A universal, operational theory of multi-user communication with fidelity criteria

by

Mukul Agarwal

B. Tech., Electrical Engineering,
Indian Institute of Technology, Bombay

M. Tech, Electrical Engineering, with a specialization in Communications and Signal Processing,
Indian Institute of Technology, Bombay

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

February 2012

Author: ______________________________

Department of Electrical Engineering and Computer Science
February 3, 2012

Certified by: ______________________________

Professor Sanjoy K. Mitter
Thesis Supervisor

Accepted by: ______________________________

Professor Leslie A. Kolodziejski
Chair of the Committee on Graduate Students

Thesis committee:

Prof. Robert Gallager, Electrical Engineering and Computer Science, MIT

Prof. John Tsitsiklis, Electrical Engineering and Computer Science, MIT

Tom Richardson, VP Engineering and Chief Scientist, Qualcomm Flarion Technologies

Prof. Sanjoy Mitter, Electrical Engineering and Computer Science, MIT

A universal, operational theory of multi-user communication with fidelity criteria

by

Mukul Agarwal

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

Abstract

This thesis has two flavors:

1. *A theory of universal multi-user communication with fidelity criteria:* We prove the optimality of digital communication for universal multi-user communication with fidelity criteria, both in the point-to-point setting and in the multi-user setting. In other words, we prove a universal source-channel separation theorem for communication with a distortion criterion, both in the point-to-point setting and the multi-user setting. In the multi-user setting, the setting is unicast, that is, the sources which various users want to communicate to each other are independent of each other. The universality is over the medium of communication: we assume that the medium might belong to a family. Both in the point-to-point setting, we assume that codes can be random: the encoder might come from a family of deterministic codes and the decoder has access to the particular realization of the deterministic code, and finally, an average is taken over all these deterministic codes. In Shannon's theory, random-coding is a proof technique. However, in our setting, random codes are essential: universal source-channel separation does not hold if codes are not allowed to be random. This happens because we are asking the universal question. We also show the partial applicability of our results to the traditional wireless telephony problem.

2. *An operational theory of communication with a fidelity criterion:* We prove the source-channel separation theorem operationally: we rely only on definitions of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a certain distortion level. We do not rely on functional simplifications, for example, mutual information expressions for the proofs. By operational, we do *not* mean that what we are doing is "practically operational". The view that we have can also be viewed as a layered black-box view: if there is a black-box that is capable of one form of communication, then the black-box can be layered in order to accomplish another form of communication.

Thesis Supervisor: Sanjoy Mitter
Title: Professor of Electrical Engineering and Computer Science

To my thesis advisor,

Prof. Sanjoy K. Mitter,

for being a rock solid pillar of support and
stability these seven and a half years,
for his continued interest in my education,
and for possibly being his last full student.

Acknowledgements

This list is supposed to be endless.

I thank my advisor Prof. Sanjoy Mitter who has been a rock solid pillar of support and stability through these seven and a half years. For starting me off on the problem that started off my thesis. And for taking keen interest in my education, in particular, the classes that I took. For numerous words of wisdom, of which I recall one, "You don't want to be a scholar; you want to be a researcher!"

I thank Prof. Anant Sahai of UC Berkeley for allowing me to spend a semester in UC Berkeley, and for working on a research problem, which was an unsolved problem in a paper of his, with me. This is what started my thesis.

Swastik Kopparty, a student here at MIT, has been a collaborator on one research problem and interactions with him have been wonderful. It has been good to have computer science perspective, and see some sheer mathematical brilliance! I take a problem formulated as a combinatorics problem to him which I have no clue what to do about, and Swastik solves it in one day!

I thank Prof. Bob Gallager, my committee member for the many insightful discussions that we have had on the topics of communications and information theory. As is said, it is best to hear from the "horses mouth." On the topics of communications and information theory, I do not think there exists a better authority, at least in academia. Also, for reading parts of my thesis very carefully which resulted in significant improvement in the main chapter.

I thank Prof. Tsitsiklis, another committee member, for reading my thesis and for some beneficial comments.

I have also learnt a lot from Tom Richardson and Rajiv Laroia at Qualcomm Flarion Technologies, where I interned. It was good to apply some of the theoretical skills I had to more practical problems. It was after I returned from Qualcomm Flarion Technologies, that I started to think about multi-user problems, and to my amazement, what I had done for the point-to-point case generalized to the multi-user setting in a very straight forward manner. Had it not been for the internship, I wonder if I would even have tried to think about multi-user problems. Tom Richardson has also been a committee member, and some of the discussions with both him and Rajiv Laroia have been wonderful. They have also been also very positive in their encouragement and their perspective has mattered a lot. I do not remember exactly what Tom and Rajiv told me, but it was something like "you want to look into the future and thus, select problems which you work on," and this has stuck with me, since then.

Discussions with Prof. Lizhong Zheng and Prof. Dave Forney have also been helpful.

Among students, two students from whom I got a lot of ideas were Shasibhushan Borade and Ashish Khisti. I would go to them with random thoughts in my mind. And suddenly, they will say something, and there would be a leap in my thesis! I should especially thank Shashibhushan Borade because the first problem that I solved for my thesis, happened with discussions with him. This is the same problem with came out of a paper of Anant Sahai's which I mentioned before. Discussions with Barış Nakiboğlu were also very helpful. Vincent Tan gets a profuse thank you, for going through the proof of the most technical part of my thesis with me, and also, for reading parts of my thesis. It is due to Vincent that I feel confident that the chances of any mistakes in the most technical part of my thesis are very low.

In general, MIT has been a wonderful place to get educated. I have had the chance to take some good classes in various departments at MIT: EECS, mathematics, physics and theater. There is no better way to learn communications than from Bob Gallager and Dave Forney! Their basic courses are probably the best existing, anywhere in the world. It is the best possible combination: Bob Gallager, the mathematical engineer, and Dave Forney, an engineer who is a very good mathematician! Learning probability from Dan Stroock in the mathematics department gave me some taste of cold-blooded difficult mathematics! The theater department at MIT has been my introduction to the arts here. Prof. Alan Brody has been a great introductory theater professor, and he has taken a keen interest in my education at MIT. I will miss our beer and wine outings into Cambridge and discussions on art, theater, philosophy and meditation! I have also learnt a lot from Prof. Jay Scheib, another theater professor at MIT, the epitome of avant-garde, for his teachings on theater, especially the view-points teaching of Anne Bogart, which are applicable much more generally, and for his statements like "you want to go in leaps, not in steps" and "you should just try to do what you want, and the support will come!". Being at MIT and being able to learn the arts from some of the very best has been a pleasure and honor.

MIT journey would not have been possible without many friends.

Two deserve special mention: Patrick Kreidl and Chintan Vaishnav. Patrick Kreidl, for his class, style, and wisdom. Our discussions on research, academia, industry, our outings to Emma's for pizza and wine, and his perspectives on life. For his simple words of encouragement all the time. And for listening to me during my most difficult times. Without Patrick, the MIT journey would have been much more difficult. Similarly, Chintan, for our long walks through the infinite corridor and discussions on research, what is good research, on life and philosophy, in particular, for our readings of Plato together. For his ever cheerful-ness, which always made me jealous: why cannot I be like that! Also, for his words of wisdom, and our humor together, which has been classic!

Barış Nakiboğlu has been another very good friend, here at MIT. Our lunch and dinner outings, and sharing of research and personal lives has been wonderful. We have gone through similar processes, I believe, and sharing them has been so helpful. Vincent Tan has also been a good friend, lunch colleague and someone, to whom I can always go, whenever I have a research question. In the earlier years, Shashibhushan Borade and Ashish Khisti have been good friends.

The fun conversations with Emmanuel Ebbe, a fellow student and friend, which I should definitely spare the audience, are very well remembered!

The list of friends cannot be complete without my three mother figures at MIT: Doris Inslee, Maria Brennan and Leslie Patton.

Doris Inslee, who was the LIDS administrator for many years, and someone, who in spite of her work load, was always happy when you walked into her office. Her ever cheerfulness, her wisdom, our lunch time outings, have been a pleasure. Of course, she has been the most helpful in all matters administrative, with all the professionalism and class.

Maria Brennan, who is the Assistant Director of the International Students Office. Someone who I always could rely on for the rather painful international students status. And someone, who I could always go to in times of agony: she would always lend a hearing ear, and her perspective on life have been wonderful. I have learnt so much from her. Our discussions, for example, the one on "Gandhi and women", I still remember! Without her, again, life at MIT could have been more difficult.

Finally, Leslie Patton, who is an administrator in the insurance department at MIT Medical. Again, for all her help with matters related to medical insurance. And for always talking to me. Especially, inviting me to her home for Christmas when I had no where to go. For her constant enthusiasm and positivity.

Jennifer Donovan, a LIDS administrator has also been a good friend. Again, helpful in all matters administrative, and a good ear, always to listen to you. And Brian Jones, for his smile, and taking me to learn kung-fu, which, fortunately or unfortunately, I learnt, only for a few days.

For the many friends, here at MIT, who I have missed, I apologize.

Coming from a different country is difficult personally. New country, new people, different people, no friends.

My first friends in Boston were Manas and Ananya, who I met on the whale watch trip which was an MIT orientation event. They came and talked to me when apparently, I was sitting in a corner on the boat, missing home. Since then, we have been friends, and we have spent many a happy times, together. A lot of gratitude to them for inviting me to their house numerous times.

I have been blessed to have a place outside MIT where I could go. That is the place of Zach and Kippy who I know through meditation. Their place, on tuesday nights, was one source of stability for many years. Zach has been a dear friend, there always to listen to me.

Zach's brother Dan, to the small extent that I have known him, has been someone with whom I have thoroughly enjoyed our intellectual discussions and his words of wisdom, especially on Buddhist thought. Also, his wife Mari, who has been a very caring person when I visited them in Berkeley. It is ironical that an Indian gets a PhD in western science and Americans and Finnish get PhDs in Buddhist thought! Those, well, are some of the ways of this world! Dan is also thanked for help with some of the hindi and the pali in this acknowledgement and the quotation.

Zach's and Dan's mother Kippy has also been another mother figure, here in Boston. She has always been very encouraging and supportive regarding matters, both professional and personal.

Another home has been Dhamma Dhara, the Vipassana Meditation Center in Western Massachusetts, where I would go on numerous weekends to meditate or to volunteer, especially during my first few years, which were very lonely.

I have had two very good Italian friends, Bruno and Alberto, who I met in US, but they are back to Italy now. The following Italian, due to my friend Tina Magazzini, is for them: Ho conosciuto Bruno quando si trovava in visita per un semestre al MIT. Siamo diventati buoni amici, ed è stato la mia guida all'italianità; in seguito, visitandolo in Italia, e conoscendo sua moglie e suo figlio, la nostra amicizia si è rafforzata. Ho conosciuto Alberto durante uno stage a Qualcomm Flarion Technologies. Senza di lui, la vita in quel posto dimenticato da dio e dagli uomini che è il New Jersey sarebbe stata impossibile. Le nostre camminate, le partite a calcio e a ping-pong, le nostre visite alla panettaria italiana, sono ricordi a cui ripenso con gratitudine, così come la sua ospitalità quando mi ha accolto nella sua casa in Italia.

As I mention New Jersey (in the above Italian: that is where I interned in Qualcomm Flarion Technologies), I remember Carol Wishart. The first day that I landed in New Jersey for my internship, I had the incorrect directions. I was roaming around clueless, and Carol came and talked to me. She said that where I am supposed to live was 2 miles from where I was! She gave me a ride, and invited me, several times to her place for food! It was good to know Carol and her mother. Without Carol also, New Jersey would have been nightmare!

I will go down memory lane, to my years in India. I will write this part in my mother-tongue, hindi.

सम्भवतः, मैं विज्ञानाध्ययन के लिये पूर्वनियत था! मुझे याद है कि जब मैं दसवीं और ग्यारहवीं कक्षा में था, मैं पिताजी की फ़ाइनमैन और आइन्सटाइन की किताबें झाँका करता था। खास समझ नहीं पाता था लेकिन कोशिश जरूर करता था! ये किताबें पढ़ने से मुझे भौतिक विज्ञान में रुचि हुई। गणित प्रतियोगिताओं की तैयारी करने हेतु मुझे गणित में रुचि हुई। इन प्रतियोगिताओं की तैयारी के शुरुआती दिनों में अध्यापक टी. एन. सेशन बहुत उत्साह से गणित सिखाते थे। वे हमें सिखाने में सम्पूर्ण समर्पित थे। इन शुरुआत के दिनों में

एक सहछात्रा और अच्छी दोस्त, रुपषा सामन्ता, याद आती है, जिसके साथ मैं गणित सवालों के बारे में सोचा करता था। जब इन प्रतियोगिताओं की तैयारी गंभीर हुई, तब मुझे प्रणेशचर, योगानंद और वेंकटचल आदि बहुत समर्पित अध्यापकों ने गणित सिखाया। अभिनव, हरिहरन्, ऋषि राज, शोहम, जयदीप, आदि अत्यंत बुद्धिमान सहछात्रों से बहुत गणित सीखा।

लेकिन सबसे महत्त्वपूर्ण वह चार साल थे जब उच्च गणित सीखने मैं गर्मियों की छुट्टियों में भारतीय सांख्यिकीय संसथान, कल्कत्ता जाया करता था। इन दिनों, भौतिक यंत्रशास्त्र का पूर्वस्नातक अध्ययन मैं भारतीय प्रौद्योगिकी संस्थान, बम्बै, में कर रहा था, लेकिन इन चार साल हर गरमी की छुट्टियों का एक महीना मैं कलकत्ते में बिताता था। भौतिक यंत्रशास्त्र और गणित का यह बहुत उत्तम मेल था। बहुत उच्च गणित सीखा! सबसे महत्त्वपूर्ण, मैंने प्रायिकता सिद्धांत सीखा जिसकी छाप मेरे इस शोध निबंध में स्पष्ट है! तीन प्राध्यापक, शोमेश बागची, बी.वी. राव तथा अमर्त्य दत्त, विशेष याद आते हैं। इन्होंने खूब समर्पण से सिखाया। प्राध्यापक राव आज तक के मेरे प्रायिकता सिद्धांत के सर्वश्रेष्ठ अध्यापक हैं!

जैसा मैंने पहले कहा, पूर्वस्नातक अध्ययन मैंने भारतीय प्रौद्योगिकी संस्थान, बम्बै, में दिद्युत यंत्रशास्त्र में किया। प्रथम वर्ष मेरा कमरा साथी उदय चेट्टियर था। उदय से अधिक बुद्धिमान कमरा साथी मिलना मुशकिल ही नहीं, नामुंकिन है! बहुत से प्राध्यापकों ने सिखाया, जिनमें प्राध्यापक नारायणन् खास याद आते हैं। प्राध्यापक नारायणन् ने पूर्वस्नातक अध्ययन के प्रथम वर्ष में अत्याधिक गणित दृष्टिकोण से विद्युत प्ररिपथों के बारे में सिखाया। तब मैंने विद्युत यंत्रशास्त्र और गणित में पहली बार साफ़ संबंध पाया। यह संबंध मेरे इस शोध निबंध में भी देखने को मिलता है। प्राध्यापक नारायणन् ने तभी से अनुसंधान की ओर भी प्रोत्साहित किया। व्यक्तिगत स्तर पर मेरे बहुत अच्छे दोस्त ध्रुव और संदीप याद आते हैं।

मैंने स्नातकोत्तर अध्ययन भी भारतीय प्रौद्योगिकी संस्थान, बम्बै, में ही किया। स्नातकोत्तर शोध प्रबंध के लिये जिस अनुसंधान सवाल के बारे में मैं सोच रहा था, उसपर मैं प्राध्यापक विवेक बोरकर के साथ वादानुवाद किया करता था। उन दिनों प्राध्यापक बोरकर टाटा मूलभूत अनुसंधान संस्थान में प्राध्यापक थे, लेकिन भारतीय प्रौद्योगिकी संस्थान, बम्बै, में सहित प्राध्यापक थे। यह मेरा प्रथम गम्भीर अनुसंधान अनुभव था। प्राध्यापक बोरकर के अनुसंधान क्षेत्र में उनसे अधिक जानकार मैंने शायद ही कोई देखा हो। यह बात साफ है कि यदि प्राध्यापक बोरकर न होते, तो मैं एम. आई. टी. में नहीं होता!

