}} \subseteq \mathcal{P}$.

Next, we prove the achievability of $\mathcal{P}$, establishing the reversed inclusion $\mathcal{P} \subseteq \mathscr{C}_{\boldsymbol{\rho}}$. To do so, it suffices to show that the extreme points of $\mathcal{P}$ are achievable,

as the rest of the region can be achieved my means of a time-sharing protocol. We claim that

$$\mathcal{P} = \operatorname{conv}\left\{ \mathbf{0}, \left\{ \frac{1}{k} \sum_{i=1}^{k} i \binom{m}{i} \mathbf{e}_i \right\}_{k=1}^{m} \right\}. \tag{3.63}$$

where the vector $\mathbf{e}_i$ denotes the ith unit vector in $\mathbb{R}^m$. To see this, consider an invertible linear transformation $L : \mathbb{R}^m \to \mathbb{R}^m$ given by

$$\begin{cases} x_m &= \frac{\rho_m}{m\binom{m}{m}}, \\ x_k &= \frac{\rho_k}{k\binom{m}{k}} - \frac{\rho_{k+1}}{(k+1)\binom{m}{k+1}}, \quad k \in \{1, \ldots, m-1\}. \end{cases} \tag{3.64}$$

It is straightforward to check that the image $\mathcal{P}$ under L is given by the oriented m-simplex $L\mathcal{P} = \{\mathbf{x} \in \mathbb{R}_+^m : \sum_{k=1}^{m} kx_k \leq 1\} = \operatorname{conv}\{\mathbf{0}, \mathbf{v}_1', \ldots, \mathbf{v}_m'\}$, wherein $\mathbf{v}_k' = \frac{1}{k}\mathbf{e}_k$. Since L is invertible, the extreme points of $\mathcal{P}$ can be obtained by applying L^{-1} to the extreme points of $L\mathcal{P}$. Thus, $\mathcal{P} = \operatorname{conv}\{\mathbf{0}, \boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_m\}$ where

$$\boldsymbol{\rho}_k = L^{-1}\mathbf{v}_k' = \frac{1}{k} \sum_{i=1}^{k} i \binom{m}{i} \mathbf{e}_i, \tag{3.65}$$

$k \in \{1, \ldots, m\}$. Hence (3.63) is proved.

Next, we show that each rate vector $\boldsymbol{\rho}_k$ given by (3.65) is achievable. Consider the following message structure:

$$\mathcal{W}_i = \{W_{i,1}, \ldots, W_{i,m}\} \tag{3.66}$$

and

$$\mathcal{W}_i(A) = \begin{cases} \cup_{j \geq |A|} W_{i,j}, & i \in A; \\ \emptyset, & i \notin A. \end{cases} \tag{3.67}$$

It is immediate to verify that the above sets satisfy conditions **A1.**, so the message structure is well defined. For every i, sender i transmits m independent messages $\{W_{i,1}, \ldots, W_{i,m}\}$ encoded at rates $\{R_{i,1}, \ldots, R_{i,m}\}$. For every $k \in \{1, \ldots, m\}$, the kth message $W_{i,k}$ is decoded at receiver A if $i \in A$ and if $|A| \leq k$, that is, if user i is active and there are less than k active users. To achieve the rate vector $\boldsymbol{\rho}_k$ it suffices to set $R_{i,k} = \frac{1}{k}$ for all i, and the other rates equal to zero, that is,

each sender i transmits a *single* message of information $W_{i,k}$ encoded at rate $\frac{1}{k}$. Encoding is performed by means of a standard multiple-access random codebook. It follows from (3.67) that receiver A decodes $W_{i,k}$ if $i \in A$ and $|A| \le k$. Thus, we have

$$r_i(A) = \begin{cases} \frac{1}{k}, & i \in A \text{ and } |A| \le k; \\ 0, & \text{otherwise.} \end{cases} \tag{3.68}$$

Observe that for every receiver A the sum of the rates of the decoded messages is at most 1. It follows that decoding can be performed by means of a standard $k$user multiple-access decoder. By plugging (3.68) into (3.21), we obtain that

$$\rho_{k,i} = \begin{cases} \frac{i}{k}\binom{m}{i}, & \text{if } i \in \{1, \dots, k\} \\ 0, & \text{otherwise,} \end{cases} \tag{3.69}$$

hence (3.65) is achievable.

3.9.4 Proof of Theorem 3.6.3

In order to prove the theorem, we first need to state two lemmas. The first lemma builds upon properties of the cumulative distribution function of the Binomial distribution.

Lemma 3.9.1. *Let $k \in \{1, \dots, m-1\}$. There exists a $p_k \in \left(0, \frac{k}{m}\right)$ such that*

$$\frac{1}{k}F_{m-1,k-1}(p) - \frac{1}{k+1}F_{m-1,k}(p) \begin{cases} > 0, & p < p_k \\ = 0, & p = p_k \\ < 0, & p > p_k \end{cases} \tag{3.70}$$

Proof. Define $f(p) = \frac{1}{k}F_{m-1,k-1}(p) - \frac{1}{k+1}F_{m-1,k}(p)$. The binomial sum $F_{m-1,k-1}(p)$ is related to the incomplete Beta function by [1, (6.6.4) page 263]

$$F_{m-1,k-1}(p) = 1 - k\binom{m-1}{k}\int_0^p t^{k-1}(1-t)^{m-1-k}dt. \tag{3.71}$$

Substituting (3.71) into the definition of $f(p)$ and differentiating, we obtain the following expression for the derivative of f with respect to p $f'(p) = -\frac{1}{p(1-p)}f_{m-1,k}(p)\left[1 - p\frac{m}{k+1}\right]$. By studying the sign of $f'(p)$ one can see that $f(p)$ is a strictly decreasing

function of p in the range $\left(0, \frac{k+1}{m}\right)$, reaches a minimum at $p = \frac{k+1}{m}$ and is a strictly increasing in the interval $\left(\frac{k+1}{m}, 1\right)$. We have $f(1) = 0$, and the Taylor expansion centered at $p = 1$ shows that $f(p)$ increases to zero as p tends to one. Thus, $f\left(\frac{k+1}{m}\right) < 0$. Note that $f(0) > 0$ so, by the monotonicity of f and by the mean value theorem, there exists a unique $p_k \in \left(0, \frac{k+1}{m}\right)$ such that

$$
f(p) \begin{cases} > 0, & p < p_k \\ = 0, & p = p_k \\ < 0, & p > p_k \end{cases} . \tag{3.72}
$$

To complete the proof, we show that $p_k < \frac{k}{m}$. Direct computation shows that $p_1 = 1/m$, while for $k \in \{2, \ldots, m-1\}$, we have that

$$
\begin{aligned}
f\left(\tfrac{k}{m}\right) &= \tfrac{1}{k} F_{m-1,k-1}\left(\tfrac{k}{m}\right) - \tfrac{1}{k+1} F_{m-1,k}\left(\tfrac{k}{m}\right) \\
&= \left(\tfrac{1}{k} - \tfrac{1}{k+1}\right) F_{m-1,k-1}\left(\tfrac{k}{m}\right) - \tfrac{1}{k+1} f_{m-1,k}\left(\tfrac{k}{m}\right) \\
&< \left(\tfrac{1}{k} - \tfrac{1}{k+1}\right) k f_{m-1,k-1}\left(\tfrac{k}{m}\right) - \tfrac{1}{k+1} f_{m-1,k-1}\left(\tfrac{k}{m}\right) \\
&= 0, \tag{3.73}
\end{aligned}
$$

where the inequality follows from the fact that $f_{m-1,i}\left(\tfrac{k}{m}\right) \le f_{m-1,k-1}\left(\tfrac{k}{m}\right)$ for $i \in \{0, \ldots, k-1\}$, with equality iff $i = k-1$, and that $f_{m-1,k-1}\left(\tfrac{k}{m}\right) = f_{m-1,k}\left(\tfrac{k}{m}\right)$ for $k \in \{2, \ldots, m-1\}$. Thus, (3.72) and (3.73) show that $p_k < \frac{k}{m}$ as claimed. $\square$

Roughly speaking, the above says that to achieve the throughput the encoding rate has to decrease as the transmission probability increases. The second lemma shows that $1/k$ is the optimal encoding rate when p is in the kth interval of the partition $\Pi_m(P)$.

Lemma 3.9.2. *Let $k \in \{1, \ldots, m\}$. Define $p_0 \triangleq 0$ and $p_m \triangleq 1$ and let $\{p_k\}_{k=1}^{m-1}$ be as in Lemma 4.4.2. Then,*

$$
\frac{1}{k} F_{m-1,k-1}(p) \ge \frac{1}{j} F_{m-1,j-1}(p), \quad j \in \{1, \ldots, m\}, \tag{3.74}
$$

for $p \in [p_{k-1}, p_k]$.

Proof. In virtue of Lemma 4.4.2, it suffices to show that $p_k < p_{k+1}$, for $k \in \{0, \ldots, m-1\}$. As $p_1 \in (0, 1/m]$, it follows that $p_0 < p_1$. Next, suppose that $k \in \{1, \ldots, m-1\}$. Lemma 4.4.2 shows that $\frac{1}{k} F_{m-1,k-1}(p_k) = \frac{1}{k+1} F_{m-1,k}(p_k)$ and that $p_k \in \left(0; \frac{k}{m}\right)$. Thus, we have

$$
\begin{aligned}
& \tfrac{1}{k+1} F_{m-1,k}(p_k) - \tfrac{1}{k+2} F_{m-1,k+1}(p_k) \\
&= \tfrac{1}{k} F_{m-1,k-1}(p_k) - \tfrac{1}{k+2} F_{m-1,k+1}(p_k) \\
&= \left(\tfrac{1}{k} + \tfrac{1}{k+2}\right) F_{m-1,k-1}(p_k) - \tfrac{1}{k+2}\left(F_{m-1,k-1}(p_k) + F_{m-1,k+1}(p_k)\right) \\
&> \tfrac{2(k+1)}{k(k+2)} F_{m-1,k-1}(p_k) - \tfrac{2}{k+2} F_{m-1,k}(p_k) \\
&= \tfrac{2(k+1)}{k(k+2)} F_{m-1,k-1}(p_k) - \tfrac{2(k+1)}{k(k+2)} F_{m-1,k-1}(p_k) \\
&= 0, \hspace{6cm} (3.75)
\end{aligned}
$$

where the inequality uses the fact that $F_{m-1,k-1}(p_k) + F_{m-1,k+1}(p_k)) < 2 F_{m-1,k}(p_k)$ for $p < k/m$. Comparing (3.70) and (3.75), we obtain the desired inequality $p_k < p_{k+1}$. □

Using the above lemma, it is immediate to prove theorem 3.6.3.

Proof. Observe that the optimum value of a linear program, if it exists, is always achieved at one of the extreme point of the feasibility set. Thus, (3.63) implies that

$$
\begin{aligned}
\underline{T}(p, m, \mathrm{P}) &= \max_{k \in \{1, \ldots, m\}} \frac{1}{k} \sum_{i=1}^{k} i \binom{m}{i} p^i (1-p)^{m-i} \\
&= \max_{k \in \{1, \ldots, m\}} m p \frac{1}{k} F_{m-1,k-1}(p) \\
&= m p \sum_{k=1}^{m} \frac{1}{k} F_{m-1,k-1}(p) \mathbb{1}_{\{p \in (p_{k-1}, p_k]\}},
\end{aligned}
$$

where the last equality follows from Lemma 3.9.2. □

3.9.5 Proof of Theorem 3.7.4

Let $c_k \triangleq \mathcal{C}(k\mathrm{P})$. In order to evaluate $\overline{T}(p, m, \mathrm{P})$, it is convenient to make the change of variable

$$
\begin{cases}
x_m &= \frac{\rho_m}{m\binom{m}{m}}, \\
x_k &= \frac{\rho_k}{k\binom{m}{k}} - \frac{\rho_{k+1}}{(k+1)\binom{m}{k+1}}, \quad k \in \{1, \ldots, m-1\}.
\end{cases}
\tag{3.76}
$$

Substituting the new variables into (3.22), (3.25a), and (3.25b) and performing a modicum of algebra, we obtain,

$$
\overline{T}(p, m, \mathrm{P}) = \max_{\mathbf{x} \in \overline{\mathscr{C}}_{\mathbf{x},m}} mp \sum_{i=1}^{m} F_{m-1,i-1}(p) \, x_i,
\tag{3.77}
$$

where $\overline{\mathscr{C}}_{\mathbf{x},m}$ denote the set of rates $\{x_k\} \in \mathbb{R}_+^m$ such that

$$
\sum_{k=1}^{K-1} k x_k + K \sum_{k=K}^{m} x_k \leq c_K
\tag{3.78}
$$

for every $K \in \{1, \ldots, m\}$. Observe that the optimum value of the linear program (3.77) is achieved at one of the extreme point of the feasibility set. Therefore, to prove the theorem it suffices to show that $\{\mathbf{v}_k\}_{k=1}^m$ as defined in (3.36-3.39) are extreme points of $\overline{\mathscr{C}}_{\mathbf{x},m}$, and that the objective function in (3.77) reaches a strict local maximum at $\mathbf{v}_k$ when p is in the kth interval of the partition $\Pi_m(\mathrm{P})$.

For every $k \in \{1, \ldots, m\}$, it is straightforward to check that $\mathbf{v}_k$ satisfies (3.78) for $K \in \{k, \ldots, m\}$, and that $\mathbf{v}_k$ has $k - 1$ zero components. Thus, we conclude that $\mathbf{v}_k$ is an extreme point of $\overline{\mathscr{C}}_{\mathbf{x},m}$.

Next, we establish that if $p \in [p_{k-1}, p_k]$, where $\{p_{k-1}\}$ are defined in Lemma 4.4.2, then the objective function reaches a local maximum at $\mathbf{v}_k$. We proceed by showing that the objective function at $\mathbf{v}_k$ is strictly greater than at any of its neighboring extreme points. By definition, two extreme points are neighbors if they are connected by an edge. It is possible to show that $\mathbf{v}_k$ has exactly m neighbor extreme points, which we denote by $\left\{\mathbf{n}_j^{(k)}\right\}_{j=1}^m$. The proof of this fact is straightforward albeit fairly lengthy, so is not reported here. For $k \in \{1, \ldots, m-1\}$,

we have that

- If $j \in \{1, \ldots, k-1\}$, then

$$
n_{j,i}^{(k)} = \begin{cases} \frac{k}{j(k-j)}c_j - \frac{1}{k-j}c_k, & i = j, \\ 0, & i \in \{1, \ldots, j-1\} \cup \{j+1, \ldots, k-1\}, \\ \frac{k-j+1}{j(k-j)}c_k - \frac{1}{k-j}c_j - c_{k+1}, & i = k, \\ v_{k,i}, & i \in \{k+1, \ldots, m\} \end{cases}
$$

(3.79)

- If $j = k$, then $\mathbf{n}_j^{(k)} = \mathbf{v}_{k+1}$.

- If $j \in \{k+1, \ldots, m-2\}$, then

$$
n_{j,i}^{(k)} = \begin{cases} \frac{3}{2}c_{j-1} - c_{j-2} - \frac{1}{2}c_{j+1}, & i = j-1, \\ 0, & i = j, \\ \frac{3}{2}c_{j+1} - c_{j+2} - \frac{1}{2}c_{j-1}, & i = j+1, \\ v_{k,i}, & i \in \{1, \ldots, j-2\} \cup \{j+2, \ldots, m\} \end{cases}
$$

(3.80)

- If $j = m-1$, then

$$
n_{m-1,i}^{(k)} = \begin{cases} \frac{3}{2}c_{m-2} - c_{m-3} - \frac{1}{2}c_m, & i = m-2, \\ 0, & i = m-1, \\ \frac{1}{2}c_m - \frac{1}{2}c_{m-2}, & i = m, \\ v_{k,i}, & i \in \{1, \ldots, m-3\} \end{cases}
$$

(3.81)

- Finally, if $j = m$ then

$$
n_{m,i}^{(k)} = \begin{cases} c_{m-1} - c_{m-2}, & i = m-1, \\ 0, & i = m, \\ v_{k,i}, & i \in \{1, \ldots, m-2\}, \end{cases}
$$

(3.82)

On the other hand, for $k = m$ and $j \in \{1, \ldots, m-1\}$, we have that

$$n_{j,i}^{(m)} = \begin{cases} \frac{m}{j(m-j)} c_j - \frac{1}{m-j} c_m, & i = j, \\ \frac{1}{m-j} c_m - \frac{1}{m-j} c_j, & i = m, \end{cases} \tag{3.83}$$

It can be immediately verified that $\left\{ \mathbf{n}_j^{(k)} \right\}_{j=1}^m$ as defined above are extreme points of $\overline{\mathscr{C}}_{\mathbf{x},m}$, and neighbors of $\mathbf{v}_k$.

Next, we establish that the objective function in (3.77) reaches a local maximum at $\mathbf{v}_k$ by comparing the value achieved at $\mathbf{v}_k$ to the one at its neighboring extreme points. First, suppose $k \in \{1, \ldots, m-1\}$.

- If $j \in \{1, \ldots, k-1\}$, we can observe, from plugging (3.79) into (3.77) and performing some algebraic manipulations, that

$$mp \sum_{i=1}^m (v_{k,i} - n_{j,i}^{(k)}) F_{m-1,i-1}(p)$$

$$= F_{m-1,k-1}(p) \left(\frac{1}{k-j} c_j - \frac{j}{k(k-j)} c_k \right) - F_{m-1,j-1}(p) \left(\frac{k}{j(k-j)} c_j - \frac{1}{k-j} c_k \right)$$

$$> F_{m-1,k-1}(p) \left(\frac{1}{k-j} c_j - \frac{j}{k(k-j)} c_k \right) - \frac{j}{k} F_{m-1,k-1}(p) \left(\frac{k}{j(k-j)} c_j - \frac{1}{k-j} c_k \right)$$

$$= 0,$$

because $F_{j-1} < \frac{j}{k} F_{m-1,k-1}(p)$ if p is in the kth interval of the partition $\Pi_m(P)$.

- If $j = k$, then

$$mp \sum_{i=1}^m (v_{k,i} - n_{k,i}^{(k)}) F_{m-1,i-1}(p)$$

$$= \frac{1}{k+1} \left(\frac{k+1}{k} c_k - c_{k-1} \right) \left(\frac{1}{k} F_{m-1,k-1}(p) - \frac{1}{k+1} F_{m-1,k}(p) \right)$$

$$> 0,$$

- If $j \in \{k+1, \ldots, m-1\}$, then

$$mp \sum_{i=1}^{m} (v_{k,i} - n_{j,i}^{(k)}) F_{m-1,i-1}(p) =$$

$$= 2 \left(c_j - \frac{c_{j-1} + c_{j+1}}{2} \right) \left(F_{m-1,k}(p) - \frac{F_{m-1,k-1}(p) + F_{m-1,k+1}(p)}{2} \right)$$

$$> 0,$$

- If $j = m$, then

$$mp \sum_{i=1}^{m} (v_{k,i} - n_{m,i}^{(k)}) F_{m-1,i-1}(p) =$$

$$= (c_m - c_{m-1}) \left(F_{m-1,m-1}(p) - F_{m-1,m-2}(p) \right)$$

$$> 0,$$

Next, suppose $k = m$. Compare the utility function at $\mathbf{v}_m$ and $\mathbf{n}_j^{(m)}$.

$$mp \sum_{i=1}^{m} (v_{k,i} - n_{m,i}^{(k)}) F_{m-1,i-1}(p) =$$

$$= F_{m-1,m-1}(p) \left(\frac{1}{m-j} c_j - \frac{j}{m(m-j)} c_m \right) - F_{m-1,j-1}(p) \left(\frac{m}{j(m-j)} c_j - \frac{j}{m-j} c_m \right)$$

$$= F_{m-1,m-1}(p) \left(\frac{1}{m-j} c_j - \frac{j}{m(m-j)} c_m \right) - \frac{j}{m} F_{m-1,m-1}(p) \left(\frac{m}{j(m-j)} c_j - \frac{j}{m-j} c_k \right)$$

$$= 0.$$

Therefore, we have established that the objective function reaches a local maximum at $\mathbf{v}_k$ and completed the proof.

3.9.6   Proof of Theorem 3.7.5

Let $c_k \triangleq \mathcal{C}(k\mathrm{P})$. For every $k \in \{1, \ldots, m\}$, if $p \in (p_{k-1}, p_k]$ we have that

$$\underline{T}(p, m, \mathrm{P}) \geq mp \frac{c_k}{k} F_{m-1,k-1}(p). \tag{3.84}$$

In particular, equality holds in (3.84) when $p \in [\max(p_{k-1}, p_{k-1}(\mathrm{P})), \min(p_k, p_k(\mathrm{P}))]$. It follows that

$$\overline{T}(p, m, \mathrm{P}) - \underline{T}(p, m, \mathrm{P}) \leq mp \sum_{i=1}^{m} v_{k,i} F_{m-1,i-1}(p) - mp \frac{c_k}{k} F_{m-1,k-1}(p). \qquad (3.85)$$

for $p \in (p_{k-1}, p_k]$. To prove the theorem, we show that the right hand side of (3.85) is upper bounded by one for every $k \in \{1, \ldots, m\}$. First, we consider the case $k = 1$. By substituting (3.36) into (3.85), we obtain that

$$mp \left[\sum_{i=1}^{m} v_{1,i} F_{m-1,i-1}(p) - \mathcal{C}(\mathrm{P}) F_{m-1,0}(p) \right]$$

$$= mp \left[3(c_2 - c_1) F_{m-1,0}(p) + \sum_{j=1}^{m} (c_{j+1} - c_j) f_{m-1,j} \right]$$

$$= mp \left[\frac{3}{2} F_{m-1,0}(p) + \sum_{j=1}^{m-1} \frac{1}{2j} f_{m-1,j} \right]$$

$$= \frac{3}{2} f_{m,1}(p) + \sum_{j=1}^{m-1} \frac{j+1}{2j} f_{m,j+1}$$

$$\leq \frac{3}{2} f_{m,1}(p) + \sum_{j=2}^{m} f_{m,j}$$

$$= \frac{3}{2} f_{m,1}(p) + (1 - f_{m,0}(p) - f_{m,1}(p))$$

$$\leq 1$$

where the second equality uses the fact that $c_{j+1} - c_j \leq 1/(2j)$, while the last equality follows from $2f_{m,0}(p) \geq f_{m,1}(p)$ for $p \in (0, 1/m]$. Similarly, from (3.37) we

obtain that, for every $k \in \{2, \ldots, m-1\}$,

$$mp \left[\sum_{i=1}^{m} v_{k,i} F_{m-1,i-1}(p) - \frac{c_k}{k} F_{m-1,k-1}(p) \right]$$

$$= mp \left[(c_k - c_{k-1}) F_{m-1,k-1}(p) + \sum_{i=k+1}^{m-1} (2c_i - c_{i-1} - c_{i+1}) F_{m-1,i-1}(p) \right.$$

$$\left. + (c_m - c_{m-1}) F_{m-1,m-1}(p) \right]$$

$$= mp \left[\left(c_k - c_{k-1} + \sum_{i=k+1}^{m-1} (2c_i - c_{i-1} - c_{i+1}) + c_m - c_{m-1} \right) F_{m-1,k-1}(p) \right.$$

$$\left. + \sum_{j=k+1}^{m} \sum_{i=j}^{m-1} (2c_i - c_{i-1} - c_{i+1} + c_m - c_{m-1}) f_{m-1,j-1}(p) \right]$$

$$= mp \sum_{j=k+1}^{m} (c_j - c_{j-1}) f_{m-1,j-1}(p)$$

$$\leq mp \sum_{j=k+1}^{m} \frac{1}{2(j-1)} f_{m-1,j-1}(p)$$

$$= mp \sum_{j=k}^{m-1} \frac{1}{2j} f_{m-1,j}(p)$$

$$\leq 1$$

The proof is concluded observing that we have $\overline{T}(p, m, \mathrm{P}) = \underline{T}(p, m, \mathrm{P})$ when $k = m$.

3.10   Bibliography

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed.   New York: Dover, 1972.

[2] N. Abramson, "The ALOHA system – Another alternative for computer communications," in *Proc. Full Joint Computer Conf.*, AFIPS Conf., vol. 37, 1970, pp. 281-285.

[3] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory*, Tsahkadsor, Armenian S.S.R., 1971, Hungarian Acad. SC., pp. 23-52, 1973.

[4] D. Angelosante, E. Biglieri, M. Lops, "Low-complexity receivers for multiuser detection with an unknown number of active users," *IEEE J. Sel. Areas Comm.*, submitted 2007.

[5] S. Avestimehr, S. Diggavi and D. N. C. Tse, "A deterministic approach to wireless relay networks", in *Proc. Allerton Conf. on Communication, Control, and Computing*, Illinois, USA, Sept. 2007.

[6] E. Biglieri, J. G. Proakis, S. Shamai (Shitz), "Fading Channels: Information-Theoretic and Communication Aspects," *IEEE Trans. on Inform. Theory*, Vol. IT-44, No. 6, pp. 2619–2692, Oct. 1998.

[7] E. Biglieri, M. Lops. "Multiuser Detection in a Dynamic Environment Part I: User Identification and Data Detection," *IEEE Trans. on Inform. Theory*, Vol. IT-53, No. 9, pp. 3158–3170, Sept. 2007.

[8] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Trans. on Inform. Theory*, vol. IT-51, no.11, pp. 3992-4003, Nov. 2005.

[9] T. M. Cover, "Broadcast channels," *IEEE Trans. on Inform. Theory*, Vol. IT-18, No. 1, pp. 2-14, Jan. 1972.

[10] Ephremides, A.; Hajek, B., "Information theory and communication networks: an unconsummated union," *IEEE Trans. on Inform. Theory*, vol. IT-44, no.6, pp.2416-2434, Oct 1998

[11] A. K. Fletcher, S. Rangan and V. K. Goyal, "On-Off random access channels: a compressed sensing framework," `arXiv:0903.1022`.

[12] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 2, pp. 124-142, Mar. 1985.

[13] S. Ghez, S. Verdú, and S. Schwartz, "Stability properties of slotted ALOHA with multipacket reception capability," *IEEE Trans. Autom. Control*, vol. 33, no. 7, pp. 640–649, Jul. 1988.

[14] C. Hwang, M. Malkin, A. El Gamal and J. M. Cioffi, "Multiple-access channels with distributed channel state information," in *Proc. of IEEE Symposium on Information Theory*, pp. 1561-1565, 24-29 June, 2007, Nice, France.

[15] H. Liao, "A coding theorem for multiple access communications," in *Proc. Int. Symp. Information Theoy*, Asilomar, CA, 1972.

[16] J. L. Massey and P. Mathys, "The collision channel without feedback," *IEEE Trans. on Inform. Theory*, Vol. IT-31, No. 2, pp. 192–204, Mar. 1985.

[17] M. Medard, J. Huang, A. J. Goldsmith, S. P. Meyn, and T. P. Coleman, "Capacity of time-slotted ALOHA packetized multiple-access systems over the AWGN channel", *IEEE Trans. on Wireless Communications*, Vol. 3, No. 2, pp. 486-499, Mar. 2004.

[18] P. Minero and D. N. C. Tse "A broadcast approach to multiple access with random states," in *Proc. of IEEE Symposium on Information Theory*, pp. 2566-2570, 24-29 June, 2007, Nice, France.

[19] V. Naware, G. Mergen and L. Tong, "Stability and delay of finite user slotted ALOHA with multipacket reception", *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2636–2656, Jul. 2005.

[20] S. Shamai (Shitz), "A broadcast approach for the multiple-access slow fading channel," in *Proc. of IEEE Symposium on Information Theory*, p. 128, 25-30 June 2000, Sorrento, Italy.

[21] A. Steiner and S.Shamai (Shitz), "The broadcast approach in communications systems," *2008 IEEE 25th Convention of Electrical and Electronic Engineers in Israel*, Dec. 3-5, 2008, Eilat, Israel.

[22] S. Shamai (Shitz) and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. on Inform. Theory*, Vol. 49, No. 10, pp. 2617–2635, Oct. 2003.

Chapter 4

Control and Communications

A data rate theorem for stabilization of a linear, discrete-time, dynamical system with arbitrarily large disturbances, over a rate-limited, time-varying communication channel is presented. Necessary and sufficient conditions for stabilization are derived, their implications and relationships with related results in the literature are discussed. The proof techniques rely on both information-theoretic and control-theoretic tools.

4.1 Introduction

In modern control theory, the *data rate theorem* refers to the smallest feedback data rate above which an unstable dynamical system can be stabilized. In its scalar form, it states that a discrete linear plant of unstable mode $|\lambda| \geq 1$ can be stabilized if and only if the data rate R over the (noise free) digital feedback link satisfies the inequality $R > \log_2 |\lambda|$ bits per sample, where $\tilde{H} = \log_2 |\lambda|$ is called the intrinsic entropy rate of the plant. From its first appearance, this result has been generalized to different notions of stability and system models, and has also been extended to multi-dimensional systems [1] [3] [5] [13] [16] [21] [24] . The survey papers [2] and [17] give an historical and technical account of the various formulations.