पूर्वस्नातक अध्ययन के दूसरे या तीसरे वर्ष मैंने पूज्य विपस्सनाचार्य सत्यनारायण गोइंका तथा उनके सहायक आचार्यों से विपस्सना साधना सीखी। इस साधना की मेरी जिंदगी में बहुत गहरी छाप है। निश्चित रूप से विपस्सना साधना ने मेरे इस शोध निबंध की अनुसंधान दिशा पर प्रभाव डाला है। विपस्सना से अधिक गहरा उपदेश शायद इस दुनिया में न हो।

विपस्सनाचार्य ऊ बा खिन गोइंकाजी के विपस्सना गुरु थे। इस कारण, विपस्सनाचार्य ऊ बा खिन को याद करता हूँ। सिद्धार्थ गोतम संयक संबुद्ध, जो सम्भवतः इस दुनिया के अब तक के सर्वश्रेष्ठ मनोवैज्ञानिक हैं, ने यह साधना प्रथम सिखायी, और यह साधना विपस्सनाचार्यों की वन्शावली द्वारा गोइंकाजी तक चली आई। इस कारण, सिद्धार्थ गोतम संयक संबुद्ध को याद करता हूँ।

अन्ततः, अपने परिवार तथा बचपन के बारे में कुछ लिखूँगा। मैं माता-पिता और बहन के साथ बम्बै में बड़ा हुआ। माता-पिता ने अपनी ओर से श्रेष्ठसंभव परवरिश की, उसका आभारी हूँ। बचपन के दोस्त याद आते हैं। और याद आता है कर्णाल, मेरा ननिहाल, जहाँ गरमी की छुट्टियों में अपने ममेरे भाई-बहनों के साथ खेला करता था!

The last two paragraphs above need to be translated into English.

I learnt Vipassana meditation during the second or third year of my undergraduate education in India from venerable Vipassana teacher Satyanarayan Goenka and his assistant teachers. I do not think I have seen a more profound teaching than this in my life. This teaching has had a profound influence on my life; for sure, it has affected the research direction of this doctoral dissertation of mine. To Goenkaji and his teacher, venerable U Ba Khin, my deepest gratitude. And to Siddharth Gotama Buddha, who taught Vipassana in the first place, and arguably, the best understander of mind, matter, and their interaction, the world has, yet produced, my profuse gratitude. I continued my Vipassana training at the Vipassana Meditation Center, Dhamma Dhara in Western Massachusetts, and I thank all the teachers there.

I grew up in Bombay (now, Mumbai), with my parents and sisters. My parents gave me the best possible upbringing they could, and for that, I am grateful. I remember my numerous friends, cousin brothers and sisters, and their love, when I was a kid.

Back to MIT years, there are four people who have had the most profound influence in my life. They are Kumkum, Ruth, Jayme and Martin. From them I have learnt so much about life, that without them, this journey would for sure have been impossible. India was where I was born, but US is where I have grown up (well, at least partly grown up!). A large part of the reason for that is them.

Christine Kerrigan has been a close friend during the last year. Her presence made life richer.

A big thanks and a round of applause to MIT, Cambridge and US. It has been a wonderful place for me; it has been the land of my dreams, in many ways.

I thank each and every interaction with each and every person and thing that I have had in my life; each of them has hopefully helped me grow.

Finally, at the end, as was, at the beginning, I would like to remember my thesis advisor, Prof. Sanjoy Mitter again. This thesis might not have been possible without him.

Contents

List of Figures

यदा हवे पातुभवन्ति धम्मा
आतापिनो झायतो ब्राह्मणस्स,
अथस्स कङ्खा वपयन्ति सब्बा
यतो पजानाति सहेतुधम्मं

-महावग्गपाळि १

When things become manifest
To the ardent meditating brahmin,
All his doubts vanish
Because he understands [each] thing with its cause.

- Mahavagga Pali 1

In his influential book *The Structure of Scientific Revolutions*, Thomas Kuhn argues that a field of scientific inquiry is made up by *paradigms* and *puzzles*. He describes paradigms as models for research, a general problem area sharing a common formulation, a framework in which it becomes possible to ask *'valid'* questions. Puzzles are concrete applications, conjectures, open problems. Most scientists piece together puzzles and it is this activity which Kuhn calls *normal science*. The term puzzle suggests *spielerei* - playing games. This negative connotation is - so it said - unintentional. By formulating puzzles, a scientist can focus on specific questions, questions lead to answers, answers are the products of scientific research.

This structure of scientific inquiry is very much present in (applied) mathematics in general and in the theory of dynamical systems in particular. However, there has been an unfortunate unexplicable total domination of puzzle solving. Paradigms have been muted, supressed, not spoken about, let alone scrutinized, rejected, updated. Examining and formulating paradigms has achieved a reputation in mathematical circles as being soft: it leads to too many definitions and not enough theorems. Solving puzzles, on the other hand, is considered a serious activity, requiring intelligence, mathematical culture, virtuosity. The ultimate of mathematical achievement is to solve a puzzle (a conjecture) formulated by someone else preferably in another century. Thus, we have attained a complete reversal in which posing paradigms is considered *spielerei*, we find ourselves in a situation in which proving theorems, not building theories, appears to be the aim of mathematical research.

-Jan Willems in his paper Models for Dynamics

Stylistic note

The main text and math formulae in this thesis have been typeset using LaTeX in Garamond font. However, the text and math formulae in the figures in this thesis are typeset using LaTeX in Computer Modern font. The major noticeable difference that this makes is with sets: sets in LaTeX, when typesetting in Garamond font are typeset with \mathscr, for example, $\mathscr{A}$. When using Computer Modern font, however, sets are typeset with \mathcal, for example, $\mathcal{A}$. These will be assumed to represent the same object. There are other minor differences, too. In general, the reader should note the slight difference in font in the main text of this thesis and the figures in this thesis.

Chapter 1

Introduction: Digital communication architectures, why or why not?

What many of us fail to realize is that the last four hundred years are a highly special period in the history of the world. The pace at which changes during these years have taken place is unexampled in earlier history, as is the very nature of these changes. This is partly the results of increased communication, but also of an increased mastery over nature, which on a limited planet like the earth, may prove in the long run to be an increased slavery to nature. For the more we get out of the world the less we leave, and in the long run we shall have to pay our debts at a time that may be very inconvenient for our own survival.

-Norbert Wiener

1.1 Introduction

Communication is a basic need of most (if not all) living beings. For humans, verbal language has developed as a very important form of communication and language is supposed to be the reason why humans have dominated other living species on this planet [Hay05]. With the advance of technology, communication has flourished over long distances in some form of language, be it speech, images or text. Also has flourished communication over long distance between machines or between humans and machines using some form of language.

This thesis is concerned with a theory of communication in scenarios where communication is desired over long distances between many users. There are various sources which various users want to communicate to each other with various guarantees over a communications medium. Examples of such media are wireless and internet. In the example of traditional wireless, telephone, the sources are voice of various users and the medium of communication is the atmosphere. In the case of the internet, the sources can take various forms, for example, text, audio, video, etc., and the medium of communication is the internet architecture. Wireless communication over the atmosphere or communication over the internet requires communication technology in the modern sense of the word, and not in the sense of smoke signals and drum rolls that primitive societies used as technological aids to communication.

In what follows in this chapter, I have taken material directly, on various instances, from the first chapter in Gallager's book [Gal08], a version of which can also be found on MIT OpenCourseWare [Gala] and the first and second video lectures of his course on Digital Communications which can be found on the MIT OpenCourseWare [Galb].

Also, in what follows throughout the thesis, when I use the word "wireless communications," I would be referring to traditional wireless telephony. Wireless communication is used today, not just for voice communication, but also for transmission of various kinds of data.

1.2 Chapter outline

This chapter is a high-level discussion of analog and digital architectures and factors which determine which technology (analog or digital) should be implemented and which should not, and whether or not any technology should be implemented. We also discuss the contributions of this thesis, which are partly motivated by the factor of cost / performance which determines in an important way, which technology is implemented.

In Section 1.3, we discuss analog and digital point-to-point communication systems.

In Section 1.4, we discuss the reasons for which technology (analog or digital) is implemented and which is not. We also discuss reasons which should be considered when determining which technology is implemented, or whether a technology should be implemented at all, and which are not considered. These reasons belong to various categories as discussed in this section. One very important reason is cost / performance.

In Section 1.4.2, we discuss, why the reason of cost / performance is a very important reason which determines which technology is implemented. In this section, we also consider the reason of human nature and how it determines what exists in this world. In Section 1.4.3 we discuss that the reason of cost / performance, just by itself is not well understood in multi-user settings. Understanding this is one of the motivations for this thesis.

In Section 1.6, we discuss multi-user analog and digital communication systems. Another motivation of this thesis is to undertand why digital architectures are good on a conceptual level. These are the two flavors of this thesis: understanding digital architectures from the point of view of cost / performance in multi-user communication systems and understanding why digital separation based architectures are good on a conceptual level.

These two flavors are discussed in Section 1.7.

At the end, Section 1.8 discusses the organization of the rest of this thesis.

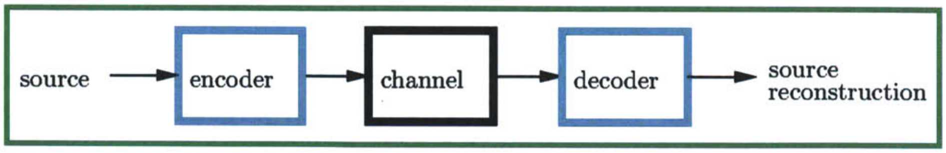


Figure 1.1. A general point-to-point communication system

1.3 Point-to-point analog and digital communication systems

For simplicity, consider a point-to-point communication problem: there are 2 users, and one user wants to communicate a source to the other user over a medium (which is synonymous with a channel in the point-to-point setting) with some guarantee. In the point-to-point setting, the medium will be called, the channel. For example, a person in Boston wants to send an e-mail to his parents in Mumbai over a simplified point-to-point internet channel. The guarantee is that the e-mail should have no errors in that it should be received exactly as it was sent. Another example is that a person in Boston wants to talk to his parents in Mumbai on the phone over a simplified point-to-point wireless channel. The guarantee is that the parents should be able to hear what the son spoke and make sense out of it even though the reproduction might not be exact.

This is done with the help of an encoder and a decoder. The encoder encodes the source. The encoded source is the input to the channel. The channel communicates the input with errors. The decoder reconstructs the source from the erroneous channel output. In the example of wireless, a cellphone acts as both an encoder and a decoder.

1.3.1 A general/analog point-to-point communication system

A general point-to-point communication system is the composition of an encoder, channel, and a decoder, see Figure 1.1.

1.3.2 A digital point-to-point communication system

Modern day communication systems are usually digital. Digital communication systems are communication systems which use a digital sequence as an interface between the source and the channel input, and similarly, between the channel output and the final destination.

A digital sequence is a sequence made up of elements from a finite alphabet, for example, the binary digits (bits) $\{0, 1\}$, the decimal digits $\{0, 1, 2, \dots, 9\}$, or the letters from the english alphabet. Binary digits are almost universally used for digital communication and storage, and hence, when we say digital communication, we would mean that the interface is binary. Communication sources, for example, speech waveforms, image waveforms, and text files, are

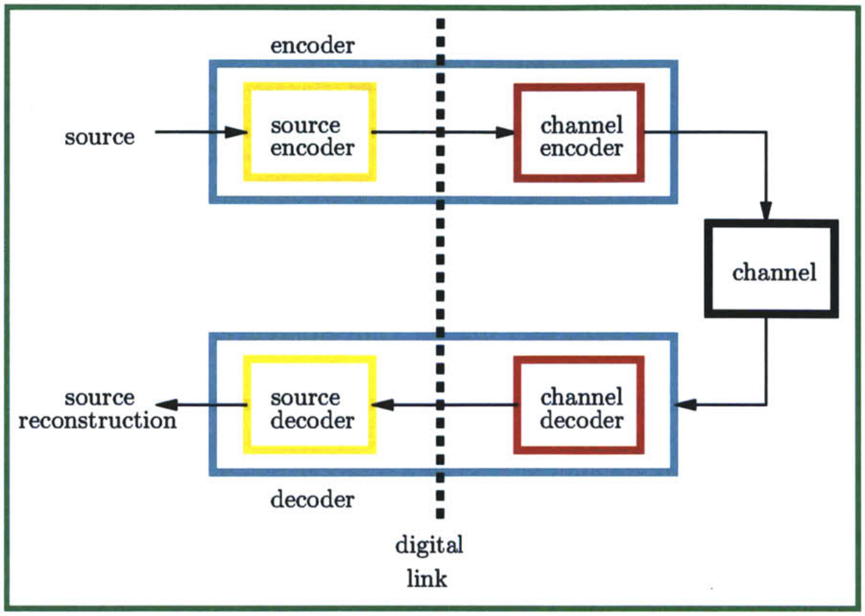


Figure 1.2. Placing a digital (usually binary) link between source and channel. The source encoder converts the source output to a binary sequence and the channel encoder (usually called a modulator) processes the binary sequence for transmission over the channel. The channel decoder (demodulator) recreates the incoming binary sequence, and the source decoder recreates the source output

represented as binary sequences. The binary sequence is then converted into a form suitable for transmission over particular physical media such as a cable, twisted wire pair, optical fiber, or electromagnetic radiation through space.

Thus, a point-to-point digital communication system is a special case of a general communication system where the encoder is the composition of a source encoder and a channel encoder, and the decoder is the composition of a channel decoder and a source decoder. See Figure 1.2.

The idea of converting an analog source output to a binary sequence was quite revolutionary in 1948, and the notion that this should be done before channel processing was even more revolutionary. By today, with digital cameras, digital video, digital voice, etc., the idea of digitizing any kind of source is commonplace. The notion of a binary interface before channel transmission is almost as commonplace. For example, we all refer to the speed of our internet connection in bits per second.

The input to the channel encoder is a digital sequence, usually a binary sequence.

Note, finally, that digital schemes are a special case of analog schemes.

1.3.3 A note on digital point-to-point communication systems

In general, an analog source can be converted into a digital sequence as follows: sample the analog source very finely and quantize the sampled source very finely. The finer the sampling and the quantization, the closer, the digital sequence is a replica of the original analog source. In the limit, the digital sequence completely represents the analog source.

In this thesis, we would consider questions related to the optimality of digital architectures: that is, whether digital architectures can perform as well as analog architectures.

Since, by very fine sampling and quantization, the digital representation will approach the analog source as closely as we want, it seems that digital architectures can be made to perform arbitrarily closely to analog architectures. This is not necessarily that direct because usually, there is a limitation on the resource consumption in the system. Very fine sampling and quantization can consume a lot of switching energy and thus, the digital scheme, built this way, might end up consuming a lot more energy than the analog scheme. Thus, it is a priori unclear whether digital schemes will perform as well as analog schemes.

Another very important point is that we do not want the digital interface to be of very large cardinality. In fact, we want that cardinality to be independent of the source. If the source quantization is very fine and depends on the source, this can lead to a very large cardinality of the digital interface, and we would like it to be as small as possible: as stated before, we prefer if it is binary.

1.3.4 Separation based architectures

In practice, digital architectures are built in the following manner:

1. The source encoder compresses the source to within an allowable distortion level. The output is a binary sequence.
2. The channel encoder and decoder help to communicate the binary sequence *reliably* over the channel. By "reliably", we mean that the error probability in the detection of the binary sequence is "very" small.
3. The source decoder reconstructs the source from the original binary sequence

Separation based architectures are called so because they separate the source and the channel: the source encoder and decoder are independent of the channel and the channel encoder and decoder are independent of the source. In addition, however, when defining separation based architectures, we require "reliable" communication of the binary sequence over the channel. A general digital communication scheme may or may not be a separation based scheme with the above definition.

Separation based architectures are thus, a special case of digital architectures. When proving results concerning the optimality of digital communication, we will always do it by the use of separation based architectures.

1.4 Factors which determine which technology is implemented and factors which should be considered when determining which technology is implemented or whether a particular technology should be implemented, but are not

In Subsection 1.4.1, we discuss the reasons which determine which technology is implemented and factors which should be considered but many times, not, when determining which technology is implemented, and whether a technology should be implemented in the first place. These are a whole variety of reasons: cost / performance (in other words, profit motive), architectural, social, etc. These reasons are discussed in subsection 1.4.1.