In many engineering applications, the aim is to control one or more dynamical systems using multiple sensors and actuators communicating over digital
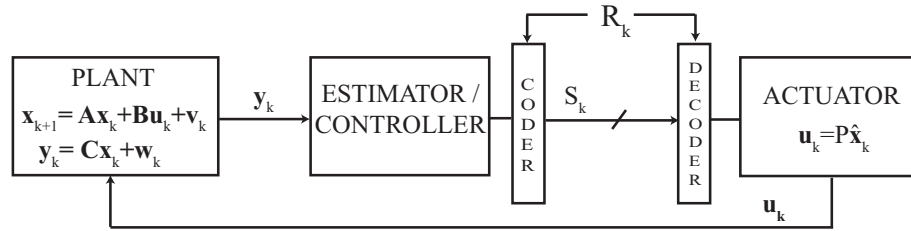
98

Figure 4.1: Feedback loop model.

links. In this framework, the data rate theorem represents a point of contact where the theories of control and communication converge, as it relates the speed of the dynamics of the plant to the information rate of the communication channel. From an information-theoretic perspective, the existence of a critical positive rate below which there does not exist any quantization and control scheme able to stabilize an unstable plant is reminiscent of Shannon's source coding theorem [20]. Stated informally, this says that if one wants to communicate with a fixed-length code over a noise free channel the output of a finite-valued stationary ergodic source process with entropy rate $H(\mathcal{X})$, then the number of bits that must be used to represent the source sequence with arbitrarily small error probability is at least $H(\mathcal{X})$. In other words, Shannon's entropy rate, representing the amount of uncertainty of the source, poses a fundamental limit on the communication rate. Similarly, the intrinsic entropy rate $\tilde{H} = \log_2 |\lambda|$ of an unstable linear dynamical system, representing the growth of the state space spanned by the open loop system, poses a fundamental limit on the minimum data rate that must be available over the feedback loop to guarantee stability.

In this work, we are concerned with the formulation of the data rate theorem over time-varying feedback channels. A motivating example is given by sensors and actuators communicating over a wireless channel for which the quality of the communication link varies over time because of random fading in the received signal. In the case of digital communication, this can reflect in a time variation of the rate supported by the wireless channel. However, if the channel variations are slow enough, transmitter and receiver can estimate the quality of the link by sending a training sequence, and can adapt the communication scheme to the channel's

condition. We ask the following question: is it possible to design a communication scheme that changes dynamically according to the channel's condition and, at the same time, is guaranteed to stabilize the system?

To answer the above question, we assume the following model. The communication channel, at any given time k, allows transmission of R_k bits without error, where R_k fluctuates randomly over time. R_k remains constant in blocks of n consecutive channel uses and then varies according to an independent and identically distributed (i.i.d) process across blocks. Furthermore, both encoder and decoder have causal knowledge of the rate supported by the communication link, see Figure 4.1. It is clear from the illustration that of the feedback loop model that the (encoded) estimated state s_k is quantized and sent to a decoder over a wireless digital link that supports error-free transmission of R_k bits per discrete unit time. We remark that such channel state information (CSI) can be obtained through feedback from the receiver to the transmitter if the fading variation is slow enough.

The model above includes the erasure channel as a special case, by allowing the rate process to have only a value $R > 0$, or zero if an erasure occurs. In this case, CSI at the transmitter can be simply obtained through one bit feedback that notifies the sender of erasures.

The model, however, does not capture the possibility of having other decoding errors beside erasures. Rather than addressing general channels with noise, our aim here is to obtain crisp results in a simple setting which can be used to understand the basic trade-offs between the intrinsic entropy rate of the system, the available rate on the communication channel, and the additional randomness due to the changing conditions of the environment. In this framework, our work directly relates to the ones in [9] [13] [16] [19] [21] and we describe this relationship in more detail next, while we refer the reader interested in more general channels with noise to the work of Sahai and Mitter [18], as well as to the works in [14] [15] [22] [23].

In an influential paper, Tatikonda and Mitter [21] have studied a model similar to ours in which the rate is fixed and system disturbances are bounded. Nair

and Evans [16] addressed the case in which the rate is still fixed, but disturbances can have an unbounded support (Gaussian disturbances are a special case of this). Finally, Martins, Dahleh, and Elia [13] considered the case of a scalar system with state feedback, random time-varying rate and bounded disturbances, and they provided necessary and sufficient conditions for m-th moment stability. In this work, we allow both the system disturbances to have unbounded support *and* the rate to vary randomly. Furthermore, the encoder has access to output feedback rather than to state feedback and we also consider the multi-dimensional case. This formulation requires the use of an adaptive quantizer, as this must be capable of tracking the state when atypically large disturbances affect the system and must dynamically adapt to the rate that is instantaneously supported by the channel. Naturally, our results can recover the ones mentioned in the above papers.

We also want to spend a few words on a different approach that has been used in the literature to model control over time-varying channels. This has a network-theoretic flavor rather than the information-theoretic one described above. In this case, the channel uncertainty is modeled using random packet dropouts. Packets are considered as single entities, each carrying the estimated state, that can be lost independently, with some probability. Furthermore, channel state information is in this case modeled as packet acknowledgement at the transmitter. An extensive survey of different works following this approach appears in [11] and we refer the reader to this work for references. The network-theoretic equivalent of the data rate theorem is the proof of existence of a *critical dropout probability* above which the closed loop system cannot be stabilized, see for example [7] [9] [12] [19].

Our present work reveals an important link between the network-theoretic, packet-loss model described above, and the information-theoretic approach. From an information-theoretic perspective, the packet loss model corresponds to an erasure channel in which the rate is infinity, with probability $1 - p$ and zero with probability p. This is because a single packet, representing the state of the system which is a real quantity, can carry an infinite amount of information, as a real number can have arbitrarily many bits within its binary expansion. Now, if we apply our results to an erasure channel, where the rate is R with probability

$1 - p$ and zero with probability p, in the high data rate limit ($R \to \infty$) this channel can be seen as communicating real numbers with random i.i.d erasures, and in this case we obtain a necessary and sufficient condition for stabilization that is the same as the one in [9], obtained under the network theoretic model, with Bernoulli packet dropouts, acknowledgement of packet reception, and Gaussian system disturbances.

The rest of the chapter is organized as follows. The main contributions are informally summarized and discussed next. Section 4.3 formally defines the problem. Section 4.4 is devoted to the proof of the necessary and sufficient conditions for stabilizability in the scalar case. These are shown via the entropy-power inequality (necessary) and the construction of an adaptive, variable length encoder (sufficiency). Section 4.5 is devoted to the more complex multi-dimensional case, for which necessary and sufficient conditions are shown to be tight in some special cases.

4.2   Overview of the results

In the scalar case, we prove that a necessary and sufficient condition to stabilize a linear system of unstable mode $|\lambda| \geq 1$ in the second moment sense over a digital link of time-varying limited rate R_k as described above, is

$$\mathbb{E}\left[\left(\frac{\lambda^2}{2^{2R}}\right)^n\right] < 1, \tag{4.1}$$

where n is the length of the block during which the rate on the digital link remains constant, and the rates R_k's are i.i.d. across blocks and distributed as a random variable R.

The condition above is amenable to the following intuitive interpretation. If no information is sent over the link during a transmission block, the estimation error at the decoder about the state of the system grows by λ^{2n}. The information sent by the encoder can reduce this error by at most 2^{2nR}, where nR is the total rate supported by the channel in a given block. However, if averaging over the fluctuation of the rate λ^{2n} exceeds 2^{2nR}, then the information sent over the channel

cannot compensate (on average) the dynamics of the system and it is not possible to stabilize the plant. Notice that if the rate is fixed over time and equal to a constant R, then the condition in (4.1) reduces to the well known inequality $R > \log_2 |\lambda|$.

Finally, it is also easy to see that when communicating over an erasure channel for which $R = \infty$ with probability $1 - p$ and $R = 0$ with probability p, then for $n = 1$ the necessary and sufficient condition for stabilization in (4.1) reduces to

$$p < \frac{1}{\lambda^2},$$

which is the same critical loss probability derived in [9] for systems with Gaussian (i.e. unbounded-support) disturbances under the network-theoretic model.

The proof of the result in (4.1) is based on an information-theoretic argument based on the entropy-power inequality (necessary condition), and on an explicit construction of an adaptive quantizer and coder-decoder pair (sufficient condition). In the latter case, the main challenge is to design a quantizer that adapts dynamically to the exogenous rate process and can handle atypically large disturbances. The construction of the coder-decoder pair is similar to the one by Nair and Evans [16]. There are, however, some key differences vis-à-vis in the way the stabilizing scheme is constructed. In [16], time is divided into cycles of fixed duration, and system state observations are quantized using a fixed number of bits, which are transmitted over the digital link for the duration of a cycle. Thus, communication between coder and decoder occurs at a fixed transmission rate. In our case, the total number of bits available in a cycle of fixed duration is random and it is not known a priori, as the rate process is known only causally at the coder and the decoder. As a consequence, the choice of an appropriate quantization rate is not immediate. Our solution consists in dividing time into cycles of *fixed duration*, but quantize the state observations using a random number of bits, which depends on the realization of the rate process. The fact that future realizations of the rate process are not known in advance is not a problem, since the quantizer we use is successively refinable, and can dynamically adapt to the rate that is instantaneously supported by the digital link. Hence, our scheme performs as if

the future realizations of the rate process were known in advance at the coder and decoder. An alternative approach consists of quantizing the observations using a fixed number of quantization points, but allowing cycles to have *variable duration.* A scheme based on this approach is outlined in section 4.5.4. Finally, we remark that, as in related works in the literature [13] [16] [21], the construction provided in this work relies on the crucial assumption that the coder and decoder can agree on the initial values of the internal states through an a priori iterative communication process.

The extension of the analysis to multi-dimensional linear systems entails the difficulty of the rate allocation to the different unstable modes. In this case, we derive necessary conditions for second moment stabilizability, which define a region with a special polymatroid structure. When the rate is fixed and equal to a constant R, the necessary conditions reduce to

$$\tilde{H} := \sum_{|\eta_i| \geq 1} \log_2 |\eta_i| < R,$$

where $\eta_1, \ldots, \eta_n$ are the open loop eigenvalues (raised to their corresponding algebraic multiplicities). Again, this recovers the well known data rate theorem for vector systems with deterministic rate [16], [21]. Finally, as in the scalar case, in the high data rate limit over an erasure channel, we also recover the necessary condition on the critical dropout probability of [9].

Finally, we provide a general coder-decoder construction for vector systems and show that this is optimal in some limiting cases. For some specific rate distributions, however, it is possible to design more efficient schemes. This latter point is shown by considering stabilization over a binary erasure channel, for which a better scheme is proposed.

4.3   Problem formulation

In the sequel, the following notation is used: vectors are written in bold-faced type and sequences $\{a_i\}_{i=0}^k$ are denoted as a_0^k; expectation with respect to

the random variable X is written as $\mathbb{E}_X[\cdot]$, the differential entropy of a continuous random vector $\mathbf{X}$ as $h(\mathbf{X}) = -\mathbb{E}[\ln f_{\mathbf{X}}(\mathbf{x})]$ and the entropy of a discrete random vector $\mathbf{X}$ as $H(\mathbf{X}) = -\mathbb{E}[\ln P(\mathbf{X})]$; the set of non-negative integers as $\mathbb{N}$, the positive integers as $\mathbb{Z}_+$, and the rational numbers as $\mathbb{Q}$; finally, the cardinality of a finite set S is denoted as $|S|$.

Consider the partially-observed, discrete-time state-space unstable stochastic linear system

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{v}_k, \ \mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{w}_k, \ \forall k \in \mathbb{N}, \qquad (4.2)$$

where $\mathbf{x}_k \in \mathbb{R}^d$ is the state process, $\mathbf{u}_k \in \mathbb{R}^m$ is the control input, $\mathbf{v}_k \in \mathbb{R}^d$ process disturbance, the measurement $\mathbf{y}_k$ and measurement noise $\mathbf{w}_k$ are random vectors in $\mathbb{R}^p$. Suppose $\mathbf{A}$ is uniquely composed by unstable modes (having magnitude greater or equal to unity). No Gaussian assumptions are made on the initial condition $\mathbf{x}_0$ and the disturbances, but the following is assumed to hold:

A0. $(\mathbf{A}, \mathbf{B})$ is reachable and $(\mathbf{C}, \mathbf{A})$ observable.

A1. $\mathbf{x}_0, \mathbf{v}_k$ and $\mathbf{w}_j$ are mutually independent for all $k, j \in \mathbb{N}$.

A2. $\exists \epsilon > 0$ such that $\mathbf{x}_0, \mathbf{v}_k$ and $\mathbf{w}_j$ have uniformly bounded $(2 + \epsilon)$-th absolute moments over $k \in \mathbb{N}$.

A3. $\inf_{k \in \mathbb{N}} h(\mathbf{v}_k) > -\infty$. Thus, $\exists \beta > 0$ such that $e^{\frac{2}{d}h(\mathbf{v}_k)} > \beta$ for all $k \in \mathbb{N}$ and $\mathbf{v}_k \in \mathbb{R}^d$.

Suppose that coder and decoder are connected by a *time-varying* digital link, see Figure 4.1. The transmission rate supported by the digital link is assumed constant over blocks of $n \in \mathbb{N}$ channel uses but changes independently from block to block according to a given probability distribution. Formally, at time $k \in \mathbb{N}$ the digital link is an identity map on an alphabet $\{1, \cdots, |S_k|\}$; $\log_2 |S_k|$ denotes the transmission rate supported by the digital link, and coincides with R_j if and only if $k \in \{jn, jn+1, \cdots, (j+1)n - 1\}$. At time $k \in \{jn, jn+1, \cdots, (j+1)n - 1\}$, coder and decoder know R_j, while the realization of the rate process in future blocks, $\{R_i\}_{i=j+1}^{\infty}$, is unknown to them. The $\{R_j\}_{j \in \mathbb{N}}$ are i.i.d random variables

distributed as R, where R is an integer-valued random variable taking values on $\mathcal{R} \subseteq \mathbb{N}$. We denote by r_{min} the minimum value in the set $\mathcal{R}$.

This definition of the rate process is motivated by communication over wireless channels. In fact, the rate supported by a block fading wireless channel can be modeled as a random variable, since this is a function of the (random) channel gain that attenuates the transmitted signal. The block fading model captures a fading scenario where the fading channel state remains invariant over a block of time but changes from block to block. If the fading variation is slow enough, feedback from the receiver to the transmitter can be used to acquire channel state information. If the channel state information is known, then the rate supported by the channel is also known at both transmitter and receiver. Finally, the rate can be modeled as an i.i.d. random process across the channel blocks if we assume that block lengths are similar to coherence time intervals (length of time over which the channel's statistical properties do not change) of the channel. For example, the i.i.d. assumption is valid for a slow frequency hopped time division multiple access channel.

Example 4.3.1. *We call* erasure channel *a digital link where $R = r$ with probability $1 - p$ and $R = 0$ with probability p, for some nonnegative integer r and $p \in (0, 1)$. If $r = 1$, we call the channel* binary erasure channel.

Each transmitted symbol can depend on all past and present measurements, the present channel state and the past symbols,

$$S_k = g_k \left(\mathbf{y}_0^k, S_0^{k-1}, R_j \right),$$
$$k \in \{jn, \ldots, (j+1)n - 1\}, \forall j \in \mathbb{N},$$

where $g_k(\cdot)$ is the coder mapping at time k. The control sequence, on the other hand, can depend on all past and present channel symbols

$$\mathbf{u}_k = \delta_k \left(S_0^k \right), \quad \forall k \in \mathbb{N},$$

where $\delta_k(\cdot)$ is the controller mapping at time k.

We want to construct a coder-decoder pair which stabilizes the plant in the mean square sense

$$\sup_{k \in \mathbb{N}} \mathbb{E}\left[\|\ \mathbf{x}_k\ \|^2\right] < \infty, \tag{4.3}$$

using the finite data rate provided by the time-varying digital feedback link.

4.4   Scalar systems

In this section it is assumed that the plant in (4.2) is scalar and has a representation of the following type:

$$x_{k+1} = \lambda x_k + u_k + v_k, \ y_k = x_k + w_k, \ \forall k \in \mathbb{N}, \tag{4.4}$$

where $|\lambda| \geq 1$, so that the system is unstable. The result for the scalar case is now stated:

Theorem 4.4.1. *Under assumptions **A0.-A3.** above, necessary and sufficient condition for stabilizing the plant* (4.4) *in the mean square sense* (4.3) *is that*

$$\mathbb{E}\left[\left(\frac{|\lambda|^2}{2^{2R}}\right)^n\right] < 1, \tag{4.5}$$

where n is the length of the channel block with the same rate.

4.4.1   Necessity

In order to prove the statement, we find a lower bound for the second moment of the state, and show that (4.5) is a necessary condition for this lower bound to be finite. We focus on the times $k = jn$ with $j \in \mathbb{N}$, i.e. on the beginning of each channel block. Let $\bar{S}_j = \{S_0, \ldots, S_{(j+1)n-1}\}$, denote the symbols sent over the noiseless channel until the end of the j-th channel block. By iteration of (4.4), we have

$$x_{(j+1)n} = \lambda^n x_{jn} + \sum_{k=jn}^{(j+1)n-1} \lambda^{(j+1)n-1-k}[\delta_k(S_0^k) + v_k].$$

Let $n_j = \frac{1}{2\pi e}\mathbb{E}_{\bar{S}_{j-1}}\left[e^{2h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1})}\right]$ be the conditional entropy power of x_{jn} conditioned on the event $\{\bar{S}_{j-1} = \bar{s}_{j-1}\}$, averaged over all possible $\bar{s}_{j-1}$. The second moment of x_{jn} is lower bounded by n_j:

$$
\begin{aligned}
\mathbb{E}[x_{jn}^2] &= \mathbb{E}_{\bar{S}_{j-1}}\left[\mathbb{E}[x_{jn}^2|\bar{S}_{j-1} = \bar{s}_{j-1}]\right] \\
&= \frac{1}{2\pi e}\mathbb{E}_{\bar{S}_{j-1}}\left[e^{\ln(2\pi e\mathbb{E}[x_{jn}^2|\bar{S}_{j-1}=\bar{s}_{j-1}])}\right] \\
&\geq \frac{1}{2\pi e}\mathbb{E}_{\bar{S}_{j-1}}\left[e^{2h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1})}\right] \\
&= n_j,
\end{aligned}
$$

where the inequality follows from the maximum entropy theorem [4, Theorem 9.4.1]. It follows that a necessary condition for (4.3) to hold is that $\sup_{j\in\mathbb{N}} n_j < \infty$. We now complete the proof by showing that this necessary condition is violated whenever (4.5) does not hold. We make use of the following technical lemma (proved in the Appendix A),

Lemma 4.4.2. *For all non-negative random variables R_j, the following inequality holds*

$$
\mathbb{E}_{\bar{S}_j|\bar{S}_{j-1},R_j}\left[e^{2h(x_{jn}|\bar{S}_j=\bar{s}_j)}\right] \geq \frac{1}{2^{2nR_j}}e^{2h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1})}.
$$

First, we show that n_j evolves according to a recursive equation. Using standard properties of entropy [4] (translation invariance, conditional version of

entropy power inequality), and assumptions **A1.** and **A3.**, it follows that

$$
\begin{aligned}
\mathbb{E}_{\bar{S}_j}\left[e^{2h(x_{(j+1)n}|\bar{S}_j=\bar{s}_j)}\right] &= \\
&= \mathbb{E}_{\bar{S}_j}\left[e^{2h(\lambda^n x_{jn}+\sum_{k=jn}^{(j+1)n-1}\lambda^{(j+1)n-1-k}v_k|\bar{S}_j=\bar{s}_j)}\right] \\
&\geq \lambda^{2n}\mathbb{E}_{\bar{S}_j}\left[e^{2h(x_{jn}|\bar{S}_j=\bar{s}_j)}\right] + \gamma \\
&= \lambda^{2n}\mathbb{E}_{\bar{S}_{j-1},R_j}\left[\mathbb{E}_{\bar{S}_j|\bar{S}_{j-1},R_j}\left[e^{2h(x_{jn}|\bar{S}_j=\bar{s}_j)}\right]\right] + \gamma \\
&\geq \lambda^{2n}\mathbb{E}_{\bar{S}_{j-1},R_j}\left[\frac{1}{2^{2nR_j}}e^{2h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1})}\right] + \gamma \\
&= \lambda^{2n}\mathbb{E}_{R_j}\left[\frac{1}{2^{2nR_j}}\right]\mathbb{E}_{\bar{S}_{j-1}}\left[e^{2h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1})}\right] + \gamma,
\end{aligned}
$$

wherein the second inequality follows from assumption **A3.** above, i.e. $e^{2h(v_k)} > \beta$. The constant γ is defined as $\gamma := \sum_{k=jn}^{(j+1)n-1}\lambda^{2[(j+1)n-1-k]}\beta$. Finally, the last inequality follows from Lemma 4.4.2 and the fact that R_j is independent of x_{jn} and $\bar{S}_{j-1}$. Thus, using the fact that the rate process is i.i.d., we have

$$
n_{j+1} \geq \mathbb{E}\left[\left(\frac{\lambda^2}{2^{2R}}\right)^n\right]n_j + \frac{\gamma}{2\pi e}.
$$

Therefore, $\mathbb{E}\left[\left(\frac{\lambda^2}{2^{2R}}\right)^n\right] \geq 1$ implies that $\sup_{j\in\mathbb{N}} n_j = \infty$.

4.4.2 Sufficiency

We first describe the adaptive quantizer that is at the base of the constructive scheme. A fundamental property of this quantizer is then stated as a lemma, whose proof appears in [16].

Quantizer

The quantizer partitions the real line into non-uniform regions, and a parameter $\rho > 1$ determines the speed at which the quantizer range increases. The quantizer generates 2^ν, $\nu \geq 0$, quantization intervals labeled from left to right by $I_\nu(0), \ldots, I_\nu(2^\nu - 1)$. Let $I_0(0) := (-\infty, \infty)$, $I_1(0) := (-\infty, 0]$ and $I_1(1) := (0, \infty)$. If $\nu \geq 2$ the quantization intervals are generated by

- partitioning the set $[-1, 1]$ into $2^{\nu-1}$ intervals of equal length,

- partitioning the sets $(\rho^{i-2}, \rho^{i-1}]$, $[-\rho^{i-2}, -\rho^{i-1})$ into $2^{\nu-1-i}$ intervals of equal length, $i \in \{2, \ldots, \nu - 1\}$.

The two open sets $(-\infty, -\rho^{\nu-2}]$ and $(\rho^{\nu-2}, \infty)$ are respectively the leftmost and rightmost intervals of the quantizer. Let

- $\kappa_\nu(\omega)$ be half-length of interval $I_\nu(\omega)$ for $\omega \in \{1, \ldots, 2^\nu - 2\}$, be equal to $\rho^\nu - \rho^{\nu-1}$ when $\omega = 2^\nu - 1$ and equal to $-(\rho^\nu - \rho^{\nu-1})$ when $\omega = 0$.

- $q_\nu(x) := \bar{\omega}_\nu(\omega)$ be midpoint of interval $x \in I_\nu(\omega)$ for $\omega \in \{1, \ldots, 2^\nu - 2\}$, be equal to ρ^ν when $\omega = 2^\nu - 1$ and equal to $-\rho^\nu$ when $\omega = 0$.

A property of this construction is that for $\nu \geq 2$ the quantization intervals $I_\nu(\cdot)$ can be generated recursively starting from $q_2(\cdot)$. In fact, for any integer $i \geq 2$ the quantizer intervals for $q_{i+1}(\cdot)$ are formed by partitioning each bounded interval $I_i(\omega)$ $\omega \in \{1, \ldots, 2^i - 2\}$ into two uniform subintervals, and partitioning the semi-infinite interval $I_i(0) = (-\infty, -\rho^{i-2}]$ into two intervals $I_{i+1}(0) = (-\infty, -\rho^{(i+1)-2}]$ and $I_{i+1}(1) = (-\rho^{(i+1)-2}, -\rho^{i-2}]$ and, similarly, partitioning the semi-infinite interval $I_i(2^i - 1) = (\rho^{i-2}, \infty)$ into two intervals $I_{i+1}(2^{i+1} - 2) = (\rho^{i-2}, \rho^{(i+1)-2}]$ and $I_{i+1}(2^{i+1} - 1) = (\rho^{(i+1)-2}, +\infty)$.

Given a real-valued random variable X, if its realization x is in $I(\omega)$ for some $\omega \in \{0, \ldots, 2^\nu - 1\}$, then the quantizer approximates x with $\bar{\omega}_\nu(\omega)$. The quantization error is not uniform over $x \in \mathbb{R}$, but is bounded by $\kappa_\nu(\omega)$ for all $\omega \in \{1, \ldots, 2^\nu - 2\}$. A fundamental property of the quantizer is that the average quantization error diminishes like the inverse square of the number of levels, $2^{-2\nu}$. More precisely, if the $(2 + \epsilon)$-th moment of X is bounded for some $\epsilon > 0$, then an upper bound of the second moment of the estimation error decays as $2^{-2\nu}$. The higher moment of X is useful to bound the estimation error (using Chebyshev's inequality) when X lies in one of the two open intervals $(\rho^{\nu-1}, \infty)$ and $(-\infty, -\rho^{\nu-1}]$.

Let $L \in \mathbb{R}_+$ be any random variable, define the functional

$$M_\epsilon[X, L] \equiv \mathbb{E}[L^2 + X^{2+\epsilon} L^{-\epsilon}]. \tag{4.6}$$

The functional $M_\epsilon[X, L]$ is an upper bound to the second moment of X:

$$\mathbb{E}[X^2] = \mathbb{E}[X^2(1_{|X| \leq L} + 1_{|X| > L})] \leq M_\epsilon[X, L]. \tag{4.7}$$

Define the conditional version of $M_\epsilon[X, L]$ given a random variable R as $M_\epsilon[X, L|R]$ $\equiv \mathbb{E}[L^2 + X^{2+\epsilon} L^{-\epsilon}|R]$. The fundamental property of the quantizer described above is given by the following result:

Lemma 4.4.3. *[16, Lemma 5.2] Let X and $L > 0$ and be random variables with $\mathbb{E}[X^{2+\epsilon}] < \infty$ for some $\epsilon > 0$, and $n \in \mathbb{N}$. If $\rho > 2^{2/\epsilon}$, then for any $R \in \mathbb{N}$ the quantization errors $X - Lq_{nR}(X/L)$ satisfy*

$$M_\epsilon[X - Lq_{nR}(X/L), L\kappa_{nR}(\omega)] \leq \frac{\zeta}{2^{2nR}} M_\epsilon[X, L],$$

where $\omega \in \{0, \dots, 2^{nR} - 1\}$ is the index of the quantizer level $q_{nR}(X/L)$, and ζ is a constant greater than 2 determined only by ϵ and ρ.

Next, the coder and decoder are described.

Coder

The first stage of the encoding process consists of computing the linear minimum variance estimator of the plant state based on the previous measurements and control sequences. The filter process satisfies a recursive equation of the same form as (4.4), namely

$$\bar{x}_{k+1} = \lambda \bar{x}_k + u_k + z_k, \quad \forall k \in \mathbb{N}. \tag{4.8}$$

where $z_k := (y_k - \bar{x}_k) l_k$ is the product of the innovation $y_k - \bar{x}_k$ and the appropriate optimal gain l_k. The $(2 + \epsilon)$-th moment of z_k can be shown to be bounded, under assumption **A2.**, for any finite k. From the orthogonality principle the stability of $\bar{x}$ is equivalent to that of x. The output $\bar{x}_k$ of the filter (or a function of it) must be transmitted using the finite number of bits supported on the digital channel. Coder and decoder share a state estimator $\hat{x}_k$ based uniquely on the symbols sent over the digital link. Since $\hat{x}_k$ is available both at the coder and decoder, while

the minimum variance estimator is available at the coder only, the encoder uses the quantizer described in the previous section to encode the error between $\bar{x}_k$ and $\hat{x}_k$. The error is scaled by an appropriate coefficient and then recursively encoded using the quantizer in section 4.4.2. An accurate approximation of the error is obtained by transmitting the quantization index across many channel blocks. The fact that the random rate available at future times is not known in advance is not a problem, as the quantizer is successively refinable and can dynamically adapt to the rate that is instantaneously supported by the channel. By transmitting for a large enough number of blocks, the error between the two estimators can be kept bounded.

Define the coder error at time $k \in \mathbb{N}$ as $f_k = \bar{x}_k - \hat{x}_k$. Times $k \in \mathbb{N}$ are divided into *cycles* $\{jn\tau, \ldots, (j+1)n\tau\}$, $j \in \mathbb{N}$, of integer duration $n\tau$, $\tau \in \mathbb{Z}_+$. Notice that each cycle consists of τ channel blocks.