Among all these reasons, there is one very important reason determines whether a technology is implemented. This is the reason of cost / performance. The reason why this is the factor of cost / performance is the a very important factor, is speculated on, in Subsection 1.4.2. Another consideration is human nature, and its effect on whether a technology is implemented or which technology is implemented, and this is also speculated on, in Subsection 1.4.2.

In Subsection 1.4.3, we argue that the cost / performance reason, just by itself, without the other reasons, is not well understood in multi-user settings. This leads us to one of the motivations for this thesis: understanding cost / performance in multi-user scenarios.

1.4.1 The reasons, both, which determine and do not determine which technology is implemented

There are a number of reasons why communication systems now usually contain a binary interface between source and channel (that is, why digital communication is now standard, and has replaced analog communication). The reasons mentioned below all into to two categories:

- those which drive what finally gets implemented in practice. In my opinion, one of the main factors here is cost / performance. This is a simplification, but still very true.
- reasons concerning simplicity of architectures, understanding related to the functioning of the communication system arising from the simplicity of architecture, and social reasons. This can be sub-catagorized into various reasons.

These reasons are discussed below:

Cost / performance considerations

1. *Cost / Performance characteristics:* This has three components

   (a) *Cost of digital hardware:* Digital hardware has become so cheap, reliable and miniaturized, that digital interfaces are eminently practical. This has been possible because of 20/30 generations of Moore's law. Very roughly, since the number of transistors that can be placed on a particular area of a chip doubles every 18 months (or two years): the cost to put a transistor on a chip decreases and thus, the cost for obtaining the same performance decreases. Things are more complicated than this, but this is the rough idea.

   (b) *Standardization:* Digital communication is a general way of communication irrespective of the source and the channel: the source is first converted to binary and these binary sequences are then communicated over the channel. Analog design can be much more complicated, and much more of an art than digital design, and hence, it costs more: an analog designer usually gets paid much more than a digital designer! The cost of a chip is approximately related to the cost of development divided by the number stamped out. Standardization leads to lowers cost of development, and thus, cheaper chips. This is partly elaborated on in Gallager's first and second video lectures in the series [Galb].

   (c) *Separation theorem*: One of the most important of Shannon's information theoretic results is that if a source can be transmitted over a channel with certain distortion guarantee in any way at all, it can also be transmitted using a binary interface between the source and the channel, without any significant change in the use of system resources like energy and bandwidth. This is known as the source/channel separation theorem. Thus, given an analog architecture which achieves some performance, the same performance can be achieved with the same energy and bandwidth consumption by using a digital architecture. As a result, the best possible digital architecture will cost the same from the point of view of energy/bandwidth consumption as the best possible analog architecture. This does require assumptions that delay in reception of the source is not a concern, but making this assumption is probably an okay approximation.

   Note 1.1. ***The above optimality of separation based architectures was proved by Shannon in [Sha48] for reliable point-to-point communication. In [Sha48], Shannon also stated the optimality of separation based architectures for communication with a fidelity criterion (distortion criterion), and proved it in [Sha59]. Separation, in fact, does not hold in the most general possible multi-user scenarios. This point of whether separation holds in multi-user scenarios is commented on in Subection 1.7.1, and proving optimality of separation in certain multi-user scenarios is one of the main foci of this thesis.***

Other reasons

The other reasons are classified as technical/technological reasons and social reasons.

Technological reasons:. This can be categorized further, as follows:

- *Layering and hence, simpler conceptualization*: This is the same as the standardization reason, but viewed very differently. Digital architectures give a simple conceptual way of building architectures: first convert the source into binary, then, communicate the binary sequences over the channel, and finally, get a reconstruction of the source. In a digital communication system, the action of the source encoder is independent of the channel (that is, depends only on the source) and the action of the channel encoder is independent of the source (that is, depends only on the channel). This is not necessary from the definition of a digital communication system, but this is how digital communication systems are constructed. in practice. A most general analog communication scheme would be a complicated non-linear function of the source and the channel. Digital communication linearizes this into source coding followed by channel coding, and hence, conceptually much simpler than a most general communication scheme. This is elaborated on in [Gal08] (or the equivalent course notes [Gala])and first and second lectures in the series [Galb]. What is questionable about this point in terms of being a reason why digital architectures are used is that layered schemes can be built even in an analog way, and hence, this cannot be a fundamental reason for why architectures are digital. I will add that others might disagree about this last point.

- *Simpler networking:* A standardized binary interface between source and channel simplifies networking, which now reduces to sending binary sequences through the network. This is elaborated on in [Gal08] (or the equivalent course notes [Gala]). This point is very important, for example, in the case of the internet because the internet architecture consists of a series of links and it is good to have a standardized interface for the input and output of each link. This point is questionable in the sense, again, that one can also have a standardized analog interface. Probably, what is good about a standardized binary interface is that its cardinality is the smallest possible cardinality that an interface can have, that is 2, instead of a standardized analog interface which will have infinite cardinality.

- *Physical performance:* This performance is *different* from Reason 1, which discussed the cost/peformance characteristics. This reasons refers to the "physical" performance of the system.

  Copying directly from [OPS48], which talks about the advantages of PCM (pulse code modulation) over analog systems, in particular, frequency modulation:

  "In most transmission systems, the noise and distortion from the individual links cumulate. For a given quality of over-all transmission, the longer the system, the more severe are the requirements on each link. For example, if 100 links are to be used in tandem, the noise power added per link can only be one-hundredth as great as would be permissible in a single link.

Because the signal in a PCM system can be regenerated as often as necessary, the effects of amplitude and phase and non-linear distortions in one link, if not too great, produce no effect whatever on the regenerated input signal to the next link. If noise in a single link causes a certain fraction p of the pulses to be regenerated incorrectly, then after m links, if $p \ll 1$, the fraction incorrect will be approximately mp. However, to reduce p to $p' = \frac{p}{m}$ requires only a slight increase in the power in each link as we have seen in the section on threshold power. Practically, then, the transmission requirements for a PCM link are almost independent of the total length of the system. The importance of this fact can hardly be overstated."

The above refers to the section on threshold power, part of which basically says that in a PCM system, as the signal power is increased, after a particular point, even a slight increase in signal power will decrease the probability of error by a huge amount. The reader is refered to [OPS48].

In a usual analog system, however, such performance cannot be achieved. This is because a usual analog system uses amplifiers to amplify the analog signal and noise is amplified in the same proportion as the signal is amplified.

I wonder, however, if there are smart ways of building analog systems which achieve the above performance of a PCM system.

These reasons are a mixture of my own understanding synthesized with talking to some experts and the reasons which I have taken from [Gal08] (or equivalently, [Gala]), [Jr.a], and [OPS48]. The reader is refered to the first chapter, each in [Gal08] or equivalently, the course notes [Gala], and the first two videos of the corresponding course [Galb] , lecture notes [Jr.a] and the first video of the corresponding course [Jr.b], and [OPS48], for a more detailed exposition.

There are societal reasons which should drive which technology gets implemented, they are:

Societal reasons: Some of them are the following:

- *Health:* World Health Organization has declared that cellphone radiation might cause cancer [WHO]. Thus, the communication schemes should also be designed in a way such that they reduce such risks. Potentially, some particular electromagnetic waveforms are more harmful than others, and this can potentially determine, what architectures to use

- *Is it interesting?:* Digital design is more modular than analog design. However, fact is also, that digital design is much less interesting than analog design. Analog design is much more of an art, whereas digital design is more, just a process which needs to be implemented. Another way of saying this is that digital design is a much more automated process (the automation might be done by human, and not necessarily, a machine) compared to analog design. I wonder about the ramifications of this in the workplace. Because of the modularity got out of digital design, and the division of labor it can produce, what is the effect of this on the individual working in the workplace? Is the individual just playing the role of a nail in a big machine, or does the individual see work as a whole, and a process of human growth?

There are further societal reasons, which should drive whether technology like wireless communications or internet gets implemented in the first place or not, whether in an analog or a digital manner. The reasons of health also falls under this bullet.

- *Health:* Again, let us take the example of wireless. As we said above, the World Health Organization has declared that cellphone radiation might cause cancer. It is totally possible that this means that wireless radiation at radio frequency, but with the amount of energy needed in the cellphone radiation is going to cause cancer irrespective of anything else, and this factor should be taken into consideration when determining whether a technology like wireless should be implemented or not. This is a reason for whether wireless technology should be implemented or not, and not a reason for whether analog or digital.
- *Concentration:* There are unverified studies which show that concentration can suffer in the presence of too much electromagnetic radiation from computers and potentially cellphones, and this should determine, whether a technology is implemented or not. This is a reason for whether wireless technology should be implemented or not, and not a reason for whether analog or digital.
- *Environment:* Again, let us take the examples of wireless. The environmental impact of manufacturing cellphones, managing the base-stations, the waste dump produced in the process, etc, is another factor which should be taken into consideration.This is a reason for whether wireless technology should be implemented or not, and not a reason for whether analog or digital.
- *Ramifications of living in a globally interconnected world:* The internet has made the world very connected and fast paced. Is this the kind of super-globally connected, fast paced, multi-tasking way of living conducive to human happiness or not? This is a reason for whether wireless technology should be implemented or not, and not a reason for whether analog or digital.

This list of societal factors is incomplete. However, the point is that as we said before, Reason 1 is the primary (and probably the only) reason which determines whether a technology is implemented and which technology is implemented. The reason for this is speculated on, in the next sub-section.

Just as a note, reasons like effect on health can definitely be studied before a technology gets implemented. Reasons like ramifications of living in an interconnected world can probably only be understood in hindsight unless someone really wise with a lot of vision ends up looking at the problem. This is another discussion, for another day.

■ 1.4.2 A very important reason for which technology is implemented: cost / performance, and human nature

Reason 1, that is cost/profit motive, is *unfortunately a very important reason* which ultimately drives which technology gets implemented. This is due to the fact that we live in a partly dysfunctional capitalist society where empathy takes a back seat. Capitalism can, in general, be good. Here, I am commenting on the present capitalist structure, and not saying that capitalism is bad in general. The way present day capitalism is structured really might be a physical manifestation of the brain anomaly known as psychopathy [Ron11]. The bigger picture of what is good for the world is lost many times because of the overemphasis on monetary gains. This is just my opinion and that of some others, and others might disagree.

I should add that things are more complicated than what I have written above. Technology is a process, and what gets implemented tomorrow is a function of what exists today. However, in my opinion, the fact still remains that at the fundamental level, a lot of things happen in the society because of some kind of monetary motive. One might want to package things in a way that it is not just monetary motive, but in my opinion, it is monetary motive which is at the base of a lot of things. Others disagree with me.

Questions like the above bring us to some of the most fundamental questions in economics, and I would leave any further discussion here.

There is one reason which might be more fundamental than cost / performance, and that is human nature: many of us want to be someone and do something, and monetary urge beyond what one needs, is but one facet of this urge. This is not a negative urge. But it can become negative without wisdom. For example, with the question of wireless communications, one can argue in a positive way that it helps people keep in contact, and hence, it is a good thing. On the flip-side, are the negative effects to health and environment, and questions about ramifications of living in a global interconnected world. Evidently, the reason of people being able to keep in touch, won. I wonder, why? I wonder whether the people who made the initial breakthroughs, both theoretical and implementation in wireless systems, really cared about people being able to keep in touch or whether their real motivations lay somewhere else. I would like to believe that their real motivations lay somewhere else: doing something exciting, fun, making something out of their life, making money, and being famous. They might make themselves "feel good" by telling themselves that this system will do something good by helping people keep in contact with each other. If this were not the case, I would like to believe that there would have been more of a debate on whether systems like wireless got implemented. I am just speculating here. Also, things, of course, are more complicated than this. The point however, is that "I am" is a huge factor in many bad things happening in this world.

Questions like the above bring us to some of the fundamental questions about human nature, and I would leave a further discussion here.

There are various other reasons. I am taking the view of a skeptic here, and stating only the negative reasons. I should say that in my opinion, these negative reasons are very fundamental reasons, probably the most fundamental.

A corollary of the fact that Reason 1 is a very important driving motive behind which technology gets implemented is that if someone came up with an analog architecture which saved billions or trillions of dollars compared to a digital architecture, and thus, someone find ways of making money, then some one would find a way of implementing the analog scheme, irrespective of anything else. Again, things are more complicated than this because overhauling a huge existing system is a non-trivial matter.

1.4.3 The cost / performance reason, even just by itself, is not understood well enough, in multi-user settings

Also, as regards Reason 1, it is Reasons 1a and 1b and not Reason 1c which have been crucial in replacing analog technology with digital technology. ***Reason 1c is in fact not understood in the biggest communication problems like wireless and internet.*** Shannon proved 1c under two assumptions:

1. The setting is point-to-point: there are two users and one user wants to communicate a source to another user
2. The action of the channel as a transition probability is known

Real situations, for example, wireless or internet, do not follow this paradigm. Let us consider the example of wireless:

1. Wireless is a multi-user problem and not a point-to-point problem.
2. The wireless medium is time-varying and only partially known, that is, its action cannot be modeled as a known transition probability. Of course, the channel state can be ascertained to some extent by exchanging messages between the users but still, the fact is, that the action of the channel as a transition probability is only partially known.

Thus, it is unclear, for example, in the wireless communications problem whether digital communication is the best thing to do as regards cost / performance related to point 1c. It might well be the case that one comes up with an analog architecture which saves billions/trillions of dollars in energy costs in the electromagnetic waves that cellphones emit, and leads to much more than the energy savings from reason 1a and 1b, and enough to overhaul the wireless communication system to analog. It takes a lot to overhaul such a big system like wireless, but the prospect of enough profit might well, do it.

In fact, it is known that in certain multi-user communication problems with correlated sources, digital communication is not optimal in the sense of reason 1c, see, for example, [Gas02].

Many people believe that the most fundamental problems in wireless communications are related to reliable communication of binary sequences over the wireless medium, in particular, for example, understanding the effect of fading. Another important question is to model the propagation of electromagnetic waves through random media.These are definitely a very important question irrespective of whether the architecture is analog or digital because fading will have to be dealt with in either architecture. However, this is a question, which is important irrespective of whether one wants to use analog architecture or digital architecture. In my opinion, an even more fundamental question is whether one wants to build architectures digitally for wireless communications in the first place. Others would disagree because they would say that things are digital, and that is how things are going to be. There is truth to that, too.

The previous discussion leads us to one of the motivations for this thesis.

■ 1.5 Analog or digital from the point of view of cost / performance: one of the motivations for this thesis

From the above discussion, it follows that from the perspective of cost / performance, it is not entirely evident that digital communication systems should be used in multi-user communication problems. This was one of the questions I had in mind when I started working on this problem. I would not say that I have answered this question, but this thesis does provide some understanding of point 1c in multi-user scenarios when the medium action might only be partially known.

Cost / Performance: If it turned out to be the case that the energy consumption by an analog architecture in certain multi-user problems like wireless are much less compared to a digital architecture, in the long run, compared to the cost of building digital architectures, then, the system might possibly get overhauled. In other words, it is possible that an analog architecture exists for which the cost savings from reason 1c are more than the cost savings of reasons 1a and 1b of the best possible digital architecture By optimality, we mean the following: digital architectures are optimal if, given an analog architecture which consumes some amount of system resources for some performance, a corresponding digital architecture exists which consumes roughly the same or lesser system resources and provides roughly the same or better performance. Others disagree with me on this point, in that it is almost next to impossible to overhaul a big system like wireless.

Another flavor of this thesis is intellectual/conceptual understanding of why separation holds, and this is discussed in Section 1.7 and Subsection 1.7.2.

Before I go into a description of our results, I'll first look at multi-user networks.

1.6 Multi-user analog and digital communication systems

Let there be N users. N might vary with time. User i wants to communicate source X_{ij} to user j, $1 \leq i, j \leq N$ to within some guarantee G_{ij}, over a medium m. X_{ij} is reconstructed as Y_{ij} at user j. This is accomplished with the help of modems (modulators/demodulators) h_i at user i, $1 \leq i \leq N$. In the example of wireless communications, the medium is the atmosphere and the modems are cellphones and base stations (the cellphone towers). This is a general multi-user communication system. See Figure 1.3. A more elaborate high-level description and rigorous description is in Chapter 3.