At time $k = jn\tau$, just before the start of the j-th cycle, the coder sets the quantization rate equal to $nR_{j\tau}$, i.e. the rate in the first channel block in the j-th cycle, and computes

$$\bar{\omega}_{nR_{j\tau}}(\omega_{jn\tau}) = q_{nR_{j\tau}}\left(f_{jn\tau}/l_j\right),$$

where l_j is a scaling factor updated at the beginning of each cycle. This factor is used to scale $f_{jn\tau}$ close to the origin, where the quantizer provides better estimates. The index $\omega_{jn\tau}$ of the quantization level is converted into a string of $nR_{j\tau}$ bits and transmitted using the n channel uses of the $j\tau$-th channel block. Denote by $I_{nR_{j\tau}}(\omega_{jn\tau})$ the quantization interval labeled by $\omega_{jn\tau}$. After the first n transmissions in the cycle, coder and decoder agree on the fact that $f_{jn\tau}/l_j \in I_{nR_{jn}}(\omega_{jn\tau})$. The remaining $(n-1)\tau$ transmissions in the cycle are devoted to reducing the size of the uncertainty interval $I_{nR_{j\tau}}(\omega_{jn\tau})$.

At time $k = jn\tau + n$, the rate $R_{j\tau+1}$ supported during the next channel block becomes known at both coder and decoder. Thus, coder and decoder divide up $I_{nR_{j\tau}}(\omega_{jn\tau})$ into $2^{nR_{j\tau+1}}$ sub-intervals in the manner described above (uniform partitions of bounded intervals and exponential partition of semi-infinite intervals), sequentially generating the partitions $I_{nR_{j\tau}+nR_{j\tau+1}}(\cdot) \subseteq I_{nR_{j\tau}}(\omega_{jn\tau})$ of the quantizer $q_{nR_{j\tau}+nR_{j\tau+1}}\left(f_{jn\tau}/l_j\right)$.

Then, the coder sends to the decoder the index of the sub-interval containing $f_{jn\tau}/l_j$. At the end of the second channel block in the cycle, coder and decoder agree on the fact that $f_{jn\tau}/l_j \in I_{nR_{j\tau}+nR_{j\tau+1}}(\omega_{jn\tau+n})$.

Continue this process until the end of the τ-th channel block. After receiving the last sequence of bits, the decoder computes the final uncertainty interval $I_{\nu_j}(\omega_{(j+1)n\tau-n})$, corresponding to the uncertainty set formed by the quantizer $q_{\nu_j}(f_{jn\tau}/l_j)$, where the random variable

$$\nu_j := nR_{j\tau} + nR_{j\tau+1} + \ldots + nR_{(j+1)\tau-1}$$

indicates the cumulative number of bits sent in the j-th cycle.

Before the beginning of the $(j+1)$-th cycle, the coder updates the state estimator as follows,

$$\hat{x}_{(j+1)n\tau} = \lambda^{n\tau}[\hat{x}_{jn\tau} + l_j q_{\nu_j}(f_{jn\tau}/l_j)]+$$
$$+ \sum_{k=jn\tau}^{(j+1)n\tau-1} \lambda^{(j+1)n\tau-1-k} P\hat{x}_k, \tag{4.9}$$

where

$$\hat{x}_{k+1} = (\lambda + P)\hat{x}_k, \quad \forall k \in \{jn\tau, \ldots, j(n+1)\tau - 2\}, \tag{4.10}$$

and $\hat{x}_0 = 0$. P is the certainty-equivalent control coefficient such that $|\lambda + P| < 1$. Finally, the scaling coefficient l_j is updated as follows

$$l_{j+1} = \max\{\sigma, l_j|\lambda|^{n\tau}\kappa_{\nu_j}(\omega_{(j+1)n\tau-n})\}, \tag{4.11}$$

with $l_0 = \sigma$ and where $\sigma^{2+\epsilon}$ is a uniform bound for the $(2+\epsilon)$-moment of

$$g_j := \sum_{i=1}^{n\tau-1} \lambda^{n\tau-i} z_{jn\tau+i}, \quad j \in \mathbb{N}. \tag{4.12}$$

Decoder

At time $k = jn\tau$ coder and decoder are synchronized and have common knowledge of the state estimator $\hat{x}_{jn\tau}$. During times $jn\tau, \ldots, (j+1)n\tau - 1$, the decoder sends to the plant a certainty-equivalent control signal

$$u_k = P\hat{x}_k, \quad \forall k \in \{jn\tau, \ldots, (j+1)n\tau - 1\}, \tag{4.13}$$

where $\hat{x}_k$ is updated as in (4.10). At the end of the each channel block in the j-th cycle, the decoder receives estimates of the states in the way described above.

At time $(j+1)n\tau - 1$, once computed $q_{\nu_j}(f_{jn\tau}/l_j)$ the decoder updates the estimator $\hat{x}_{(j+1)n\tau}$ using (4.9). Synchronism between coder and decoder is ensured by the fact that the initial value $\hat{x}_0$ is set equal to zero at both coder and decoder, and by the fact that the digital link is noiseless.

Analysis

In this section it is shown that the coder-decoder pair described above ensures that the second moment of $\bar{x}$ is bounded if (4.5) is satisfied.

The analysis is developed in three steps. First, we show that f_k is bounded for all times $k = jn\tau$, $j \in \mathbb{N}$, i.e. the beginning of each cycle. Next, the analysis is extended to all $k \in \mathbb{N}$. Finally, the stability of f_k for all $k \in \mathbb{N}$ is shown to imply that $\bar{x}$ (and so x) is bounded.

First we show that the coder error f_k is bounded in the mean square sense for all times $k = jn\tau$, $j \in \mathbb{N}$. Instead of looking at $\mathbb{E}[|f_{jn\tau}|^2]$, it is more convenient to consider the functional $M_\epsilon[X, L]$ defined in (4.6), with $X = f_{jn\tau}$ and $L = l_j$. Thus, let

$$\theta_j := M_\epsilon[f_{jn\tau}, l_j] \equiv \mathbb{E}[l_j^2 + f_{jn\tau}^{2+\epsilon} l_j^{-\epsilon}].$$

Equation (4.7) implies that $\mathbb{E}[f_{jn\tau}^2] \leq \theta_j$. Therefore, it suffices to show that $\sup_{j \in \mathbb{N}} \theta_j < \infty$.

Substituting (4.13) into (4.8), and iterating over the duration of a cycle, we

have

$$\bar{x}_{(j+1)n\tau} = \lambda^{n\tau}\bar{x}_{jn\tau} + g_j + \sum_{k=jn\tau}^{(j+1)n\tau-1} \lambda^{(j+1)n\tau-1-k}(P\hat{x}_k), \qquad (4.14)$$

where g_j is defined in (4.12). Subtracting (4.9) from (4.14), we have

$$f_{(j+1)n\tau} = \bar{x}_{(j+1)n\tau} - \hat{x}_{(j+1)n\tau}$$
$$= \lambda^{n\tau}[f_{jn\tau} - l_j q_{\nu_j}(f_{jn\tau}/l_j)] + g_j.$$

Notice that, by assumption **A2.**, the $(2+\epsilon)$-th moment of $f_{jn\tau}$ is bounded for any finite $j \in \mathbb{N}$. Next, $f_{(j+1)n\tau}$ is used to derive an expression for θ_{j+1}. From the inequality $(|x| + |y|)^\alpha \leq 2^{\alpha-1}(|x|^\alpha + |y|^\alpha) \; \forall \alpha > 0$, we obtain

$$f_{(j+1)n\tau}^{2+\epsilon} \leq \phi\left(|\lambda^{n\tau}|^{2+\epsilon}\,|f_{jn\tau} - l_j q_{\nu_j}(f_{jn\tau}/l_j)|^{2+\epsilon}\right.$$
$$\left. + g_j^{2+\epsilon}\right),$$

with $\phi = 2^{1+\epsilon}$. Dividing by l_{j+1}^ϵ, taking expectations and using (4.11), we have

$$\mathbb{E}[f_{(j+1)n\tau}^{2+\epsilon}\,l_{j+1}^{-\epsilon}]$$
$$\leq \phi\left(|\lambda^{n\tau}|^{2+\epsilon}\mathbb{E}\left[\frac{|f_{jn\tau} - l_j q_{\nu_j}(f_{jn\tau}/l_j)|^{2+\epsilon}}{l_{j+1}^\epsilon}\right] + \mathbb{E}\left[\frac{g_j^{2+\epsilon}}{l_{j+1}^\epsilon}\right]\right)$$
$$\leq \phi\left(|\lambda^{n\tau}|^{2+\epsilon}\mathbb{E}\left[\frac{|f_{jn\tau} - l_j q_{\nu_j}(f_{jn\tau}/l_j)|^{2+\epsilon}}{[l_j\,|\lambda^{n\tau}|\kappa_{\nu_j}(\omega_{(j+1)n\tau-n})]^\epsilon}\right] + \mathbb{E}\left[\frac{g_j^{2+\epsilon}}{\sigma^\epsilon}\right]\right)$$
$$= \phi\left(\lambda^{2n\tau}\mathbb{E}\left[\frac{|f_{jn\tau} - l_j q_{\nu_j}(f_{jn\tau}/l_j)|^{2+\epsilon}}{[l_j\,\kappa_{\nu_j}(\omega_{(j+1)n\tau-n})]^\epsilon}\right] + \mathbb{E}\left[\frac{g_j^{2+\epsilon}}{\sigma^\epsilon}\right]\right). \qquad (4.15)$$

Next, observe that

$$\mathbb{E}[l_{j+1}^2] \leq \sigma^2 + \lambda^{2n\tau}\mathbb{E}\left[|l_j\kappa_{\nu_j}(\omega_{(j+1)n\tau-n})|^2\right]. \qquad (4.16)$$

Summing (4.15) and (4.16), using $\mathbb{E}[g_j^{2+\epsilon}] \leq \sigma^{2+\epsilon}$ and the definition of θ_j, we have

$$
\begin{aligned}
\theta_{j+1} &\leq \phi\left(2\sigma^2 + \right.\\
&\quad + \lambda^{2n\tau}\mathbb{E}\left[\frac{|f_{jn\tau} - l_j q_{\nu_j}(f_{jn\tau}/l_j)|^{2+\epsilon}}{[l_j \kappa_{\nu_j}(\omega_j)]^\epsilon} + |l_j \kappa_{\nu_j}(\omega_j)|^2\right]\right)\\
&= \phi\left(2\sigma^2 + \right.\\
&\quad + \lambda^{2n\tau}\mathbb{E}_{\nu_j}\left[M_\epsilon\left[f_{jn\tau} - l_j q_{\nu_j}\left(\frac{f_{jn\tau}}{l_j}\right), l_j \kappa_{\nu_j}(\omega_j)\middle|\nu_j\right]\right]\right)\\
&\leq \phi 2\sigma^2 + \phi\lambda^{2n\tau}\mathbb{E}_{\nu_j}\left[\frac{\zeta}{2^{2\nu_j}}M_\epsilon\left[f_{jn\tau}, l_j\right]\right]\\
&= \phi 2\sigma^2 + \phi\zeta\left(\mathbb{E}\left[\frac{\lambda^{2n}}{2^{2nR}}\right]\right)^\tau \theta_j,
\end{aligned}
$$

where the second inequality follows from Lemma (4.4.3), and the last equality uses the fact that the rate process is i.i.d. and that f_{jn} and l_j are independent of $R_{j\tau}, R_{j\tau+1}, ..., R_{(j+1)\tau-1}$ because of the causality constraint. Therefore, θ_j evolves according to the following recursive equation

$$
\theta_{j+1} \leq \phi 2\sigma^2 + \phi\zeta\left(\mathbb{E}\left[\left(\frac{\lambda^2}{2^{2R}}\right)^n\right]\right)^\tau \theta_j.
$$

It follows that if $\mathbb{E}\left[\left(\frac{\lambda^2}{2^{2R}}\right)^n\right] < 1$, then by making τ sufficiently large we can ensure that the coefficient of θ_j is strictly less than 1. Thus we have established that θ_j remains bounded in the limit of j going to infinity and therefore $\sup_{j\in\mathbb{N}} \theta_j < \infty$. Hence, from (4.7) it follows that $\sup_{j\in\mathbb{N}}\mathbb{E}[f_{jn\tau}^2] < \infty$.

Next, for any $k \in \{0, \ldots, n-1\}$ the triangle inequality implies $|f_{jn+k}| \leq |\lambda|^k|f_{jn}| + \sum_{i=0}^{k-1}|\lambda^{k-1-i}P|\,|z_{jn+k}|$, so the error f_k is bounded for all $k \in \mathbb{N}$. Finally, by rewriting (4.8) as $\bar{x}_{k+1} = (\lambda + P)\bar{x}_k - Pf_k + z_k$, the fact that f_k and z_k are bounded and that $|\lambda + P| < 1$ ensures that $\mathbb{E}[\bar{x}_k^2] < \infty$ for all $k \in \mathbb{N}$.

4.5 Vector Systems.

In this section, we consider the case of multi-dimensional unstable linear systems. A necessary condition for stabilizability is derived using an information-

theoretic approach. It is proved that the stabilizability region is contained inside a polytope with a polymatroid structure. A sub-optimal coder-decoder construction is provided and its optimality is shown in some limiting cases. The main difficulty in stabilizing a multi-dimensional system over time-varying channels consists of allocating optimally the bits to each unstable sub-system. The scheme proposed can be applied to *any* rate distribution. For some specific rate distributions, however, it is possible to design more efficient schemes. We illustrate this point at the end of this section, studying the specific problem of stabilization over a binary erasure channel, for which a better scheme is proposed.