The guarantee depends on the particular situation. When talking on the phone, the guarantee is that communication should happen to within a distortion level. When sending an e-mail, it is the e-mail is received perfectly (this is not possible in a noisy system, and the way this is abstracted is by building systems where the probability of error is very small).

At time t, the modem h_i at user i takes inputs $X_{i1}(t)$, $X_{i2}(t)$, ..., X_{ij}, ..., $X_{iN}(t)$. At time t, h_i produces an output $O_i(t)$ which is an input to the medium of communication m. The various inputs $O_i(t), 1 \leq i \leq N$ into the medium "mix" and noise is added on top of it. The medium produces an output $I_i(t)$ which is an input to the modem h_i. Based on O_i, the modem produces outputs Y_{1i}, Y_{2i}, ..., Y_{ji}, ..., Y_{Ni}, where Y_{ji} is the reproduction of X_{ji}, the source that user j wants to communicate to user i. X_{ij}, Y_{ij}, O_i, I_i, are evolving in time. The story is the same at each user i.

In a digital multi-user communication system, each modem is digital. A digital modem h_i is portrayed in Figure 1.4. At user i, the sources $X_{i1}, X_{i2}, \ldots, X_{iN}$ are first converted to random binary sequences by the source encoders. These binary sequences are communicated reliably over the medium with the help of medium modems at the various users. Finally, at user i, the source decoders help produce the reproductions Y_{1i}, ..., Y_{Ni}, of X_{1i}, ..., X_{Ni}, respectively. The story is the same at each modem h_i.

Next I discuss the two flavors of this thesis.

1.7 The questions asked and answered in this thesis/ the *two* flavors of this thesis

This thesis, as stated in the abstract, has two flavors, and here, they are stated in the order which is the reverse of the order in the abstract:

1. *Optimality of separation from the perspective of cost/peformance:* Understanding reason 1c, that is, whether source-channel separation holds in multi-user communication problems when the medium description as a transition probability is only partially known. This has already been stated in Section 1.5. Discussion on the nature of our results concerning this point is there in Subsection 1.7.1

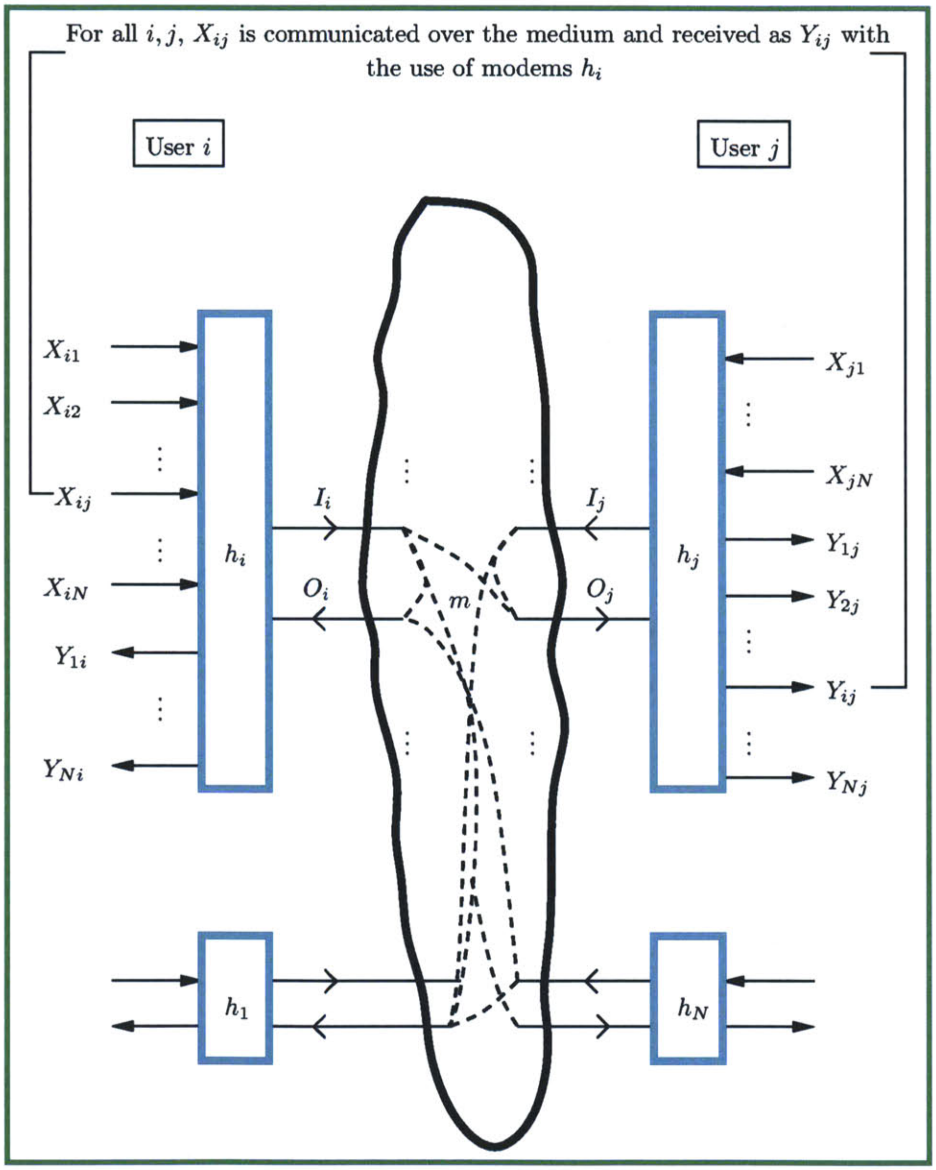


Figure 1.3. A general multi-user communication system

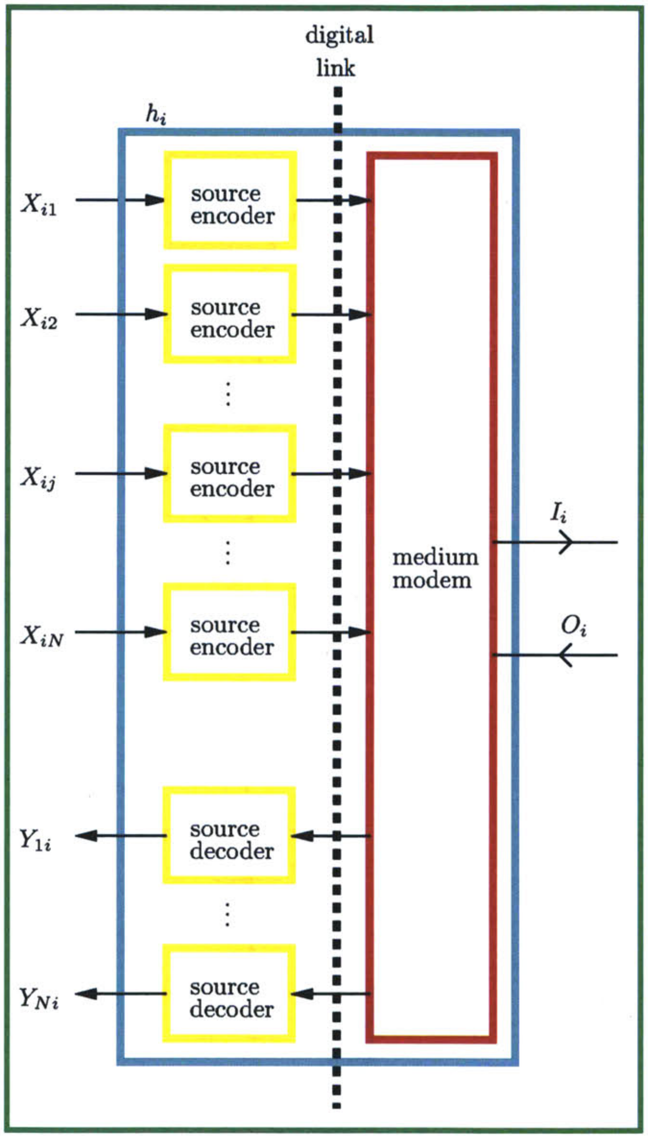


Figure 1.4. A digital modem

2. *Intellectual/Conceptual* There is no fundamental understanding, as regards a separation theorem, even in the point-to-point setting, when the channel behavior as a transition probability is only partially known. Also, there is no fundamental understanding, as regards a separation theorem in multi-user problems. We will make some statements about separation in both the point-to-point case when the channel is only partially known and also in the multi-user case when the medium is only partially known. Also, Shannon's proof of source-channel separation for a known channel in the point-to-point setting is, in my opinion, not very transparent, and we believe that our proof, more general in setting in many ways, is also more transparent. Thus, we want to understand, why separation holds, on a conceptual level. This is discussed in Subsection 1.7.2.

1.7.1 Understanding reason 1c: does separation hold in multi-user communication problems?

Recall the multi-user communication problem discussed in the previous section. We consider the problem of communicating soures X_{ij}, $1 \leq i, j \leq N$, from user i to user j within a guarantee G_{ij} over a medium m. The guarantee that we will use is that the source X_{ij} is communicated to within a distortion level D_{ij} under a distortion metric d_{ij}. The medium m is only partially known. Mathematically, this is abstracted out by saying that the medium m might belong to a family of transition probability matrices.

We will make the following assumptions:

1. The sources X_{ij} are independent of each other
2. The users can generate random codes
3. In order to prove rigorous results, we will assume that the distortion measures d_{ij} are additive. However, as we shall see from the nature of the proofs, the results should hold for permutation invariant distortion measures also. Permutation invariant distortion measures are those for which rearranging the input and the output in the same way does not change the distortion between the input and the output
4. Delays do not matter: more precisely, an arbitrary but finite delay is allowed between the transmission of the sources and their reception

Under these assumptions, we prove that digital communication is optimal from the point of Reason 1c: that is, we will prove that assuming random-coding is permitted, if there exists some architecture to communicate independent sources X_{ij} over a partially known medium m to within a distortion D_{ij} under a distortion metric d_{ij}, then, there exists a digital communication scheme which accomplishes the same, and has the same bandwidth requrements as the original scheme and essentially the same energy/power consumption as the original scheme.

Thus, from the point of view of reason 1c, for communication problems which satisfy the above assumptions, digital architectures are as good as analog architectures. Since digital architectures are already better from point of view of reason 1a and 1b, and since, it follows that from the view-point of cost / performance, digital architectures are optimal.

Assumption that the sources X_{ij} are independent of each other is necessary. Gastpar [Gas02] provides examples of multi-user problems where sources are correlated and separation is not optimal from the point of view of reason 1c.

In Chapter 2, we will prove the optimality of digital communication in the point-to-point setting. This has been done by Shannon in [Sha59]. Shannon, however assumed that the action of the channel as a transition probability is known. We do not make this assumption. In Chapter 3, we prove the optimality of digital communication in the multi-user setting under the above assumptions. This is done by reducing the problem to a point-to-point problem and an inductive argument. In Chapter 4, we will see, how this result is partially applicable to the wireless problem. The wireless problem is complex, and we can only capture some features into a mathematical abstraction. As we shall see, the assumption of independence of sources only holds partially. Also, delays matter. However, our modeling does offer partial justification in terms of reason 1c for the use of digital architectures.

I'm not claiming that I'm solving a practical problem or that I am solving the wireless problem. To summarize my view-point:

1. First I claim that the question, "Are digital architectures optimal from the point of view of cost / performance in multiuser settings, for example wireless" is a question that is not entirely well understood. As I have said some others disagree with me in that they believe that the implemented architectures, for example, in the wireless problem, will be digital, no matter, what, and hence, this is not a question of any concern at all.

2. Then, I prove that under various assumptions (stated before), digital architectures are in fact optimal from the point of view of reason 1c, and thus, from the point of view of cost / performance. Finally, we see, to what extent the assumptions hold in the wireless example.

Next, we come to the second point of this thesis, which is, why does separation hold in the first place, on a conceptual/intuitive level

1.7.2 Why does separation hold on a conceptual level?

Shannon proved the optimality of digital communication in the sense of Reason 1c (source-channel separation) for reliable communication in [Sha48] and communication with distortion in [Sha59] in the point-to-point setting.

The very important contribution of these works was not just proving source-channel separation but also, simple expressions for the information content of a source, the minimum number of binary sequences needed to compress a source to within a distortion level (the rate-distortion function) and the maximum rate of reliable communication over a channel.

The problem of maximum rate of reliable communication over a channel is an infinite dimensional optimization problem for which, Shannon provided that a corresponding finite dimensional optimization problem expression (called single letterization) exists. This is a mutual information expression. The enormous importance of this expression is that it can in fact be calculated by a machine (or by hand using pen and paper), and thus, it can be figured out, which channels allow reliable communication at which rate. This view has been taken in practice, and a lot of research has, for example, been devoted to finding capacity achieving codes over the AWGN channel, as is clear from Dave Forney's course notes and video lectures [Jr.a] and [Jr.b], respectively.

Similarly, the problem of the minimum number of binary sequences needed to compress a source to within a particular distortion level is an infinite dimensional optimization problem. Shannon provided a corresponding single letterization, the rate-distortion function, which is a mutual information expression, and which can be calculated by machine or by hand. Thus, we know the minimum number of binary sequences needed to represent a source.

The disadvantage of reliance on mutual information expressions in my view (and the view of some others, and there are others who disagree with this view) is that the proof of why separation holds is not evident. As a mathematical proof, one can see the correctness of the result, but on a more intuitive level, it is unclear why separation holds. This is especially true for the rate-distortion problem, and Shannon's proof can be found in [Sha59].

In short, in my view, mutual information expressions and proofs based on mutual information expressions are good to make calculations on which sources can be communicated over which channels but they do not lend much insight into the nature of separation.

Reality can be understood at various levels. A very good example is gravity. The various levels of understanding are:

1. Newton came up with a formula for the force of attraction between two bodies. This is helpful in terms of making calculations and predictions. However, it is totally unclear, how, this action at a distance happens physically
2. There is the postulate of gravitons or gravitational waves which machines are trying to detect. If true, this will tell, how the bodies exchange information, and the force of attraction is produced
3. The third reality is that of direct experience: as humans who can feel, if we can develop the capability to feel that there, indeed, is a force of attraction caused just due to attraction based on masses. This is the best form of understanding reality

The question of why digital communication does not fit exactly into the above framework, but there are some similarities:

1. A mathematical proof that separation holds
2. Intuition based on the mathematical proof, of why separation holds
3. Direct experience based on examples and real engineering problems. No direct experience at the level of self-experience is possible in some of these engineering problems

In my opinion and that of some others, Shannon's proof in [Sha59] provides a mathematical proof that separation in problem of communication with distortion holds, but it does not give much intuition as to why separation holds.

The proof of the separation theorem for communication with distortion in [Sha59] has two parts:

1. *Achievability:* If the channel capacity is larger than the rate-distortion function for a particular source, then communication of the source to within the distortion level over the channel is possible by the separation architecture of source-coding followed by channel-coding
2. *Converse:* If communication of a particular source to within a distortion level is possible over a channel by some architecture, then the channel capacity is larger than or equal to the rate-distortion function for the source for that distortion level. Thus, by achievability, the communication of the source to within the distortion level is also possible by using a source-channel separation based architecture

The direct part is fairly intuitive. Shannon's proof of the converse in [Sha59] seems to work based on a lot of mathematical manipulations with entropy and mutual information and using their properties like convexity. The ideas are those of a standard information-theoretic converse proof.

We provide a proof of the separation for communication with distortion which uses only the definitions of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a particular distortion level. We do not use any simplified mathematical expressions for channel capacity or the rate-distortion function like the single-letter mutual information characterizations. We call this an *operational* proof because it only uses the *operational* meanings of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress the source to within a certain distortion level and not any simplified mathematical quantifications of these quantities. Our usage of the word operational should not be confused with "being operational in practical implementations". The proof of the direct part is the same as that of Shannon. *However, for us, the converse is an achievability: we prove the converse using achievability techniques*, and in my opinion, lends much more insight into the nature of separation than Shannon's proof. The proof also demonstrates a duality between source and channel coding and our proof of separation is built on ideas from that duality. Why separation fundamentally holds is, in my opinion, hidden in this duality.

■ 1.8 Organization of this thesis

In Chapter 2, we prove the optimality of digital communication for communication with a fidelity criterion in the sense of reason 1c (in other words, a universal source-channel separation theorem for rate-distortion) assuming that random-coding is permitted.

In Chapter 3, we prove the optimality of digital communication for communication with fidelity criteria in the multi-user setting assuming that the various sources which the users want to communicate to each other are independent of each other (the setting is unicast) and that, random coding is permitted at the various encoders and decoders.

In Chapter 4, we discuss the partial applicability of the results of Chapter 3 to the traditional wireless telephony problem.

In Chapter 5, we provide an operational perspective on the source-channel separation theorem for rate-distortion. We also provide an alternate proof of the rate-distortion theorem for certain sources which we believe is more insightful than Shannon's original proof.

in Chapter 6, we recapitulate and discuss potential research directions.

■ 1.9 In the next chapter ...

In the next chapter, we prove the optimality of digital communication for universal communication with a fidelity criterion in the sense of reason 1c in the point-to-point setting assuming random-coding is permitted.

Chapter 2

Optimality of digital communication for communication with a fidelity criterion: universal, point-to-point setting

The fundamental problem of communication is that of reproducing at one point exactly or approximately a message selected at another point.

-Claude Shannon

Our difficulty is not the proofs, but in learning what to prove.

-Emil Artin

2.1 In this chapter ...

2.1.1 Introduction

In this chapter, we prove that digital communication is optimal for rate-distortion communication in the sense of reason 1c stated in Section 1.4.1 in Chapter 1. That is, we prove a source-channel separation theorem in the rate-distortion context. A source-channel separation theorem for rate-distortion in the point-to-point setting was hinted at by Shannon in [Sha48] and proved rigorously in [Sha59]. Shannon [Sha59] assumes that the action of the channel as a transition probability is known. We call these, fully known channels. *The main contribution of our source-channel separation theorem for rate-distortion is that we assume that the channel is only partially known. This is abstracted by saying that the channel probability may belong to a set. In mathematical terms same encoding-decoding schemes should work for all channels in the set.*

This is important because real life media like the internet and wireless are only partially known. One way of modeling these situations is to say that the medium action is only partially known as a transition probability. Internet and wireless are multiuser networks. In this chapter, we only prove the optimality of digital communication in the point-to-point setting. *Another important contribution of this thesis is to prove the optimality of digital communication* *in certain multiuser settings where the medium is only partially known.*. This is the subject of the Chapter 3.

Our formulations will be information-theoretic, see Section 2.2. In the information theory literature, encoding-decoding schemes which work for a partially known channel are called *universal. Note that the universality is over the channel, and not the source. We will assume that the source statistics are known.* The reason why we assume the knowledge of source statistics is commented on, in Section 2.19.

In our work, we use the probability of excess distortion criterion instead of the expected distortion criterion for the definition of "communication to within a distortion D". The probability of excess distortion criterion is (2.29) and the expected distortion criterion is (2.28). This use of the probability of excess distortion criterion instead of the excess distortion criterion is crucial to our work. This is commented on in Section 2.12.

Also, throughout, we will assume that random-coding is permitted. That is, the encoder and the decoder are allowed to generate random codes. That is, the encoder can belong to a family of encoders and the decoder has access to the particular realization of the encoder. Errors get averaged out over the random code. Over a fully known channel, if there exists a random code which achieves a particular performance, usually, there also exists a deterministic encoder-decoder which achieves the same performance. However, over partially known channels, random codes can enhance the performance of the system. Mathematically, this happens because since we model a partially known channel as coming from a set of fully known channels, some of the deterministic codes that make up the random encoder-decoder may work well for some channels, and others will work well for other channels, and they can be constructed in such a way that taking an average over all these deterministic encoder-decoder work well for all the channels and achieve a performance which a single deterministic encoder-decoder cannot achieve. This point is discussed precisely in Section 2.17. We emphasize again, that we insist in good performance over "all channels" in the set which make up the partially known channel, rather than good performance averaged over the channels.

The channel model as a transition probability will be very general in the sense that the present channel output can depend, in an arbitrary manner, on all the past channel inputs, all the past channel outputs, and possibly, an initial channel state.

We will hint at the essential connection between source and channel coding that allows separation to be true. This essential connection or duality will not use the definition of channel capacity as a maximum mutual information or the rate-distortion function as a minimum mutual information. Only the meanings of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a certain distortion level will be used. This discussion will be high level. This is the subject of Section 2.14.

Rigorous results will be proved when the source is i.i.d., and the time evolves discretely for the source and channels and additive distortion measures. This is done in Section 2.15.

High level remarks will be made when the source and channel evolve continuously in time, the source is stationary ergodic and the distortion measure is permutation invariant. Of course, separation will not hold for arbitrary stationary ergodic sources and arbitrary channels. The point of this discussion will be to bring out the idea of why results are expected to generalize to this more general setting. Permutation invariant distortion measures are discussed in Sections 2.14 and 2.16. Generalization to stationary ergodic sources evolving in continuous time is discussed in Section 2.18.

2.1.2 A high-level statement of universal source-channel separation for rate-distortion in the point-to-point setting

The following is a high-level statement of the universal source-channel separation theorem for rate-distortion:

High level statement 2.1 (Universal source-channel separation for rate-distortion or the optimality of digital communication for universal communication with a fidelity criterion). *Assuming random-coding is permitted, in order to communicate a random source universally over a **partially known** channel to within a particular distortion level, it is sufficient to consider source-channel separation based architectures, that is, architectures which first code (compress) the random-source to within the particular distortion level, followed by universal reliable communication over the partially known channel. There is sufficiency in the sense if there exists some architecture to communicate the random source to within the required distortion level, universally over the partially known channel, and which consumes certain amount of system resources (for example, energy and bandwidth), then there exists a separation based architecture to universally communicate the random source to within the same distortion universally over the partially known channel, and which consumes the same or lesser system resources as the original architecture.*

We emphasize, again, that universal means that the *same* encoding-decoding scheme should work irrespective of the particular action of the partially known channel, which, as we have said above, belongs to a set of transition probabilities. Also, we emphasize again, that universality is over the channel, not the source.

2.1.3 Chapter outline

The following is the outline of this chapter:

The first part deals with the set-up, definitions, comments on the definitions and a small discussion on important past literature:

Section 2.2 discusses the information theoretic set up and the assumptions that we make. Some assumptions like "delays do not matter" (in other words, arbitrarily large delays are allowed) are crucial to our results whereas other assumptions like the channel input and output sets are finite and time evolves discretely are not crucial and are made only to simplify mathematical technicalities.

Section 2.3 discusses a very important notation that we use: the superscript notation. It is important enough to merit a separate section.

Section 2.4 discusses sources and section 2.5 discusses channels. We will assume that the channel is only partially known, and the channel model is a very general channel model. These sections also discuss in brief, the validity of our source and channel models.

Sections 2.6 and 2.7 discuss analog and digital communication systems respectively. They discuss analog encoders and decoders, digital encoders (source encoder and channel encoder) and digital decoders (channel decoder and source decoder), the make up of point-to-point analog and digital communication systems and resource consumption in point-to-point communication systems. The sections also discuss the problem of point-to-point communication problem, and the particular problem of communication with a fidelity criterion that we study in this thesis. In part, this section makes rigorous, the discussion in Section 1.3.

This is followed by a discussion of distortion in Section 2.8. We discuss two kinds of distortion measures: permutation invariant and additive. Additive distortion measure is a special case of permutation invariant distortion measure. Sections 2.9 defines what it means for a channel to be capable of universally communicating a random source to within a certain distortion level. Channels which communicate a random source to within a certain distortion level are defined with the probability of excess distortion criterion. Section 2.10 defines source codes which code (compress) a source to within a certain distortion level. Two criteria are used when defining source codes which compress a source to within a certain distortion level: the expected distortion criterion and the probability of excess distortion criterion.

Section 2.11 defines universal reliable communication over a partially known channel.

This is followed by a small discussion of why we use the probability of excess distortion criterion instead of the expected distortion criterion in Section 2.12.

Section 2.13 discusses the important past literature on the problem of point-to-point communication with fidelity criterion.

Then, we come to the results:

Section 2.14 discusses why universal source-channel separation in the point-to-point setting holds in the point-to-point case. We use what we call the uniform X source for this discussion. The uniform X source consists of sequences with type precisely p_X, and this helps avoid a lot of ϵs and δs in the proofs. This discussion holds for permutation invariant distortion measures. We assume a technical condition on the rate-distortion function, and with this assumption, we have a rigorous proof of universal source-channel separation theorem for communication with a fidelity criterion in order to communicate the uniform X source when the distortion metric is permutation invariant. *This section is the most important section of the whole thesis and is the main idea why separation/optimality of digital communication holds for* *universal communication with a fidelity criterion.*

This is followed, in Section 2.15, by a rigorous proof of the universal source-channel separation theorem for rate-distortion when the source is i.i.d. and the distortion measure is additive.

The results are followed by various discussions:

Section 2.16 discusses the high-level idea for proving universal source-channel separation rigorously for permutation invariant distortion measures .

Throughout, we have made the assumption that random-coding is permitted. This assumption is crucial Section 2.17 discusses why this assumption is crucial.

Throughout, we have assumed that the source evolves in discrete time. Section 2.18 discusses high-level ideas for generalization to continuous time sources.

Section 2.19 comments on the assumptions described in Section 2.2, which was discussed above in brief.

Finally, we recapitulate this chapter in Section 2.20.

2.2 Our set up: Information theoretic, and various assumptions made

We will use an information-theoretic set up. In particular, we will assume that

- *Delays do not matter*: That is, it is okay if the source is reproduced with arbitrarily large (but finite) delay. This assumption makes sense in certain cases, but not in others. For example, it makes more sense when sending a text message compared to real-time voice communication over a cellphone. In any case, it is a good assumption from the point of view of getting insight into the nature of communication architectures. *This assumption is crucial.* The cruciality of this assumption is in the sense that if this assumption is not made, in fact, source-channel separation based architectures are not optimal

- *The source can be modeled as a random process, in particular, a stationary ergodic random process:* This is the usual assumption made in communications theory, that the source can be modeled as a random process. Also, it is the usual assumption in communications theory, that the source is stationary, ergodic.

  Without these assumptions, it is very difficult to prove any rigorous results. Comments are made in Subsection 2.4.3, on this assumption

- *The channel can be modeled as a (partially known) transition probability:* In information theory, channels are modeled as transition probabilities.

  We will assume that the channel is only partially known, and thus, we will model the channel as coming from a set of transition probabilities. We will assume that this modeling makes sense. The motivation behind this modeling is that real channels like wireless and internet are time-varying, and their action is not entirely known: at any point of time, the action of the wireless channel is not entirely known as a transition probability, and the exact internet architecture and its behavior is not entirely known as a transition probability. If the channel were adversarial, it can usually be modeled as a partially known channel. Comments are made in Section 2.5.4, on modeling a channel as a known transition probability or a partially known transition probability. The more usual language of saying that the channel is modeled as a transition probability is that we will be solving the universal/compound problem.

In most of the information theory literature, further assumptions on the behavior of the channel, for example, memorylessness, Markoff nature or some assumptions on the memory of the channel (for example, indecomposability in the sense of Gallager [Gal68]) are needed. We will not require any such assumptions, and our model will be a very general channel model a la Verdu-Han [VH94]

- *The source can incur distortion, and the distortion can be modeled as a distortion metric:* We will allow the source to incur distortion, and we will work in the framework of rate-distortion theory as developed by Shannon [Sha59]. Our treatment will be different in the sense that we will use the probability of excess distortion criterion instead of the average distortion criterion. We will assume that the distortion can be modeled as a distortion metric, and this is the usual assumption made in communications theory

The above assumptions are the basis on which information theory is built. We would require further assumptions, and they are stated below

The following crucial assumptions are made concerning the nature of the distortion metric and the distortion criterion, and on random-coding:

- *The distortion metric can be modeled as a permutation invariant distortion metric, and the distortion criterion is the "probability of excess distortion" criterion:* A permutation invariant distortion function is the following: the distortion between the source input signal and its reproduction does not change under the *same* rearrangement of the source and its reproduction. We require the distortion metric to be permutation invariant in the sense that our results are not true if this assumption is not made. Also, we use, what we call the "probability of excess distortion criterion (2.29), the same as the one used in the book of Csiszar and Korner in [CK97], instead of the expected distortion criterion (2.28) which was used by Shannon in [Sha59]. Again, our results are not directly true if we use the expected distortion criterion.
- *Random-coding is permitted:* We assume that random-coding is permitted. This assumption is crucial in the sense that if random-coding is not permitted, the universal source-channel separation theorem for rate-distortion that we prove is not true. In Shannon's random-coding argument, random-coding is a proof technique: assuming that a random code exists, a deterministic code exists. For us, random-coding is *not* just a proof technique. It is necessary. This is because we assume that the channel is only partially known

unlike Shannon who assumed that the channel is fully known. This is commented on further, after we have proved our results, in Section 2.17.

We make the following assumption concerning the statistics of the source. We conjecture that we do not require this assumption, but we are not sure.

- *The source statistics are known:* That is, the source distribution is known. As stated in the previous line, we conjecture that we do not require this assumption. This is commented on further in Section 2.19.

The remaining assumptions are made to prove results rigorously and to avoid mathematical complications, and we are quite sure that modulo some technical assumptions, they can be removed.

We will make the following assumptions on the cardinality of the source and source reproduction alphabet:

- *The source alphabet and source reproduction alphabet are finite*: (in fact, to make any kind of physical sense, it is enough to assume that source alphabet is finite: it does not make sense to have the reproduction alphabet cardinality larger than the source alphabet cardinality): this assumption is needed to prove rigorous results. It not crucial to our work, and the results can be generalized to many sources with infinite alphabet size.

We make also make the following assumption on the random process corresponding to the source:

- *The source that needs to be communicated over the channel is i.i.d.*: this assumption is made only to prove the results rigorously and for simplicity of presentation. We believe (in fact, we are sure), that the results will generalize to many stationary ergodic sources.

We make the following assumption concerning the distortion measure:

- *The distortion measure is additive*: this assumption is made to prove rigorous results. We will hint at how to generalize the results to permutation invariant distortion functions. Permutation invariant distortion functions and additive distortion functions are defined in Section 2.8.

We will make the following assumptions on the time evolution of the source and the channel:

- *The channel evolves discretely in time*: this assumption is made for simplicity of presentation. Continuous time stochastic models for channels involve a lot of technicalities. As will be clear, the results generalize without any change in proofs, even when the channel evolves in time continuously.

- *The source evolves discretely in time*: this assumption is made for simplicity of presentation. Continuous time models for sources involve a lot of technicalities. We will point out ideas for generalizing to sources which evolve continuously in time.
- *The source and the channel evolve on the same time scale:* We will assume that the source and channel evolve on the same time scale. In particular, we will assume that the source and the channel evolve at every integer time. In practice, source can be evolving faster than the channel or vice-versa. For example, when sending a text message on a cell-phone, the source is in fact, evolving in time discretely, whereas the wireless medium evolves in time continuously. We make this assumption for simplicity of presentation. Our results can be generalized to the case when the source and channel evolve on different time scales.

The assumptions which are crucial are:

- Delays do not matter
- The source can be modeled as a stationary ergodic random process and the channel can be modeled as a transition probability
- Allowed distortion between the source and its reproduction can be modeled as a permutation invariant distortion metric
- The source statistics are known,

and an assumption which we have made and is crucial to proving our results, but which we believe can be removed is

- Source statistics are known

The rest of the assumptions are made in order to prove rigorous results and to avoid mathematical complications. Of course, universal source-channel separation will not hold for arbitrary stationary ergodic sources and arbitrary permutation invariant distortion metrics; however, we believe that it should hold for a wide variety of stationary ergodic sources and permutation invariant distortion metrics. These assumptions are commented on in Section 2.19 after we have have provided the the rigorous proof of the universal source-channel separation for rate-distortion.

2.3 The superscript notation

Superscript n will denote a quantity with or related to sequence length (or block length) n. For example, For example, x^n will denote a sequence of length n. Y^n denotes a random variable on the set $\mathcal{Y}^n$. Note that Y^n need not be i.i.d.: the superscript n refers to the block length being n. k^n and c^n will denote n-length channel transition probabilities. e^n will denote an encoder which encodes n length sequences and f^n will denote a decoder which decodes n length sequences.

We will denote sequences of these quantities for various block lengths with $< >$. For example, sequences $< x^n >_1^\infty$, $< y^n >_1^\infty$, $< k^n >_1^\infty$, $< c^n >_1^\infty$. We will denote these infinite sequences by their single letters: for example, $x =< x^n >_1^\infty$, $y =< y^n >_1^\infty$, $k =< k^n >_1^\infty$, $c =< c^n >_1^\infty$.