4.5.1 Real Jordan form

As usual, it is convenient to put $\mathbf{A}$ into real Jordan canonical form [10] so as to decouple its dynamical modes. Denote the system matrix in real Jordan canonical form as $\mathbf{J}$. The matrices $\mathbf{J}$ and $\mathbf{A}$ are related via a similarity matrix $\mathbf{T}$ such that $\mathbf{T}^{-1}\mathbf{J}\mathbf{T} = \mathbf{A}$. Let $\lambda_1, \ldots, \lambda_u \in \mathbb{C}$ be the distinct unstable eigenvalues (if λ_i is non-real, we exclude from this list the complex conjugates λ_i^*) of $\mathbf{A}$, and let m_i be the algebraic multiplicity of each λ_i. The real Jordan canonical form $\mathbf{J}$ then has the block diagonal structure $\mathbf{J} = \text{diag}(\mathbf{J}_1, \ldots, \mathbf{J}_u) \in \mathbb{R}^{d \times d}$, where the block $\mathbf{J}_i \in \mathbb{R}^{\mu_i \times \mu_i}$ and $\det \mathbf{J}_i = \lambda_i^{\mu_i}$, with

$$\mu_i = \begin{cases} m_i & \text{if } \lambda_i \in \mathbb{R} \\ 2m_i & \text{otherwise.} \end{cases}$$

As $\mathbf{A}$ is uniquely composed by unstable systems, we have that $\sum_{i=1}^{u} \mu_i = d$. Let $\mathcal{U} := [1, \ldots, u]$ denote the index set of unstable systems. Then, the dynamical system equation can be written as

$$\mathbf{x}_{k+1} = \mathbf{J}\mathbf{x}_k + \mathbf{T}\mathbf{B}\mathbf{u}_k + \mathbf{T}\mathbf{v}_k \ \in \mathbb{R}^d, \ \mathbf{y}_k = \mathbf{C}\mathbf{T}^{-1}\mathbf{x}_k + \mathbf{w}_k \ \in \mathbb{R}^p, \tag{4.17}$$

with $\mathbf{x}_k = \left[\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(u)T}\right]^T \in \mathbb{R}^d$, and where each sub-system $\mathbf{x}_k^{(i)}$ evolves according to

$$\mathbf{x}_{k+1}^{(i)} = \mathbf{J}_i\mathbf{x}_k^{(i)} + (\mathbf{T}\mathbf{B}\mathbf{u}_k)^{(i)} + (\mathbf{T}\mathbf{v}_k)^{(i)} \ \in \mathbb{R}^{\mu_i}, \ i \in \mathcal{U}.$$

As the states of (4.17) and (4.2) are related through the transformation matrix $\mathbf{T}$, in the following we will assume that the system evolves according to (4.17).

4.5.2 Necessity

Theorem 4.5.1. *Under assumptions **A0.-A3.** above, necessary condition for stabilizability of the system in (4.17) in the mean square sense (4.3) is that $(\log_2 |\lambda_1|, \ldots, \log_2 |\lambda_u|) \in \mathbb{R}_+^u$ satisfy, for all $s_i \in \{0, \ldots, m_i\}$ and $i \in \mathcal{U}$,*

$$\sum_{i \in \mathcal{U}} a_i s_i \log_2 |\lambda_i| < -\frac{\mu'(\mathbf{s})}{2n} \log_2 \mathbb{E}\left[2^{-\frac{2n}{\mu'(\mathbf{s})}R}\right], \tag{4.18}$$

wherein $\mu'(\mathbf{s}) \equiv \sum_{i \in \mathcal{U}} a_i s_i$, and $a_i = 1$ if $\lambda_i \in \mathbb{R}$, and $a_i = 2$ otherwise.

The following example highlights the special geometric structure of the region defined by (4.18):

Example 4.5.1. *Consider a two-mode system with two distinct eigenvalues (λ_1, λ_2), where λ_1 is complex and has dimensionality $m_1 = 1$ (so $\mu_1 = 2$) while λ_2 is real and has dimensionality $m_2 = \mu_2 = 1$. Suppose that the digital channel in the feedback link is an erasure channel. Computing the bounds in (4.18) we obtain the following necessary conditions on $(\log_2 |\lambda_1|, \log_2 |\lambda_2|)$ for stabilizability:*

$$2 \log_2 |\lambda_1| < -\frac{1}{n} \log_2 \left[p + (1-p)2^{-nr}\right],$$
$$\log_2 |\lambda_2| < -\frac{1}{2n} \log_2 \left[p + (1-p)2^{-2nr}\right], \tag{4.19}$$
$$2 \log_2 |\lambda_1| + \log_2 |\lambda_2| < -\frac{3}{2n} \log_2 \left[p + (1-p)2^{-2n/3r}\right].$$

In general, these three bounds define a pentagon in the $(\log_2 |\lambda_1|, \log_2 |\lambda_2|)$ domain. In Figure 4.2 the boundaries of this pentagon are plotted as dashed lines in the case $n = 6$, $r = 1$ and $p = \frac{1}{3}$. In some limiting cases, however, the pentagon reduces to a square or a triangle. On the one hand, in the limit of r going to infinity the third constraint in (4.19) becomes inactive and the pentagonal region reduces to the square determined by the first two inequalities. On the other hand, in the limit
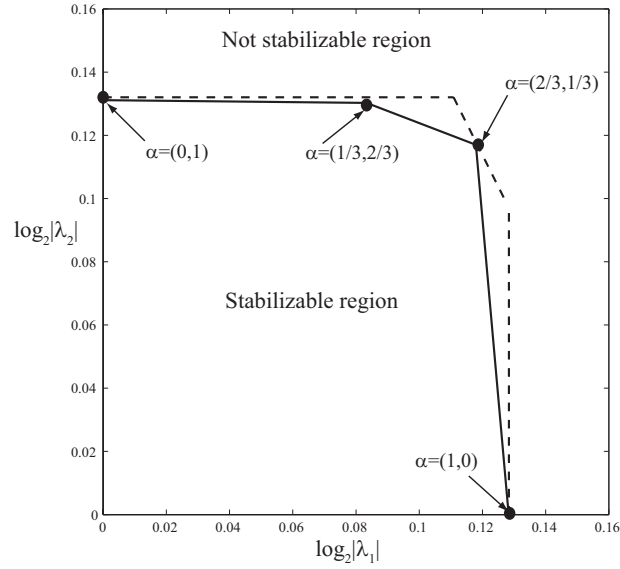
Figure 4.2: Stabilizability region for the system described in Example 4.5.2.

of p going to zero the only active constraint is the third inequality, and the region determined by (4.19) is triangular.

Proof. Consider the system in (4.17). Notice that each block $\mathbf{J}_i$ has an invariant real subspace of dimension $a_i s_i$, for *any* $s_i \in \{0, \ldots, m_i\}$. Consider the subspace $\mathcal{A}$ formed by taking the product of any of the invariant real subspaces for each real Jordan block. The total dimension of $\mathcal{A}$ is $\mu'(\mathbf{s}) = \sum_{i=1}^{u} a_i s_i$, for some $s_i \in \{0, \ldots, m_i\}$. Denote by $S = \{e_1, \ldots, e_n\} \subseteq \{1, 2, \ldots, d\}$ the index set of the components of $\mathbf{x}$ belonging to $\mathcal{A}$.

Suppose that a genie helps the decoder by stabilizing all the unstable states that are not in $\mathcal{A}$. Thus, stack the remaining unstable subsystem states to construct a new state $\mathbf{x}_k^S \in \mathbb{R}^{\mu'(\mathbf{s})}$.

$$\mathbf{x}_k^S = \left[\mathbf{x}_k^{(e_1)T}, \ldots, \mathbf{x}_k^{(e_n)T}\right]^T := \mathbf{R}\mathbf{x}_k \in \mathbb{R}^{\mu'(\mathbf{s})}.$$

where $\mathbf{R}$ is some transformation matrix. Observe that $\mathbf{x}_k^S$ evolves as follows

$$\mathbf{x}_{k+1}^S = \mathbf{J}^S \mathbf{x}_k^S + \mathbf{R}\mathbf{T}\mathbf{u}_k + \mathbf{R}\mathbf{v}_k, \text{ where } \det \mathbf{J}^S = \prod_{i \in S} \lambda_i^{a_i s_i}.$$

In order to prove the statement, we find a lower bound for $\mathbb{E}\left[\|\mathbf{x}_k^S\|^2\right]$, and show that (4.18) is a necessary condition for the lower bound to be finite. As in Theorem 4.4.1, the lower bound is given by $n_j^S = \frac{1}{2\pi e}\mathbb{E}_{\bar{S}_0^{k-1}}\left[e^{\frac{2}{\mu'(\mathbf{s})}h(\mathbf{x}_k^S|\bar{S}_0^{k-1}=\bar{s}_0^{k-1})}\right]$.

Proceeding as in Theorem 4.4.1, one can derive a recursive formula for n_j^S of the form

$$n_{j+1}^S \geq \mathbb{E}\left[\frac{|\det \mathbf{J}^S|^{\frac{2n}{\mu'(\mathbf{s})}}}{2^{\frac{2n}{\mu'(\mathbf{s})}R}}\right]n_j^S + \gamma^S.$$

for some constant $\gamma^S < \infty$. Therefore, $\left|\prod_{i\in S}|\lambda_i|^{a_i s_i}\right|^{\frac{2n}{\mu'(\mathbf{s})}}\mathbb{E}\left[2^{-\frac{2n}{\mu^S}R}\right] \geq 1$ implies that $\sup_{j\in\mathbb{N}}n_j^S = \infty$.

$\square$

The region determined in Theorem 4.5.1 has a special combinatorial structure. The polytope (4.18) is defined by the set function $f(S) := -\frac{|S|}{2n}\log_2\mathbb{E}[2^{-2n/|S|R}]$, $\forall S \subset E := \{1,\ldots,d\}$. It is shown in the following proposition, proved in the Appendix, that this set function defines a polymatroid.

Proposition 4.5.2. *The polytope defined by (4.18) is a polymatroid.*

Remarks:

1. When the rate process is constant, the constraints in (4.18) reduce to the well known condition [16],[21]

$$\sum_{|\eta_i|\geq 1}\log_2|\eta_i| \equiv \sum_{i\in\mathcal{U}}\mu_i\log_2|\lambda_i| < R,$$

and the stabilizability is contained in the region in the positive orthant strictly inside the hyperplane $\sum_{i\in\mathcal{U}}\mu_i\log_2|\lambda_i| = R$.

2. Notice that the right hand side of (4.18) can be rewritten as

$$-\frac{\mu'(\mathbf{s})}{2n}\log_2 2^{-\frac{2n}{\mu'(\mathbf{s})}r_{min}}\mathbb{E}\left[2^{-\frac{2n}{\mu'(\mathbf{s})}(R-r_{min})}\right]$$

$$= r_{min} - \frac{\mu'(\mathbf{s})}{2n}\log_2\mathbb{E}\left[2^{-\frac{2n}{\mu'(\mathbf{s})}(R-r_{min})}\right]$$

$$\rightarrow r_{min}$$

as $n \to \infty$. Thus, in the limit of n going to infinity, (4.18) reduce to

$$\sum_{i\in\mathcal{U}}\mu_i\log_2|\lambda_i| < r_{min}, \tag{4.20}$$

and the stabilizability region is determined uniquely by r_{min}. The intuitive justification of this latter fact is that the digital link supports the same rate for an arbitrarily long time interval, so stability has to be guaranteed under the worst possible rate. In the limit, stabilization is not possible for those channels where $r_{min} = 0$ (e.g. erasure channels).

3. In an erasure channel, for a fixed n, as r goes to infinity the stabilizability reduces to the n-dimensional cube described by

$$\log_2|\lambda_i| < \frac{1}{2n}\log_2\frac{1}{p} \quad \forall i \in \mathcal{U}. \tag{4.21}$$

In other words, the system in (4.17) cannot be stabilized if the erasure probability is such that

$$p \geq \frac{1}{\max_{i\in\mathcal{U}}\lambda_i^{2n}},$$

In the case $n = 1$, this is the same condition derived in [9] in the context of the LQG problem with erasures.

4.5.3 Sufficiency

We now present a sufficient condition for mean-square stabilizability of the multi-dimensional system (4.17). The scheme is based on the adaptive quantizer introduced in section 4.4.2. We introduce a rate allocation vector which indicates

what fraction of the available rate is allocated to each unstable sub-system.

Theorem 4.5.3. *Under assumptions **A0.-A3.** above, sufficient condition for stabilizability of the system in (4.17) in the mean square sense (4.3) is that $(\log_2 |\lambda_1|, \ldots, \log_2 |\lambda_u|) \in \mathbb{R}_+^u$ are inside the convex hull of the region determined by*

$$\log_2 |\lambda_i| < -\frac{1}{2n} \log_2 \mathbb{E}\left[2^{-\frac{2n}{\mu_i}\alpha_i(R)R}\right], \forall i \in \mathcal{U}, \tag{4.22}$$

for some rate allocation vector $\boldsymbol{\alpha}(R) := [\alpha_1(R), \ldots, \alpha_u(R)]^T$ satisfying

$$\begin{cases} \alpha_i(r) \in [0,1] \\ \frac{1}{\mu_i}\alpha_i(r)\, nr \in \mathbb{N} & , \quad \forall r \in \mathcal{R} \setminus \{0\}, i \in \mathcal{U}. \\ \sum_{i=1}^u \alpha_i(r) \le 1 \end{cases} \tag{4.23}$$

Suppose that transmission of r bits per channel use is supported on the digital link in a given block. The rate allocation vector $\boldsymbol{\alpha}(r)$ indicates what fraction of the total nr bits transmitted in a block is allocated to each sub-system. All μ_i modes in the i-th sub-system are quantized using $\frac{\alpha_i(r)nr}{\mu_i}$ bits. Condition (4.23) requires that $\frac{\alpha_i(r)nr}{\mu_i}$ is an integer number for all $i \in \mathcal{U}$, and that the total number of bits used in each block should not exceed nr. Such conditions define finitely many rate allocation vectors, and for each allocation vector (4.22) defines a cube in the space of $(\log_2 |\lambda_1|, \cdots, \log_2 |\lambda_u|)$. By using a time-sharing protocol among different rate allocation vectors it is possible to stabilize those points inside the convex hull of the union of such cubes. Before looking at the proof of the Theorem, consider the following Example:

Example 4.5.2. *Consider the system in Example 4.5.1 and assume that $n = 6$ and $r = 1$. Under this channel model, (4.23) defines four allocations vectors, namely $\alpha_1(1) = 1 - \alpha_2(1) = \frac{j}{6}, j \in \{0, 2, 4, 6\}$. For each allocation vector, (4.22) defines a cube in the space of $(\log_2 |\lambda_1|, \log_2 |\lambda_2|)$, and the stability region defined by Theorem 4.5.3 is the convex hull of the union of such cubes. Figure 4.2 shows the boundaries of the achievable stabilizability region in the case $p = \frac{1}{3}$: vertexes of the cube defined by (4.22) are represented as dots, while the solid lines show the convex hull of the*

union of such cubes. Notice that the outer bounds defined by (4.19) are achieved in three points, two of which lie on the two axis and correspond to the case where only one of the two sub-systems is unstable. In these cases the optimal rate allocation consists of allocating all the available bits to the unstable mode. The third optimal point corresponds to the case where the two eigenvalues have the same magnitude, i.e. $|\lambda_1| = |\lambda_2|$, and the optimal allocation in this case is to allocate one bit to each unstable mode. We will see that a protocol that time-shares among these three points is optimal in the limit $r \to \infty$.