In general, the superscript n just denotes block length n and is not supposed to indicate a cartesian product or nesting of sorts. For example, Y^n need not be i.i.d., and in general, there might be no relation whatsoever between Y^n and Y^{n+1}: it need not even be the case that the first n components of Y^{n+1} are the same as Y^n. In other words, there might be no nesting.

In certain cases, there will be nesting. For example, when we discuss physical channels, k^n, the transition probability corresponding to the channel for n length sequences will be the first n components of k^{n+1}. However, this will *not* be the case when we discuss abstract channels: c^n might be unrelated to c^{n+1}. Similarly, for encoders, e^n, the encoder used to encode n-length sequences might be completely unrelated to e^{n+1}, the encoder used to encoder $n+1$ length sequences.

In certain cases, superscript n will indeed denote a cartesian product of sorts. For example, X^n will be reserved for an i.i.d. X sequence of length n. For a set $\mathcal{A}$, $\mathcal{A}^n$ will many denote the cartesian product of $\mathcal{A}$ with itself, n times. There are other cases when this will not be true: we will discuss sets $\mathcal{U}^n$ consisting of sequences of a particular type: in that case, there need not be any cartesian product relation between $\mathcal{U}^n$ for various n.

The i^{th} component of x^n will be denoted by $x^n(i)$. In particular, $x^n(n)$ will denote the value of the sequence x^n at time n. Usual literature uses the notation x_n for $x^n(n)$. The reason why we do not want to use this notation is because this notation makes sense only when x^n is the first n components of x^{n+1}. In our set-up, nesting will happen in certain situations and not in others. For example, output sequences y^n will not be nested, in which case, $y^{n+1} \neq (y^n, y^{n+1}(n+1))$, and the notation of y_i for the i^{th} component will not make sense because it would be unclear, we are talking about the i^{th} component of y^n for which block length n. For that reason, we use the notation $y^n(i)$. The same discussion holds for abstract channels, encoders and decoders which are not nested in general. There are certain cases when quantities will be nested. For example, when we consider the model of physical channel $k =< k^n >_1^\infty$, the k^n will be nested, that is, $k^{n+1} = (k^n, k^{n+1}(n+1))$. In this case, we will also denote $k^n(i)$ (which is the same irrespective of the block length n, that is, $k^n(i) = k^{n'}(i) \forall n, n'$) by k_i: the notations k_i and k_i^n will be used interchangeably in such situations.

We would have occasion to require the part of a sequence corresponding to block length n. $a^n(p..q)$ will denote $(a^n(p), a^n(p+1), \ldots, a^n(q))$.

There will be one case when we will not use the superscript notation. That will be when we are dealing with real numbers related to certain block lengths. Then, we will use sub-scripts. For example, ω_n will denote a real number related to block length n and correspondingly, the sequence $< \omega_n >_1^\infty$. We will not use the superscript notation in this case to prevent the possibility of confusion of ω^n with the n^{th} power of ω.

2.4 Source and source reproduction

In this section, we describe our abstraction of a source and source reproduction. The problem of point-to-point communication is to communicate a source from a sender to a receiver over a possibly noisy channel with some guarantee.

Sources will be modeled as random processes.

First, we start with some notation.

2.4.1 Some notation

Notation 2.1 (Source space). The source alphabet is $\mathscr{X}$. We assume that $\mathscr{X}$ is a finite set. $\mathscr{X}^n$ denotes the cartesian product of $\mathscr{X}$ with itself n times. An element of $\mathscr{X}$ is x. An element of $\mathscr{X}^n$ is denoted by x^n.

Notation 2.2 (Source Reproduction space). The source reproduction alphabet is $\mathscr{Y}$. We assume that $\mathscr{Y}$ is a finite set. $\mathscr{Y}^n$ denotes the cartesian product of $\mathscr{Y}$ with itself n times. An element of $\mathscr{Y}$ is y. An element of $\mathscr{Y}^n$ is denoted by y^n.

Note 2.1. The source reproduction space is not the same as the source space for the purpose of abstraction, and also, for the purpose that the source reproduction need not always be a perfect replica of the source. The latter is the case, for example, with voice communication where the received voice need not be the same as the transmitted voice for the recipient to make out, what the speaker said.

2.4.2 i.i.d. X source

We will assume that sources evolve discretely in time.

We will model sources as random processes. For simplicity, we will use the i.i.d. X source.

Notation 2.3 (i.i.d. X source). X is a random variable on $\mathscr{X}$. p_X is the corresponding probability distribution. X^n denotes the i.i.d. X source of block length n. X^n denotes i.i.d. X sequence of length n. $< X^n >_1^\infty$ is the i.i.d. X source.

2.4.3 Discussion: are "real" sources really stationary ergodic

We stated in the previous section that our results stated in the further sections and chapters can be generalized to stationary ergodic sources. The question arises: can practical sources be modeled as stationary ergodic sources.

The answer is, no.

For example, the sources related to language, for example, written text in some language, or spoken language, are not stationary ergodic. The use of stationary ergodic sources is made only to understand the problem of communication and get some hints into the nature of things, and hopefully, design systems.

Exact models of real world are difficult to make, and even if made, it is difficult to come up with any theory about them.

See, for example, [Galb], for a discussion.

2.5 Physical channels

In this section, we describe the physical channel model that we use.

The kind of model we are interested in is the following: Consider the example of wireless channel. The exact behavior of the wireless medium is unknown, even though we might have some knowledge. One criticism to this argument is the following: the wireless channel changes behavior over time scales which are large compared to the block lengths used in wireless communications and thus, over each block length, the wireless channel is essentially fixed. This is true, but what we are saying is that this fixed channel over a particular block length is not known, and this is what we want to model.

We will abstract a channel as a transition probability. This will be called a known channel. As we would like to model channels whose action is not entirely known. These channels are the partially known channels and will be modeled as coming from a set of transition probabilities. Before we state these definitions, we state some notation.

Time will be assumed to evolve discretely.

2.5.1 Some notation

Notation 2.4 (Channel input space). The channel input space is the set is $\mathscr{I}$. We assume that $\mathscr{I}$ is a finite set. $\mathscr{I}^n$ denotes the cartesian product of $\mathscr{I}$ with itself n times. An element of $\mathscr{I}$ is ι. An element of $\mathscr{I}^n$ is denoted by ι^n. We do not denote an element of $\mathscr{I}^n$ by i^n because i is used for indexing.

Notation 2.5 (Channel output space). The channel output space is the set is $\mathscr{O}$. We assume that $\mathscr{O}$ is a finite set. $\mathscr{O}^n$ denotes the cartesian product of $\mathscr{O}$ with itself n times. An element of $\mathcal{O}$ is o. An element of $\mathcal{O}^n$ is denoted by o^n.

2.5.2 A fully known physical channel, k

A fully known physical channel is one whose action as a transition probability is known. This is mathematically abstracted as follows:

Definition 2.1 (A fully known physical channel). We want to use a very general channel model: the output of the channel at time i can depend on the inputs of the channel up to and including time $i-1$ and the outputs of the channel up to and including time $i-1$. Let the block length be n. The channel transition probability at time i is denoted by k_i:

$$k_i(o^n(i)\,|\iota^n(1..i-1), o^n(1..i-1)), \text{ if } i \leq n$$
$$k_i(o^n(i)\,|\iota^n, o^n(1..i-1)), \text{ if } i > n \tag{2.1}$$

is the probability that the channel output at time i is $o^n(i)$ given that the channel inputs and outputs up to and including time $i-1$ are $\iota^n(1..i-1)$ (or ι^n if $i > n$) and $o^{n-1}(1..i-1)$ respectively.

Note that k_i is independent of the block-length n. When the block-length is n, the channel evolves until some time $t_n \geq n$. The channel is $k^n = (k_1, k_2, \ldots, k_{t_n})$. For simplicity, we assume that $t_n = n$. Thus, when the block-length is n, the channel $k^n = (k_1, k_2, \ldots, k_n)$. Assumption that $t_n = n$ can be made without loss of generality: it is related to the issue of time scales, discussed in Note 2.6.4.

In the superscript notation as defined earlier, k_n is also denoted as $k^n(n)$. Note that $k^{n+1} = (k^n, k^{n+1}(n+1))$, that is, this model of a physical channel is nested, as it should be. It is for this reason, as stated in Section 2.3 that we can use the notation k_n. The full channel evolution in time, as stated in Section 2.3, is denoted by $k =< k^n >_1^\infty$.

Note 2.2 (Is there no dependence on the initial channel state?). In general, there is a dependence of k_i on the initial channel state. However, we do not show this dependence. This is because, as we shall see, the model of the channel that we will use is a partially known channel, in that, the channel can belong to a family. For that reason, we will treat the same channel with different initial states as different channels and assume that all these channels belong to the family which make up the partially known channel. A partially known physical channel is the subject of the next subsection.

Note, also, that even though the model of a physical channel is nested, the channel inputs, and hence also, the channel outputs, may not be nested.

Note 2.3 (A very general channel model). This is a very general model of a "physical" channel as a transition probability evolving in discrete time. It is in fact, the most general possible model of a "physical channel" evolving in discrete time other than the fact that we have not made the dependence on the initial channel state for the reason described in Note 2.2. In particular, we do not impose any memorylessness or Markoff assumptions on the channel.

■ 2.5.3 A partially known physical channel

Note that since the initial state s may not be entirely known, in general, and also, since the transition probability k^n might not be entirely known, we want to model a channel as belonging to a set of transition probabilities. This motivates the definition of a partially known physical channel:

Definition 2.2 (Partially known physical channel). A physical channel is said to be partially known if it belongs to a set of transition probabilities $\mathscr{A}$.

Notation 2.6 (Notation for a partially known channel). A partially known channel k which comes from a set of transition probabilities $\mathscr{A}$ is denoted by $k \in \mathscr{A}$, and it is will be referred to as, "partially known channel $k \in \mathscr{A}$."

Note 2.4 (Partially known channels and compound channels). In the information theory literature, partially known channels are referred to as compound channels. See for example [CK97] for a discussion of memoryless compound channels. Our model is different from [CK97] in the sense that our channels are not memoryless.

■ 2.5.4 Discussion: Can "real" channels be modeled as a transition probability or a family of transition probabilities?

We have modeled channels which are partially known as coming from a set of transition probabilities.

Question arises: is modeling "real" channels or media as coming from a sets of transition probabilities, a good model? For example, can wireless medium or the internet be modeled as a coming from a set of transition probabilities.

The answer is that we do not know.

Some people like to model channels adversarially. Adversarial models can usually be modeled as a set of transition probabilities. Again, it is unclear if this is the right thing to do.

The important thing, from our perspective will be, as we shall see, is that the channel model will be quite irrelevant: we want to prove the optimality of source-channel separation. From the nature of the proof, it will be clear that we will convert any given architecture into a digital architecture in order to prove the optimality of a digital scheme, and thus, the channel model will be quite irrelevant.

The optimality of digital communication is probably much more fundamental than the underlying channel model. This will be commented on further in Subsection 2.14.10, after a proof of universal source-channel separation for rate-distortion for the uniform X source.

Of course, it would fail to hold for very pathological channel models. And of course, it can still happen that real channels belong to the family where digital communication is not optimal. We leave these questions unanswered.

Multi-user media will be modeled and results of this chapter, generalized to the multiuser setting, in Chapter 3

2.5.5 The problem of communication over a partially known channel

The problem of communication over a fully known channel is to construct an encoder and a decoder in order to communicate the source over the channel with a particular guarantee.

The problem of communication over a partially known channel $k \in \mathscr{A}$ is to construct an encoder and a decoder such that the source is communicated over the channel withsome guarantee irrespective of the particular $k \in \mathscr{A}$. The encoder and the decoder should be independent of the particular channel.

The guarantee that we will use is communication to within a distortion level.

The question we want to answer is: given a partially known channel $k \in \mathscr{A}$, are source-channel separation based architectures optimal to communicate i.i.d. X source over this partially known channel to within a distortion level D. In other words, if there exists some general architecture to communicate i.i.d. X source over the partially known channel to within a distortion D, does there exist a digital scheme which consumes the same or lesser system resources and accomplishes the same?

We discussed analog and digital point-to-point communication architectures in Section 1.3. In Sections 2.6 and 2.7, we make these rigorous.

2.6 An analog point-to-point communication system

In this section, we describe rigorously, the general point-to-point communication system described on a high-level in Section 1.3. A general point-to-point communication system consists of an encoder, a channel and a decoder. The encoder encodes the source and the encoder input is communicated over the channel. The decoder reconstructs the source from the channel output. Hopefully, end-to-end, the source has been communicated to within the required guarantee. The guarantee that we will use is communication to within a distortion level.

In this section, we define the action of the encoder, channel and decoder rigorously.

The precise definition of communication to within a distortion level are left for later sections.

As stated before, throughout, we will assume that time evolves discretely, both for the source and the channel, and that, the source, source reconstruction and channel input and output alphabet are all finite.

2.6.1 Discussion: Is this not already digital?

We have assumed that the source space, the channel input space, the channel output space and the source reproduction space are all finite. Thus, from the definition of Chapter 1, this architecture is already digital. The question comes, why are we calling this an analog communication system. The answer is that we have assumed the spaces to be finite only for the sake of avoiding mathematical technicalities. All definitions made so far and that will be made for the rest of this section can also be made with infinite sets. Our results can be generalized to the setting where the input and output spaces are infinite. The fact that we use a digital interface consisting of a finite alphabet (two) will remain unchanged even when the rest of the alphabets are infinite.

2.6.2 Encoder and decoder

The encoder takes input from the source. Thus, $\mathscr{X}$ is also the encoder input space. The encoder produces an output into the channel. Thus, $\mathscr{I}$ is also the encoder output space. The output of the encoder is transmitted over the channel and the channel output is an input to the decoder. Thus, $\mathscr{O}$ is also the decoder input space. The decoder reconstructs the source. Thus, $\mathscr{Y}$ is also the decoder output space. Definitions 2.5 and 2.6 state these definitions precisely.

Definition 2.3 (Random codes and random coding). Random codes are codes where the encoder can belong to a family of deterministic codes and the decoder has access to the particular realization of the deterministic encoder that happened. The performance of the code is judged by averaging over the family of deterministic codes under a certain probability distribution of the particular deterministic code. Random codes can be generated by using common randomness which is defined next.

Note 2.5. i.i.d. random codebook generation, as done by Shannon in his random-coding argument, is a special case of above random codes.

Definition 2.4 (Common randomness). The encoder and decoder have access to common randomness. This means that they have access to a continuous valued random variable independent of all other random variables in the system. Common randomness is used to generate random codes. The common randomness input is denoted by r.

Definition 2.5 (Encoder). When the block length is n, the encoder acts as e^n:

$$e^n(i^n \mid x^n, r) \tag{2.2}$$

is the probability that the encoder output is i^n given the encoder input is x^n and the common randomness input is r.

Note that the encoder is *not* necessarily nested. Also, note that in the encoder model, we assume that the input x^n is available at the beginning of time. This assumption can be made because we allow arbitrary delays and the input can be thought of as buffered.

Definition 2.6 (Decoder). When the block length is n, the decoder acts as f^n:

$$f^n(y^n \mid o^n, r) \tag{2.3}$$

is the probability that the decoder output is y^n given the decoder input is o^n and the common randomness input is r.

Note 2.6 (How can common randomness be used to generate random codes). We defined random codes as codes where the encoder belongs to a family of deterministic encoders and the decoder has access to the particular realization of the deterministic encoder that happened. Note that such codes can be generated using common randomness. For each r, assume that the encoder $e^n(\cdot \mid x^n, r)$ is deterministic in the sense that $e^n(i^n \mid x^n, r)$ is 1 for some particular i^n and zero otherwise. This happens for all x^n and i^n might depend on x^n. Similarly, for each r assume that the decoder $f^n(\cdot \mid o^n, r)$ is deterministic in the sense that $f^n(y^n \mid o^n, r)$ is 1 for some particular y^n and zero otherwise. This happens for all o^n and y^n might depend on o^n. Such an encoder-decoder pair e^n, f^n is a deterministic encoder-decoder pair. r can vary, and is available both at the encoder and the decoder. Thus, we have generated codes when the encoder can belong to a family of deterministic encoders and the deterministic decoder can depend on the particular choice of the deterministic encoder. Shannon used random-coding arguments by generating codes i.i.d. from a particular distribution. As we stated before, random codes of this variety are a special case of random-codes as defined by us. *There is a big difference however: for Shannon, random-coding was a proof technique, whereas for us, random-coding is not just a proof technique.*

Note 2.7 (A note on real-time evolution and the corresponding mathematical abstraction). Note that the encoder and decoder are *not* necessarily nested. Also, note that in the encoder model, we assume that the input x^n is available at the beginning of time. This assumption can be made because we allow arbitrary delays and the input can be thought of as buffered. Also, note that in the decoder model, we assume that the channel output o^n is available at the beginning of time. This assumption can be made because we allow arbitrary delays and the decoder output can be buffered before making an estimate of y^n.

The real time-evolution will probably happen as follows:

- Time 1 to n: x^n arrives
- Time $n+1$ to $2n$: The encoder produces i^n
- Time $n+1$ to $2n$: The channel produces output o^n
- Time $2n+1$ to $3n$: The decoder produces the estimate y^n

The mathematical model that we have abstracts this real-time evolution.

2.6.3 The composition of the encoder, channel and decoder: the point-to-point communication system

Definition 2.7 (Composition of encoder, channel and decoder). The composition of the encoder, channel and decoder is the point-to-point communication system. When the block length is n, this composition is a transition probability $e^n \circ k \circ f^n$, which we denote by c^n.

$$c^n(y^n \mid x^n) \tag{2.4}$$

is the probability that the composite channel output is y^n given that the input is x^n.

Notation 2.7. Since we are interested in block lengths as they become larger and larger, the composition of the encoder, channel and decoder will be denoted by $< e^n \circ k^n \circ f^n >_1^\infty$.

Since we model channels as belonging to a set of transition probabilities $\mathscr{A}$, we would like to think of the point-to-point communication system as

$$\{< e^n \circ k^n \circ f^n >_1^\infty \mid < k^n >_1^\infty \in \mathscr{A}\} \tag{2.5}$$

2.6.4 Time scales

Note that the encoder has been defined as a transition probability

$$e^n(i^n \mid x^n, r) \tag{2.6}$$

In effect, this means that the source and the channel are evolving on the same time scale. This is because, until time n, the number of source inputs is n and the number of encoder outputs which is the same as the channel inputs is also n: thus, the rate of source input is the same as the rate of channel input. Another way of saying this is that the source and the channel are evolving on the same time scale.

This is an assumption for modeling convenience. We can state all definitions with source and channel evolving on different time scales. It only leads to notational difficulties. For this reason, throughout, we will assume that the source and the channel evolve on the same time scale. Our results can be generalized to the case when the time-scale of source and channel evolution are different.

2.6.5 The view of a point-to-point communication system as an abstract channel, c

Definition 2.8 (Composite or abstract channel model, $< c^n >_1^\infty$). The point-to-point communication system can thus be viewed as a composite transition probability, $< c^n >_1^\infty$. When the block length is n, the point-to-point system acts as c^n:

$$c^n(y^n \mid x^n) \tag{2.7}$$

is the probability that the composite system output *up to* time n is y^n given that the system inputs up to time n are x^n. $c =< c^n >_1^\infty$ is the composite or the abstract channel.

Note 2.8 (Why do we call this channel model, abstract?). Note that the abstract channel model is ***noncausal*** and ***nonnested.*** This is the reason for calling this model abstract. Of course, we also call it the composite channel model because it is the composition of the encoder, the channel, and the decoder.

This is the abstract channel model we will use: a channel should be thought of as a sequence of transition probabilities $< c^n >_1^\infty$ where, when the block length is n, the channel acts as c^n, and $c^n(y^n|x^n, s)$ is the probability that the channel output is y^n given that the channel input is x^n and the initial state is s. We will denote the composite channel by $c =< c^n >_1^\infty =< e^n \circ k^n \circ f^n >_1^\infty$.

Similarly, when the channel belongs to a family of transition probabilities, we will think of the abstract communication system as

$$\mathscr{C}_{\mathscr{A}} \triangleq \{< e^n \circ k^n \circ f^n >_1^\infty \mid < k^n >_1^\infty \in \mathscr{A}\} \tag{2.8}$$

Note 2.9. Note that the input space of the abstract channel c is $\mathscr{X}$ instead of $\mathscr{I}$ and the output space is $\mathscr{Y}$ instead of $\mathscr{O}$

■ 2.6.6 Communication of a random source over a point-to-point communication system

Let the block length be n. The steps of communication are the following

1. The input to the encoder e^n is a realization x^n of the i.i.d. X source X^n (for this description, the source need not be i.i.d., it might be some general random variable X^n). The encoder produces the source encoding which is a realization ι^n of the random variable I^n. i^n is the input to the channel
2. i^n is communicated over the channel k^n and the channel output is a realization o^n of the random variable O^n
3. The decoder d^n reconstructs the source from o^n. The decoder output is a realization y^n of the random variable Y^n.

For block length n, this results in the joint random variable X^nY^n on the source-source reproduction space $\mathscr{X}^n \times \mathscr{Y}^n$ with the corresponding probability distribution $p_{X^nY^n}$. See Figure 2.1.

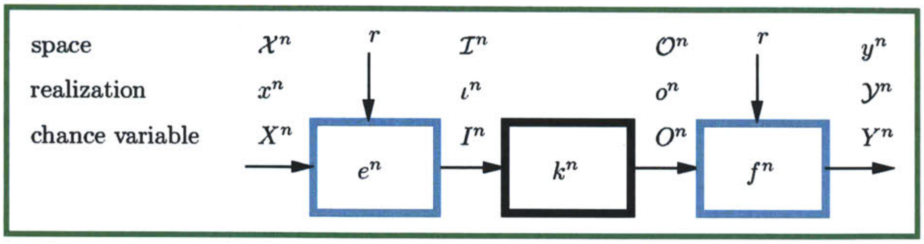


Figure 2.1. The action of a point-to-point communication system

■ 2.6.7 Resource consumption in the point-to-point communication system

In this subsection, we discuss the resource consumption in this point-to-point communication system. We want to think of system resources like energy and bandwidth. Note that our set-up is abstract, in that the sets $\mathscr{X}, \mathscr{Y}, \mathscr{I}$, and $\mathscr{O}$ are finite sets. Thus, defining energy consumption or bandwidth consumption physically is not possible. What we will instead state, is a sufficient condition for two system resources to consume the same resources, and see, why this abstraction makes sense.

A sufficient condition for two systems to consume the same system resources

We want to think of point-to-point communication systems in terms of the previous subsection (Subsection 2.6.6). Let s_1 and s_2 denote two communication systems with the same spaces $\mathscr{X}, \mathscr{I}, \mathscr{O}$ and $\mathscr{Y}$, which are used to communicate a random source over a channel k. The encoder-decoder for the system s_1 are $< e_1^n, f_1^n >_1^\infty$ and the encoder and decoder for system s_2 are $< e_2^n, f_2^n >_1^\infty$. In abstract terms, the two systems are

$$s_1 = < e_1^n \circ k \circ f_1^n >_1^\infty \tag{2.9}$$

$$s_2 = < e_2^n \circ k \circ f_2^n >_1^\infty \tag{2.10}$$

The source $< X^n >_1^\infty$ which needs to be communicated over the two systems is the same for both systems. The random-variables defined in the previous subsection for the system s_1 are X^n, I_1^n, O_1^n and Y_1^n. The random-variables defined in the previous subsection for the system s_2 are X^n, I_2^n, O_2^n and Y_2^n.

A sufficient condition for the systems s_1 and s_2 to consume the system resources is: $\forall n$, I_1^n has the same distribution as I_2^n as random variables (note that I_1^n and I_2^n are n-length random variables).

Note that the above is a sufficient condition for two systems to consume the same system resources; it is not necessary. Even if the distribution of I_1^n and I_2^n is not the same, it does not directly imply that the two systems consume different system resources.

In the next subsection, we argue, why this abstract sufficiency of the equality of consumption of system resources makes sense.

Why does this sufficient condition for two systems to consume the same system resources make sense?

The two resources which we are concerned with are energy and bandwidth. In general, these definitions make sense only for systems evolving in continuous time.

Consider a system evolving in continuous time. The input to the system at time t is $X(t)$, the output of the encoder at time t is $I(t)$ is the input to the channel, the output of the channel at time t is $O(t)$ is the input to the decoder and the output of the decoder at time t is $Y(t)$ which is the source reproduction.

Energy is consumed in the point-to-point communication system in two ways:

1. Energy consumed in the processing and computations in the encoder and the decoder.
2. Energy transmitted into the channel, that is, the energy in the signal.

In many practical scenarios, the energy consumed in the processing in the encoder and the decoder is much less than the energy consumed in the signal transmitted into the channel. For example, in the wireless example, most of the energy is consumed in the signal transmitted by the cellphone into the air, and not in voice processing.

The instantaneous power input into the system at time t is $\iota^2(t)$ where $\iota(t)$ is a realization of $I(t)$. The expected instantaneous power consumption is $E[I^2(t)]$. Clearly, the power consumption depends only on the distribution of the random process I. Energy consumption is an average over time, of the power consumption, and thus, again, depends only on the distribution of I.

Thus, assuming that the energy consumed in the processing and computations in the encoder and the decoder can be neglected, the energy consumed in the system depends only on the distribution of the channel input process I.

Now, let us consider bandwidth consumption. Let one particular realization to the channel over time t be $\iota(t)$. The bandwidth requirement is the support of the fourier transform of $\iota(t)$. Again, if two systems with inputs $I_1(t)$ and I_2 to the channel have the same distribution for the processes I_1 and I_2, the bandwidth requirements on the channel will be the same.

Thus, energy and bandwidth consumption are a function, only of the channel input distribution. For this reason, we have made the abstract condition under which two systems consume the same energy. Of course, this is a sufficient condition, and not a necessary condition.

As we have stated before, we believe our results can be generalized to systems evolving in continuous time; we are restricting to discrete spaces to avoid mathematical technicalities.

When generalizing our results to continuous time system evolution, the only fact that would be needed about the consumption of system resources is the above sufficient condition for the equality of consumption of system resources, and for that reason, we just restrict to this rather general condition. We will use this condition when proving the optimality of digital communication for communication with a fidelity criterion when we construct a digital system which consumes the same system resources as an analog system by showing that the channel input distribution for both the analog and the digital system is the same.

Consumption of "lesser" system resources

Suppose a communication system needs to be built to meet certain communication guarantees. Suppose this can be done with certain consumption of system resources. Then, we will say, abstractly, that the same guarantee can be met by consumption of the same or "lesser" system resources. This, again, is an abstract definition because we have not defined the consumption of a system resource; we have only stated a sufficient condition for the equality of consumption of the same system resources by two systems. However, the reason for this abstract use of the word "lesser" is done because for physical systems where resource consumption can in fact be defined, this would be the right usage of "the same guarantee can be met by consumption of the same or lesser system resources."

2.6.8 The point-to-point communication problem

The systems problem of point-to-point communication is to construct encoder-decoder pair $< e^n, f^n >_1^\infty$ which satisfy certain constraints on resource consumption and such that the source is communicated over the partially known channel (the channel belongs to a set $\mathscr{A}$) with a certain guarantee.

The guarantee that we will use is communication to within a distortion level. The precise definition of communication with distortion is defined in Section 2.9.

The question that we will ask is: can the encoder-decoder $< e^n, f^n >_1^\infty$ be constructed digitally without loss of optimality. First, we define digital architectures rigorously in the next section.

2.7 A point-to-point digital communication system

In a digital point-to-point communication system, the encoder is broken down into source encoder and channel encoder, and the decoder is broken down into channel decoder and the source decoder. This was described, on a high-level, in Section 1.3.

The source encoder - source decoder pair is called the source code.

The channel encoder - channel decoder pair is called the channel code.

The source encoder converts the source into a binary sequence, that is, a sequence of 0s and 1s. The binary sequence is communicated reliably (with a small probability of error) over the channel with the help of the channel encoder and the channel decoder. The source decoder reconstructs the source from the output of the channel decoder. Hopefully, end-to-end, the source has been communicated with the required guarantee. The guarantee that we will use is communication to within a distortion level.

In this section, we define mathematically, the action of source encoder-source decoder pair, and the channel encoder-channel decoder pair. The channel has already been rigorously defined in the previous section.

The precise definitions of communication to within a distortion level and reliable communication are left to later sections.

In this section, we will only talk about things abstractly, and thus, the models will be noncausal. However, behind every noncausal model, there is a causal model.

2.7.1 A rate R binary sequence

When the block length is n, a rate R binary sequence would be a binary sequence of length nR. Physically, this means a sequence of nR 0s and 1s.

The source encoder produces a binary sequence as output. If the source is random, the binary sequence is also random. This is abstracted as follows:

Definition 2.9 (The binary sequence set or the message set). The set of rate R binary sequences evolving for n units of time (block length is n) is the set

$$\mathcal{M}_R^n = \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\} \tag{2.11}$$

Each message $\in \mathcal{M}_R^n$ should be thought of as being associated with a particular binary sequence. A generic element of $\mathcal{M}_R^n$ will be denoted by m^n.

The binary sequence is communicated over the channel, hopefully, reliably, with the help of the channel encoder and the channel decoder. The output of the channel decoder is the reconstructed binary sequence. This gives the binary sequence or the message reproduction set.

Definition 2.10 (The binary sequence reproduction or the message reproduction set). The message reproduction set should be thought of as the message set along with an error message, if the message could not be reconstructed

$$\hat{\mathcal{M}}_R^n = \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\} \cup \{e\} \tag{2.12}$$

e is the error message, and one should think of another possible binary sequence being associated to e. A generic element of $\hat{\mathcal{M}}_R^n$ will be denoted by $\hat{m}^n$

In general, decoding will happen with a delay.

2.7.2 Source code

The source code consists of a source encoder and a source decoder.

When the block length is n, a realization of the source input is $x^n \in \mathcal{X}^n$. This is mapped by the source encoder into m^n, an element of the message set $\mathcal{M}_R^n$. m^n is communicated, hopefully reliably over the channel with the help of the channel encoder and the channel decoder. The output of the channel decoder is $\hat{m}^n$. The source decoder reconstructs the source from the output $\hat{m}^n$ of the channel decoder and the source reconstruction is y^n.

The mechanism by which the source encoder and the source decoder act is abstracted as follows:

Definition 2.11 (Rate R deterministic source code). Rate R deterministic source code is a sequence $s =< s^n >_1^\infty =< e_s^n, f_s^n >_1^\infty$, where e_s^n is a function with domain $\mathcal{X}^n$ and range $\mathcal{M}_R^n$, and f_s^n is a function with domain $\hat{\mathcal{M}}_R^n$ and range $\mathcal{Y}^n$.

We usually think of encoders and decoders as transition probabilities. In the above definition, we have defined e_s^n and f_s^n as functions. This is because deterministic functions can be thought of as transition probabilities.

Note 2.10 (Interpretation of a rate R deterministic source code). It would be helpful to refer to Figure 2.2, except that there is no common randomness input. $e =< e_s^n >_1^\infty$ is the source encoder and $f =< f_s^n >_1^\infty$ is the source decoder in Figure 2.2. When the block length is n, the source encoder is e_s^n and the source decoder is f_s^n. $x^n \in \mathcal{X}^n$ is source coded by e_s^n as $m^n = e_s^n(x^n)$. This message is communicated over the channel with the help of the channel encoder and the channel decoder. The output of the channel decoder is the message reconstruction $\hat{m}^n$ of m^n. In a "good" digital communication system, we would like $\hat{m}^n$ to be equal to m^n with high probability. $\hat{m}^n$ is source decoded as $y^n = f_s^n(\hat{m}^n)$. y^n is the reconstruction of x^n. In a "good" digital communication system, y^n would be with an acceptable distortion of x^n. Note that the sets $\mathcal{M}_R^n$ and $\hat{\mathcal{M}}_R^n$ are the same. However, we have used different notation just to emphasize that $\hat{\mathcal{M}}_R^n$ is the message reconstruction set and $\mathcal{M}_R^n$ is the message set.

We would assume that there is common randomness at the transmitter and the receiver. In other words, the codes can be random. This is made precise as follows.