Proof. The proof is divided into two parts. First it is shown that the linear dynamical system in (4.17) is stabilizable if (4.22) holds for some rate allocation vector $\boldsymbol{\alpha}(R)$ satisfying (4.23). Second it is shown that, by using a time-sharing protocol, all the points in the convex hull can be stabilized.

The coder computes a minimum variance estimator $\bar{x}^{(i,h)}$ for the h-th component of the i-th unstable mode, $h \in \{1, \ldots, \mu_i\}$ and $i \in \mathcal{U}$. Similarly, coder and decoder compute an estimator $\hat{x}^{(i,h)}$. Define $f_k^{(i,h)} = \bar{x}_k^{(i,h)} - \hat{x}_k^{(i,h)}$ as the error between these two estimators at time k. Let the stacked vector of unstable subsystems errors be $\mathbf{f}_k = \hat{\mathbf{x}}_k - \bar{\mathbf{x}}_k$.

Suppose that coder and decoder agree, ahead of time, on some rate allocation $\boldsymbol{\alpha}(R)$ satisfying (4.23). As in the case of a scalar system, divide times $k \in \mathbb{N}$ into cycles of integer duration $n\tau$, $\tau \in \mathbb{Z}_+$. Let

$$R_k^{(i)} := \frac{\alpha_i(R_k)R_k}{\mu_i} \in \mathbb{N}, \quad i \in \mathcal{U}, k \in \mathbb{N},$$

denote the number of bits allocated to the transmission of $f_k^{(i,h)}$ during the k-th channel block. By (4.23), $\sum_{i \in \mathcal{U}} \mu_i nR_k^{(i)} \leq nR_k \ \forall k \in \mathbb{N}$.

Therefore, at time $k = jn\tau$, the coder computes, for all $h \in \{1, \ldots, \mu_i\}$ and for all $i \in \mathcal{U}$,

$$\bar{\omega}_{nR_{j\tau}^{(i)}}(\omega_{jn\tau}^{(i,h)}) = q_{nR_{j\tau}^{(i)}}\left(f_{jn\tau}^{(i,h)}/l_j\right).$$

The scaling factor l_j is updated at the beginning of each cycle as follows,

$$l_{j+1} = \max_{\substack{h\in[1,\dots,\mu_i] \\ i\in\mathcal{U}}} \{\sigma, l_j|\lambda_i|^{n\tau}\kappa_{\nu_j^{(i)}}(\omega_{(j+1)n\tau-n}^{(i,h)})\},$$

where the random variable

$$\nu_j^{(i)} = \sum_{k=0}^{\tau-1} nR_{j\tau+k}^{(i)}$$

indicates the cumulative number of bits allocated to the i-th sub-system during the j-th cycle, and where $l_0 = \sigma$ and $\sigma^{2+\epsilon}$ is a uniform bound on the $(2+\epsilon)$-moment of $\mathbf{g}_j := \sum_{i=0}^{n\tau-1} \mathbf{J}^{n\tau-1-i}\mathbf{z}_{jn\tau+i}$, $j\in\mathbb{N}$. After the first block in the cycle, the decoder identifies an uncertainty interval $I_{nR_{j\tau}^{(i)}}(\omega_{jn\tau}^{(i,h)})$ for each unstable sub-system. The remaining $(n-1)\tau$ transmissions in the cycle are devoted to reducing the size of the uncertainty interval. After receiving the last $R_{(j+1)\tau-1}^{(i)}$ bits, the decoder can compute the final uncertainty interval $I_{\nu_j^{(i)}}(\omega_{(j+1)n\tau-n}^{(i,h)})$, corresponding to the uncertainty set formed by the quantizer $q_{\nu_j^{(i)}}(f_{jn\tau}^{(i,h)/l_j})$. For each unstable subsystem, the decoder sends to the plant a certainty-equivalent control signal $\mathbf{u}_k = P\hat{\mathbf{x}}_k$ where as in (4.13).

Let $\theta_j := M_\epsilon[\|\mathbf{f}_{jn\tau}\|, l_j]$. Proceeding along the same lines as in the scalar case, it can be shown that θ_j evolves according to the following recursive equation,

$$\theta_{j+1} \le 2\phi\sigma^2 + \phi\sum_{i\in\mathcal{U}}\mu_i|n\tau^{\mu_i-1}|^{2+\epsilon}\zeta\left(\mathbb{E}\left[\frac{\lambda_i^{2n}}{2^{2n\alpha_i(R)R/\mu_i}}\right]\right)^\tau\theta_j.$$

Hence, if (4.22) is satisfied, by choosing a τ sufficiently large, the coefficient of θ_j can be made strictly less than 1. Therefore, the recursion above is stable and yields uniformly bounded θ_j. The same argument used for the scalar case applies *sic et simpliciter* and it is now straightforward to show that the system is second moment stable.

It remains to show that, by time-sharing, all the points in the convex hull can be stabilized. Since the union of finite cubes in $\mathbb{R}^u$ is a connected compact set, by the Fenchel-Eggleston theorem [8, Theorem 18] each point in its convex closure can be represented as a convex combination of at most u points in the union, and thus each point is in the convex closure of the union of no more than u cubes

in (4.22). Given u rate allocation vectors $\boldsymbol{\alpha}^{(l)}(R)$, $l = 1, \cdots, u$, satisfying (4.23), and any $\gamma_l \in [0, 1]$ such that $\sum_{l=1}^{u} \gamma_l = 1$, it suffices to construct a scheme that stabilizes all modes $(\log_2 |\lambda_1|, \ldots, \log_2 |\lambda_u|)$ inside the region

$$\log_2 |\lambda_i| < -\sum_{l=1}^{u} \frac{\gamma_l}{2n} \log_2 \mathbb{E}\left[2^{-2n\alpha_i^{(l)}(R)R/\mu_i}\right], \forall i \in \mathcal{U}. \tag{4.24}$$

Divide times $k \in \mathbb{N}$ into cycles of duration τn, in such a way that $\gamma_l \tau n \in \mathbb{N}$ for $l = 1, \cdots, u$. During a fraction γ_l of the cycle allocate bits utilizing rate allocation vector $\boldsymbol{\alpha}^{(l)}(R)$. Repeating the analysis above, it can be proved that the crucial recursion for θ_j evolves as follows:

$$\theta_{j+1} \leq 2\phi\sigma^2 + \phi \sum_{i \in \mathcal{U}} \mu_i |n\tau^{\mu_i - 1}|^{2+\epsilon} \zeta \times$$
$$\left(\lambda_i^{2n} \prod_{l=1}^{u} \mathbb{E}\left[\frac{1}{2^{2n\alpha_i^{(l)}(R)R/\mu_i}}\right]^{\gamma_l}\right)^{\tau} \theta_j.$$

If (4.24) holds, we can choose τ large enough to make the recursion stable. Therefore, (4.24) are sufficient conditions for stabilizability. $\square$

Remarks:

1. If $\frac{rn}{d} \in \mathbb{N}$ for all $r \in \mathcal{R}$, then the rate allocation $\alpha_i(r) = \frac{\mu_i}{d} \ \forall r \in \mathcal{R} \setminus \{0\}$ is optimal when $\lambda := \lambda_1 = \ldots = \lambda_u$. In fact, from (4.22) sufficient condition for stabilizability is that

   $$\log_2 |\lambda| < -\frac{1}{2n} \log_2 \mathbb{E}\left[2^{-\frac{2n}{d}R}\right].$$

   On the other hand, this condition is also necessary, as we can see from (4.18) by letting $s_i = m_i$ for all $i \in \mathcal{U}$. For example, in Example 4.5.2 we have that $\frac{nr}{d} = 2$, so the rate allocation $\boldsymbol{\alpha} = (2/3, 1/3)$ is optimal (See Figure 4.2).

2. The scheme in Theorem 4.5.3 is optimal in the limit of n going to infinity, and the optimal coding scheme consists of a time-sharing protocol among the rate allocations $\boldsymbol{\alpha}^{(i)}(R) = \boldsymbol{e}_i$ for all $i \in \mathcal{U}$, where $\{\boldsymbol{e}_i\}_{i=1}^{d}$ are the canonical basis vectors of $\mathbb{R}^d$.

3. In an erasure channel, for a fixed n, as r goes to infinity the proposed achievable scheme is asymptotically optimal. The stabilizability region is given by the cube (4.21), and the optimal coding scheme consists of time-sharing among the rate distributions $\boldsymbol{\alpha}^{(i)}(r) = \boldsymbol{e}_i$ for all $i \in \mathcal{U}$ and the allocation given in Remark 1., i.e. $\alpha_i^{(u+1)}(r) = \frac{\mu_i}{d}$.

4. When the rate process is constant, Nair and Evans [16] showed that the necessary and sufficient conditions coincide. Once again, the optimal coding scheme consists of a time-sharing protocol among the rate distributions $\boldsymbol{\alpha}^{(i)}(R) = \boldsymbol{e}_i$ for all $i \in \mathcal{U}$.

5. A more general scheme is easily derived by allowing the rates allocated to each component of the same sub-system to be different. For ease of exposition, in Theorem 4.5.3 we assumed these rates to be equal.

4.5.4   Binary Erasure Channel

The stabilizing scheme proposed in the previous section provides an achievability result for stabilization over time-varying channels, and is optimal is some limiting cases. However, the scheme is not optimal in general. In this section, we improve the stabilizability region defined by Theorem 4.5.3 in the specific case of stabilization over a binary erasure channel. Before stating the result, we outline the main difference between the coding scheme used in this section and the construction in Theorem 4.5.3. In Theorem 4.5.3 time is divided into slots of fixed duration, and system state observations are quantized using a random number of bits dependent on the realization of the rate process. In this section, instead, we present a coder/decoder construction which is based on an alternative approach: state observations are quantized using a fixed number of bits per unstable mode; in turn, these are transmitted to the decoder over a random number of discrete time units which depends on the realization of the rate process. Based on this approach, it is possible to enlarge the set of feasible rate allocation vectors and, as a consequence, the stabilizability region. In this section, the following simplifying assumptions are made:

A4. The decoder has access to state feedback, i.e. in (4.2) we have that $C = I$ and $\mathbf{w}_k = 0$ for all $k \in \mathbb{N}$.

A5. $\exists W < \infty$ such that $|\mathbf{v}_k^{(i)}| \leq W$ uniformly in $k \in \mathbb{N}$, and $\mathbf{x}_0 \in \left[-\frac{1}{2}; \frac{1}{2}\right]^d$.

A6. The feedback digital link is a binary erasure channel, and the block length is $n = 1$.

We have the following proposition:

Proposition 4.5.4. *Under assumptions **A0.-A6.** above, sufficient condition for stabilizability of the system in (4.2) in the mean square sense (4.3) is that $(\log_2 |\lambda_1|, \ldots, \log_2 |\lambda_u|)\mathbb{R}_+^u$ are inside the convex region determined by*

$$\log_2 |\lambda_i| < -\frac{1}{2} \log_2 \mathbb{E}\left[2^{2\frac{\alpha_i}{\mu_i}R}\right], i \in \mathcal{U}, \tag{4.25}$$

for some rate allocation vector $\boldsymbol{\alpha} := [\alpha_1, \ldots, \alpha_u]$ such that,

$$\begin{cases} \alpha_i \in [0, 1] \cap \mathbb{Q}, \ \forall i \in \mathcal{U}, \\ \sum_{i=1}^u \alpha_i \leq 1. \end{cases} \tag{4.26}$$

Comparing (4.23) and (4.26), notice that while in Theorem 4.5.3 only a *finite* number of rate allocation vectors satisfy (4.23), the region defined by Proposition 4.5.4 is given by the union of a *countable* number of u-dimensional cubes, each of which is defined by (4.25) for some rate allocation vector satisfying (4.26). We also notice that the stabilizability region defined by Proposition 4.5.4 is convex, so a time-sharing protocol among different rate allocation policies is not required.

Example 4.5.3. *Consider a system with two distinct modes of dimensionality one, having unstable real eigenvalues λ_1 and λ_2, respectively. Figure 4.3 shows the achievable stabilizability region under this channel model, assuming $p = 2/3$. The boundaries of the region defined by Proposition 4.5.4 are represented as a solid curve, and each point on this curve is obtained by (4.25) for some choice of the rate allocation vector. The region in Theorem 4.5.3 is delimited by a dotted line, which*
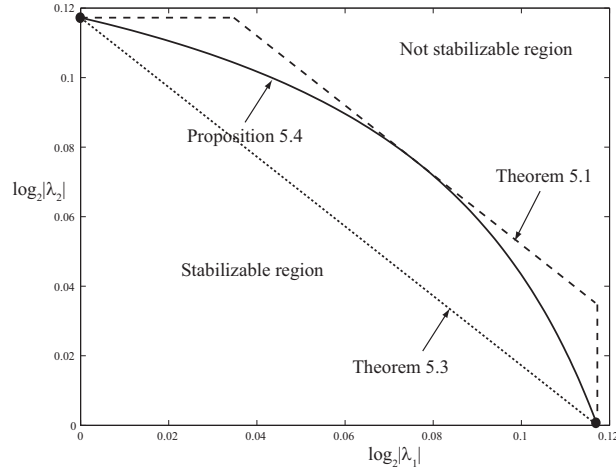
Figure 4.3: Stabilizability region for the system described in Example 4.5.3.

represents the convex combination of two points (bold dots), obtained by (4.23) with $\boldsymbol{\alpha}(1) = [1, 0]$ and $\boldsymbol{\alpha}(1) = [0, 1]$. Finally, the necessary conditions derived in Theorem 4.5.1 define a pentagon that is delimited by a dashed line. The region in Proposition 4.5.4 is optimal at the intersections with the two axis and at one point on the bisectrix $|\lambda_1| = |\lambda_2|$.

Proof. Fix an $\boldsymbol{\alpha} \in [0, 1]^u$ satisfying (4.26) and $m \in \mathbb{Z}_+$ such that $\alpha_i m / \mu_i \in \mathbb{N}$ for all i. A renewal process $\{t_k\}_{k=1}^\infty$ determines the times at which the encoder quantizes the state observations. The random interarrival times of this renewal process are denoted by the sequence $\{\tau_k\}_{k=1}^\infty$, such that $t_k = \tau_1 + \cdots + \tau_k$ for all $k \in \mathbb{N}$.

The stability in the mean square sense of the system in (4.2) is proved by showing that for each unstable sub-system $x^{(i,h)}$, $h \in [0, \ldots, \mu_i]$ and $i \in \mathcal{U}$, there exists a mean square stable sequence $\{z_k^{(i,h)}\}_{k=0}^\infty$ such that $|x_{t_k}^{(i,h)}| \leq z_k^{(i,h)}$, for all $k \in \mathbb{N}$. We define $\{z_k^{(i,h)}\}_{k=0}^\infty$ recursively as follows:

$$\begin{cases} z_1^{(i,h)} & = 1 + W \\ z_{k+1}^{(i,h)} & = \frac{|\lambda_i|^{\tau_k}}{2^{\alpha_i m / \mu_i}} z_k^{(i,h)} + \eta, \end{cases} \qquad (4.27)$$

where $\eta = \frac{W}{1 - |\lambda_i|}$ if $|\lambda_i| > 1$ and $\eta = \tau_k W$ if $|\lambda_i| = 1$. At the random time t_k, the encoder partitions the interval $[-z_k^{(i,h)}; +z_k^{(i,h)}]$ into $2^{\alpha_i m / \mu_i}$ uniform intervals, and

computes $\hat{x}_{t_k}^{(i,h)}$ as the center of the interval containing $x_{t_k}^{(i,h)}$. By construction, the approximation error satisfies

$$|x_{t_k}^{(i,h)} - \hat{x}_{t_k}^{(i,h)}| \leq \frac{z_k^{(i,h)}}{2^{\alpha_i m/\mu_i}}. \tag{4.28}$$

The time required for transmission of the cumulative m bits describing the quantized source symbols from coder to decoder is denoted by the interarrival time τ_k. We define $\tau := \inf\left\{k : \sum_{l=1}^k R_l = m\right\}$ as the time of the m-th 'success' in the Bernoulli process $\{R_i\}_{i=1}^\infty$; for any $p > 0$, we have that $\Pr(\tau < \infty) = 1$, and τ has negative binomial distribution with parameters m and p. The interarrival times $\{\tau_k\}_{k=1}^\infty$ are independent non-negative random variables, identically distributed as τ.

At time $t_{k+1} = t_k + \tau_k$, upon reception of the m binary source symbols the decoder computes the control signal

$$u_{t_{k+1}}^{(i,h)} = -\lambda_i^{\tau_k} \hat{x}_{t_k}^{(i,h)}. \tag{4.29}$$

Making use of (4.27), (4.28) and (4.29), we have the following chain of inequalities,

$$
\begin{aligned}
|x_{t_{k+1}}^{(i,h)}| &\leq |\lambda_i^{\tau_k} x_{t_k}^{(i,h)} + u_{t_{k+1}}^{(i,h)}| + \sum_{j=0}^{\tau_k - 1} |\lambda_i|^{\tau_k - 1 - j} |v_{t_k + j}|, \\
&\leq |\lambda_i|^{\tau_k} |x_{t_k}^{(i,h)} - \hat{x}_{t_k}^{(i,h)}| + \eta, \\
&\leq z_{k+1}^{(i,h)}.
\end{aligned}
\tag{4.30}
$$

From (4.30) and proceeding by induction, it follows that $|x_{t_k}^{(i,h)}| \leq z_k^{(i,h)}$, for all $k \in \mathbb{N}$. Next, we show that (4.25) is a sufficient condition for the sequence $\{z_k^{(i,h)}\}_{k=0}^\infty$ to be mean square stable, i.e. $\sup_{k \in \mathbb{N}} \mathbb{E}[|z_k^{(i,h)}|^2] < \infty$, for all $h \in [0, \ldots, \mu_i]$ and $i \in \mathcal{U}$. From (4.27) and the triangle inequality, it follows that

$$\left(\mathbb{E}\left[|z_{k+1}^{(i,h)}|^2\right]\right)^{\frac{1}{2}} \leq \left(\mathbb{E}\left[\frac{\lambda_i^{2\tau_k}}{2^{2\alpha_i m/\mu_i}}\right]\right)^{\frac{1}{2}} \left(\mathbb{E}|z_k^{(i,h)}|^2\right)^{\frac{1}{2}} +$$
$$+ \left(\mathbb{E}\left[|\eta|^2\right]\right)^{\frac{1}{2}}, \ \forall i \in \mathcal{U}, \tag{4.31}$$

wherein $\mathbb{E}\left[|\eta|^2\right] < \infty$ as $\mathbb{E}\left[\tau^2\right] = \frac{n(1-p)}{p^2} < \infty$ for all $p \in (0,1)$. By writing explicitly the expectations in (4.25), we obtain that

$$\lambda_i^2 \left[\frac{p}{2^{2\alpha_i/\mu_i}} + (1-p) \right] < 1, \ \forall i \in \mathcal{U}. \tag{4.32}$$

Making use of (4.32) simple algebra shows that

$$\mathbb{E}\left[\frac{\lambda_i^{2\tau_k}}{2^{2\alpha_i m/\mu_i}} \right] = \left(\frac{1}{2^{2\alpha_i/\mu_i}} \frac{p|\lambda_i|^2}{1-(1-p)e^{it}} \right)^m < 1, \ \forall i \in \mathcal{U}. \tag{4.33}$$

From (4.33) it follows that the recursive formula in (4.31) is stable. Therefore, (4.25) is a sufficient condition to ensure $\sup_k \mathbb{E}[|z_k^{(i,h)}|^2] < \infty$.

Finally, the convexity of the region described by (4.25) follows from Jensen's inequality applied to the concave function $\phi(x) = -\frac{1}{2}\log_2\left((1-p) + p\,2^{-2x}\right)$. $\quad\square$

4.6 Conclusion

Motivated by control problems over time-varying channels, we considered mean square stabilizability of a discrete-time, linear system with a noise free time-varying digital communication link. Process and observation disturbances were allowed to occur over an unbounded support. Necessary conditions were derived employing information-theoretic techniques, while a stabilization scheme based on an adaptive successively refinable quantizer was constructed. In the scalar case, this scheme was shown to be optimal. Furthermore, we have shown that in the vector case the necessary condition for stabilization has an interesting polymatroid structure, and have proposed a stabilization scheme that is optimal in some limiting regimes. An additional contribution is that we bridged the information-theoretic results of stabilization over rate limited channels, with the corresponding network-theoretic ones on critical dropout probabilities in systems with unbounded disturbances. We have done so by recovering the latter results as a special case of our analysis.

Chapter 4, in part, is a reprint of the material as it appears in P. Minero, M. Franceschetti, S. Dey and G. N. Nair, "Data Rate Theorem for Stabilization

Over Time-Varying Feedback Channels," *IEEE Trans. on Automatic Control*, vol 54, no. 2, pp. 243-255, February 2009. The dissertation author was the primary investigator and author of this paper.

4.7 Appendix

4.7.1 Proof of Lemma 4.4.2

Proof. First, observe that the following chain of inequalities holds:

$$
\begin{aligned}
\mathbb{E}_{\bar{S}_j|\bar{S}_{j-1},R_j} & h(x_{jn}|\bar{S}_{j-1} = \bar{s}_{j-1}, \bar{S}_j = \bar{s}_j, R_j) \\
&= h(x_{jn}, \bar{S}_j|\bar{S}_{j-1} = \bar{s}_{j-1}, R_j) - H(\bar{S}_j|\bar{S}_{j-1} = \bar{s}_{j-1}, R_j) \\
&\geq h(x_{jn}|\bar{S}_{j-1} = \bar{s}_{j-1}, R_j) - H(\bar{S}_j|\bar{S}_{j-1} = \bar{s}_{j-1}, R_j) \\
&\geq h(x_{jn}|\bar{S}_{j-1} = \bar{s}_{j-1}, R_j) - \ln 2^{nR_j} \\
&= h(x_{jn}|\bar{S}_{j-1} = \bar{s}_{j-1}) - \ln 2^{nR_j},
\end{aligned}
\tag{4.34}
$$

where $h(x, A|B)$ with A discrete denotes $-\mathbb{E}[\ln(p_{A|B}f_{x|A,B})]$. The last inequality follows from the fact that, given R_j, the cardinality of $\{S_{jn}, \ldots, S_{(j+1)n-1}\}$ is 2^{nR_j}, and where the last equality follows from the fact that $x_{jn} \to \bar{S}_{j-1} \to R_j$ is a Markov chain. Then,

$$
\begin{aligned}
\mathbb{E}_{\bar{S}_j|\bar{S}_{j-1},R_j} & e^{2h(x_{jn}|\bar{S}_j=\bar{s}_j)} \\
&\geq \mathbb{E}_{\bar{S}_j|\bar{S}_{j-1},R_j} e^{2h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1},\bar{S}_j=\bar{s}_j,R_j)} \\
&\geq e^{2\mathbb{E}_{\bar{S}_j|\bar{S}_{j-1},R_j} h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1},\bar{S}_j=\bar{s}_j,R_j)} \\
&\geq e^{2[h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1})-\ln 2^{nR_j}]} \\
&= \frac{1}{2^{2nR_j}} e^{2h(x_{jn}|\bar{S}_{j-1}=\bar{s}_{j-1})},
\end{aligned}
$$

where the first inequality follows from the fact that conditioning reduces the entropy; the second inequality follows from Jensen's inequality; finally, (4.34) implies the third inequality. □

4.7.2 Proof of Proposition 4.5.2

Proof. Let $E = \{1, \ldots, d\}$ and $f(S) = -\log_2 \left(\mathbb{E} \left[(2^{-R})^{2n/|S|} \right] \right)^{|S|/2n} 1_{|S|>0} = -\log \| 2^{-R} \|_{\frac{2n}{|S|}} 1_{|S|>0}$.

Following the definition in [6], in order for the polytope

$$\mathcal{B}(f) := \left\{ (x_1, \ldots, x_u) : \sum_{i \in S} x_i \leq f(S) \; \forall S \subset E, x_i \geq 0 \; \forall i \right\}$$

to be a polymatroid, we need to show the following properties:

1. $f(\emptyset) = 0$: this is immediate from the definition of $f(\cdot)$.

2. $f(S) \leq f(T)$ if $S \subset T$: this follows from the fact that $\| \; X \; \|_{\frac{1}{m}} \leq \| \; X \; \|_{\frac{1}{n}}$ if $n \leq m$.

3. $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$: this can be proved as follows. Note that $f(S)$ is a function only of $|S|$, i.e. $f(S) := g(|S|)$. W.l.o.g., assume that $j := |S| \leq |T| =: k$. Let $i := |S| - |S \cap T|$ and note that this is never negative. Further note that $|S \cup T| = |S| + |T| - |S \cap T| = k + i$. The desired property is then that $g(j) - g(j-i) \geq g(k+i) - g(k)$ for all integers $i \leq j \leq k$. Now, from the fundamental theorem of calculus $\exists a \in [j-i, j] \cap \mathbb{R}$ and $\exists b \in [k, k+i] \cap \mathbb{R}$ such that $g(j) - g(j-i) = g'(a)i$ and $g(k+i) - g(k) = g'(b)i$. Thus proving the desired inequality is equivalent to proving that $g'(a) \geq g'(b)$ for all $0 < a \leq b$. On the other hand, this inequality follows from the concavity of the function $g(x)$ for $x > 0$.

□

4.8 Bibliography

[1] J. Baillieul, "Feedback designs in information-based control," *Proceedings of the Workshop on Stochastic Theory and Control, Lawrence, Kansas, B. Pasik-Duncan*, ed. Springer, pp. 35–57, Oct. 2001.

[2] J. Baillieul and P. Antsaklis, "Control and communication challanges in networked real time systems," *Proc. IEEE, Special issue on emerging technologies of networked control systems*, vol. 95, no. 1, pp. 9–28, Jan. 2007.

[3] R.W. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Trans. Autom. Control*, vol. 45, no.7, pp. 1279–1289, Jul. 2000.

[4] T. Cover and J. Thomas, *Elements of Information Theory*, New York: Wiley, 1987.

[5] D.F. Delchamps, "Stabilizing a linear system with quantized state feedback," *IEEE Trans. Autom. Control*, vol. 35, no. 8, pp. 916–924, Aug. 1990.

[6] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Proceedings of Calgary Int. Conf. Combinatorial Structures and Applications, Calgary, Alta, June 1969*, pp. 69–87.

[7] N. Elia, "Remote stabilization over fading channels," *Systems and Control Letters*, vol. 54, no. 3, pp. 237–249, Mar. 2005.

[8] H. G. Eggleston, *Convexity*, Cambridge University Press, Cambridge, England, 1963.

[9] V. Gupta, B. Hassibi and R.M. Murray, "Optimal LQG control across packet-dropping links," *Systems and Control Letters*, vol. 56, no. 6, pp. 439–446, Jun. 2007

[10] R.A. Horn and C.R. Johnson, "Matrix Analysis", Cambridge University Press, 1985.

[11] J.P. Hespanha, P. Naghshtabrizi and Y. Xu, "A survey of recent results in networked control systems," *Proc. IEEE, Special issue on emerging technologies of networked control systems*, vol. 95, no. 1, pp. 138–162, Jan. 2007.

[12] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M.I. Jordan and S.S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sept. 2004.

[13] N.C. Martins, M.A. Dahleh and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *IEEE Trans. Autom. Control*, vol. 51, no.3, pp. 438–447, Mar. 2006.

[14] A.S. Matveev and A.V. Savkin, "Comments on 'Control over noisy channels' and relevant negative results," *IEEE Trans. Autom. Control*, vol. 50, no.12, pp. 2105–2110, Dec. 2005.

[15] A.S. Matveev and A.V. Savkin, "An analogous of Shannon information theory for networked control systems," in *Proc. IEEE Conference on Decision and Control*, 2004, pp. 4491–4496.

[16] G.N. Nair and R.J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43,no. 2, pp. 413–436, Jul 2004.

[17] G.N. Nair, F. Fagnani, S. Zampieri and R.J. Evans, "Feedback control under data rate constraints: an overview," *Proc. IEEE, Special issue on emerging technologies of networked control systems*, vol. 95, no. 1, pp. 108–137, Jan. 2007.

[18] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link Part I: Scalar Systems," *IEEE Trans. Inf. Theory*, vol. 52, no.8, pp. 3369–3395, Aug. 2006.

[19] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla and S.S. Sastry, "Foundations of control and estimation over lossy networks." *Proc. IEEE, Special issue on emerging technologies of networked control systems*, vol. 95, no.1, pp.163–187, Jan. 2007.

[20] C. Shannon, "A mathematical theory of communication", *Bell Systems Technical Journal* 27, 379–423, 623–656. Reprinted as: The mathematical theory of communication. *University of Illinois Press, Champaign.*

[21] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, Jul. 2004.

[22] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp.1196–1201, Jul. 2004.

[23] S. Yuksel and T. Basar, "Control over Noisy Forward and Feedback Channels," submitted to *IEEE Trans. Autom. Control*, 2007.

[24] W.S. Wong and R.W. Brockett, "Systems with finite communication bandwidth constraints, II: Stabilization with limited information feedback," *IEEE Trans. Autom. Control*, vol. 44, no. 5, May 1999.