Definition 2.12 (Rate R random source code). A rate R random source code is a sequence $s =< s^n >_1^\infty =< e_s^n, f_s^n >_1^\infty$. There is a source of randomness which we denote by r. which is available both at the encoder and the decoder. e_s^n is a transition probability

$$e_s^n(m^n | x^n, r) \tag{2.13}$$

is the probability given that the source input is x^n and the common randomness is r.

$$f_s^n(y^n | \hat{m}^n, r) \tag{2.14}$$

is the probability that the source reconstruction is y^n given that the message reconstruction is m^n and the common randomness input is r

Note 2.11. If there is no common randomness input (mathematically this can be thought of as the same source encoder and source decoder being used irrespective of the common randomness input), a random source code reduces to a deterministic source code.

Note 2.12 (Interpretation of a rate R random source code). It would be helpful to refer to Figure 2.2. $e =< e_s^n >_1^\infty$ is the source encoder and $f =< f_s^n >_1^\infty$ is the source decoder in Figure 2.2. When the block length is n, the source encoder is e_s^n and the source decoder is f_s^n. There is a common randomness input at the encoder and the decoder. Recall that the encoder consists of a source encoder and a channel encoder and a decoder consists of a channel decoder and a source decoder. There is a common randomness input at the encoder and the decoder. This means that both the encoder and decoder have access to a common random variable. This random variable is used to generate random codes. $x^n \in \mathcal{X}^n$ is source coded by e_s^n as m^n with probability $e_s^n(m^n|x^n,c)$. This message m^n is communicated over the channel with the help of the channel encoder and the channel decoder. The output of the channel decoder is the message reconstruction $\hat{m}^n$ of m^n. In a "good" digital communication system, we would like $\hat{m}^n$ to be equal to m^n with high probability. $\hat{m}^n$ is source decoded as y^n with probability $f_s^n(y^n|\hat{m}^n,c)$. y^n is the reconstruction of x^n. In a "good" digital communication system, y^n would be within an acceptable distortion of x^n.

Definition 2.13 (Transition probability corresponding to a source code). If there were perfect reproduction through the channel, that is, m^n is always reproduced as $\hat{m}^n$, the final distribution of the source reproduction given the source is given by the transition probability $e_s^n \circ f_s^n$. $e_s^n \circ f_s^n(y^n|x^n)$ denotes the probability that the source reproduction is y^n given that the source input is x^n. $< e_s^n \circ f_s^n >_1^\infty$ is called the transition probability corresponding to the source code s.

Discussion 2.1 (Why define the transition probability corresponding to a source code?). The transition probability corresponding to a source code is defined by taking the composition $< e_s^n \circ f_s^n >_1^\infty$ of the source-encoder $< e_s^n >_1^\infty$ and the source-decoder $< f_s^n >_1^\infty$. However, in a digital communication system, the source-encoder and the source-decoder are not interconnected to each other, directly. The channel encoder, the channel and the channel-decoder exist between the source-encoder and the source-decoder. The question arises: does taking the composition of the source encoder and the source-decoder make physical sense? The answer is that it does make physical sense. This is because, in a digital communication system, the channel encoder, the channel and the channel decoder act in a way so that the input to the channel encoder is communicated reliably, that is, with a very small error, and received at the channel decoder. Thus, the composition of the channel encoder, the channel, and the channel decoder can be thought of as a point-to-point communication sub-system which does almost perfect transmission. For this reason, the composition of the source encoder, the channel encoder, the channel, the channel decoder and the source-decoder will be "close to" the composition of the source-encoder and the source-decoder, and in effect, the source reproduction from of a source after passing through the whole communication system consisting of the source encoder, channel encoder, channel, channel decoder and the source decoder will be the same, with high probability, to the source reproduction, as if the the channel encoder, the channel and the channel decoder did not exist, and the source encoder were directly connected to the source decoder. Further in particular, the distortion incurred, end-to-end (defined rigorously in Section 2.9), by a point-to-point communication system consisting of the source-encoder, channel encoder, channel, channel decoder and source decoder for a particular source will be "close to" the distortion incurred by the system consisting of the composition of the source encoder and the source decoder with the source input.

Discussion 2.2 (Construction of source codes via random-coding arguments, and directly defining the transition probability corresponding to the source code). In usual random-coding arguments for source-coding in the information theory literature, and also, the random-coding arguments for source-coding that we will use, source codes are usually constructed by directly defining the transition probability $< e_s^n \circ f_s^n >_1^\infty$ corresponding to the source code $< e_s^n, f_s^n >_1^\infty$, and not the source-encoder $< e_s^n >_1^\infty$ and the source-decoder $< f_s^n >_1^\infty$, separately. If one is not defining the source encoder $< e_s^n >_1^\infty$ and the source decoder $< f_s^n >_1^\infty$, separately, it might not be clear, what the rate of the source code is. In the random-coding arguments, usually, the encoding is done in the following way: when the block length is n, generate $2^{\lfloor nR \rfloor}$ codewords using a particular distribution. This means that the set on which $e_s^n \circ f_s^n$ puts nonzero probability for any possible source distribution has cardinality $2^{\lfloor nR \rfloor}$. It follows that $e_s^n \circ f_s^n$ factors through a set of cardinality $2^{\lfloor nR \rfloor}$, and thus, has rate R. We will use such a random-coding construction by defining the transition probability $e_s^n \circ f_s^n$ directly for constructing a source code in Sections 2.14 and 2.15.

2.7.3 Channel code

The channel code consists of the channel encoder and the channel decoder.

The source encoder produces a message $m^n \in \mathcal{M}_R^n$ as output which is an input to the channel encoder. The message m^n is encoded by the channel encoder into ι^n, an element of the channel input space. The channel produces an output o^n. The message is reconstructed by the channel decoder from the channel output o^n, and the reconstructed message is $\hat{m}^n$ which is hopefully equal to m^n.

The mechanism by which the channel encoder and the channel decoder act is abstracted as follows:

Definition 2.14 (Channel encoder). On an abstract level, the action of the channel encoder is a transition probability.

$$e_c^n(\iota^n | m^n, r) \tag{2.15}$$

is the probability that the encoder output is ι^n given that the channel input message is m^n and the common randomness is r.

The channel encoder should be thought of as the sequence $< e_c^n >_1^\infty$.

Definition 2.15 (Channel decoder). On an abstract level, the action of the channel decoder is a transition probability.

$$f_c^n(\hat{m}^n|o^n, r) \tag{2.16}$$

is the probability that the output of the channel decoder is $\hat{m}^n$ given that the input is o^n and the common randomness is r.

The channel decoder should be thought of as the sequence $< f_c^n >_1^\infty$.

Note 2.13. Since the source-encoder and the channel encoder are at the same location physically when building a communication system, and similarly, since the channel decoder and the source decoder are at the same location physically when building a communication system, the common randomness input r can be thought to be the same for both the channel code and the source code.

2.7.4 Digital communication system

See Figure 2.2.

Definition 2.16 (Digital encoder). The digital encoder consists of the composition of the source encoder and the channel encoder:

$$< e^n >_1^\infty = < e_s^n \circ e_c^n >_1^\infty \tag{2.17}$$

Definition 2.17 (Digital decoder). The digital decoder consists of the composition of the channel decoder and the source decoder

$$< f^n >_1^\infty = < f_c^n \circ f_s^n >_1^\infty \tag{2.18}$$

Definition 2.18 (Digital point-to-point communication system). The digital communication system is the composition of the digital encoder, channel and the digital decoder:

$$< e^n \circ k^n \circ f^n >_1^\infty = < (e_s^n \circ e_c^n) \circ k^n \circ (f_c^n \circ e_c^n) >_1^\infty \tag{2.19}$$

The channel model that we will use is that of a partially known channel $k \in \mathscr{A}$. Thus, we would like to think of the set of digital communication systems

$$\{< e^n \circ k^n \circ f^n >_1^\infty = < (e_s^n \circ e_c^n) \circ k^n \circ (f_c^n \circ e_c^n) >_1^\infty : k = < k^n >_1^\infty \in \mathscr{A}\} \tag{2.20}$$

2.7.5 Communication of a random source over a point-to-point digital communication system

Let the block length be n. The steps of point-to-point digital communication are the following:

1. The input to the source-encoder e_s^n is a realization x^n of the i.i.d. X source X^n. The output of the source encoder is a realization m^n of the random binary source M_R^n. m^n is input to the channel encoder.
2. The channel encoder encodes the random binary source realization m^n into a sequence ι^n which is a realization of the random variable I^n, and is the input to the channel
3. The channel acts on ι^n and produces the output o^n which is a realization of the random variable O^n
4. The channel decoder decodes o^n into the reconstruction of the random binary source $\hat{m}^n$ which is a realization of the random variable $\hat{M}_R^n$.
5. The source decoder reconstructs the source from $\hat{m}^n$. The source reconstruction is y^n which is a realization of Y^n.

See Figure 2.2.

■ 2.7.6 Resource consumption in a digital point-to-point communication system

A digital point-to-point communication system is a special case of a general point-to-point communication system: the speciality lies in that the encoder and decoder are both digital.

For this reason, and the fact that the sufficient condition that we have stated for equality of consumption of system resources depends only on the channel input, this discussion is the same as for a general point-to-point communication system, see Subsection 2.6.7.

■ 2.7.7 Since all the spaces are finite, is the point-to-point communication system not already digital?

We discussed before that the assumption that all the spaces are finite is made only for technical simplifications. Our results stated in the further sections and chapters will generalize to the case when the source, channel input, channel output and source reconstruction alphabets are infinite.

There is another reason why we have to go through this whole discussion. This is the following:

As described before, the way digital communication systems are constructed is the following: the source is first coded by the source encoder into a binary sequence. This step usually compresses the source to within the desired guarantee of communication. The binary sequence is communicated *reliably over the channel*. This reliably transmitted binary sequence is reconstructed back to the source by the source decoder. Thus, digital architectures, at least

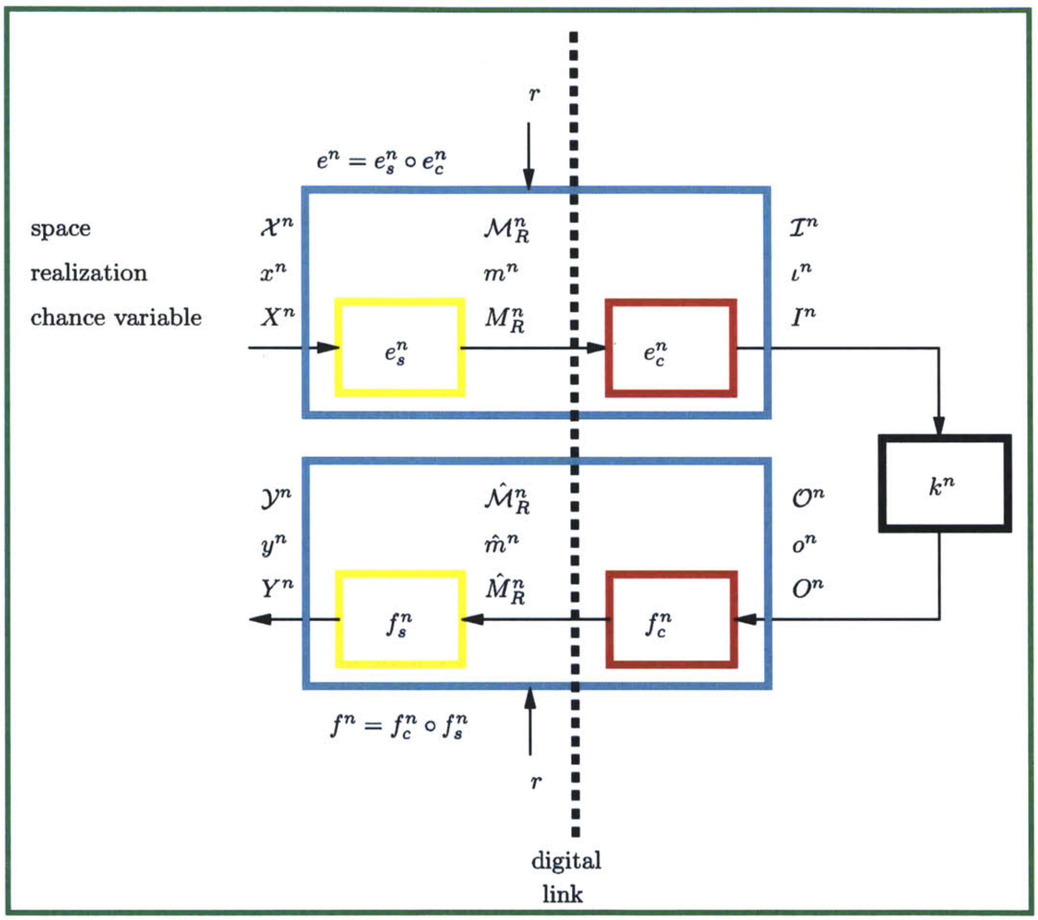


Figure 2.2. Action of a point-to-point digital communication system

the ones used in practice, have the special feature that the channel communicates the binary sequence reliably, and this needs to be described irrespective of whether the alphabets are finite or infinite. This is another reason for this whole description of digital communication systems.

We will take this as the definition of a digital communication system:

1. There is a digital interface (usually binary) between the source and the channel
2. The channel encoder and decoder perform in a way which accomplishes reliable communication over the channel

2.7.8 The point-to-point communication problem

As stated before, the systems problem of point-to-point communication is to construct encoder decoder pair $< e^n, f^n >_1^\infty$ which satisfy certain constraints on system resources, and such that the source is communicated over the partially known channel (the channel $k \in \mathscr{A}$ with a certain guarantee.

The question is: can this be done with digital encoders and decoders which have the same system resource consumption (or, digital has lesser system resource consumption). Advantages of digital architectures have been discussed, to some extent, in Chapter 1.

We will prove the optimality of digital architectures for the guarantee of communication to within a distortion level. The digital architecture that we will construct to communicate a random source to within a particular distortion level will function as follows (this is the usual way an architecture is constructed for communication with distortion):

1. A source encoder which will code (compress) the source to within the distortion level D
2. A channel encoder and a channel decoder which will help communicate the coded (compressed) source over the partially known channel, reliably
3. A channel decoder which will reconstruct the source

End-to-end, the source will be communicated to within the required distortion level.

From the descriptions in Sections 2.6 and 2.7, it follows that we still need to define the following rigorously:

- A general point-to-point communication system which communicates a random source to within a distortion level over a partially known channel
- A source code which codes (compresses) a source to within a particular distortion level

- Reliable communication over a partially known channel

This is the subject of the next few sections: Sections 2.8, 2.9, 2.10, and 2.11.

2.8 Distortion

We will allow the source reconstruction to be distorted compared to the source. In this section, we state some definitions related to distortion.

Notation 2.8 (D). $D > 0$ denotes a distortion level.

Definition 2.19 (Single letter distortion metric, d). $d : \mathcal{X} \times \mathcal{Y} \to [0,\infty)$ is the single letter distortion metric. Let $x \in \mathcal{X}, y \in \mathcal{Y}$. $d(x,y)$ is the distortion incurred if x is reconstructed as y.

Definition 2.20 (n letter distortion measure/metric, d^n). $d^n : \mathcal{X}^n \times \mathcal{Y}^n \to [0,\infty)$ is the n-letter distortion metric. Let $x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$. $d(x^n,y^n)$ is the distortion incurred if x^n is re-constructed as y^n.

Definition 2.21 (Permutation invariant n letter distortion metric). Let π^n be a permutation (rearrangement) of $(1,2,\ldots,n)$. That is, for $1 \le i \le n$, $\pi^n(i) \in \{1,2,\ldots,n\}$ and $\pi^n(1)$, $\pi^n(2),\ldots, \pi^n(n)$ are all different. For $x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$, denote

$$\pi^n(x^n) \triangleq (x^n(\pi^n(1)), x^n(\pi^n(2)), \ldots, x^n(\pi^n(n))) \tag{2.21}$$

$$\pi^n(y^n) \triangleq (y^n(\pi^n(1)), y^n(\pi^n(2)), \ldots, y^n(\pi^n(n))) \tag{2.22}$$

We will denote $\pi^n(x^n)$ as $\pi^n x^n$, and similarly, for the action of π^n on any sequence.

An n-letter distortion measure d^n is said to be permutation invariant if

$$d^n(\pi^n x^n, \pi^n y^n) = d^n(x^n, y^n) \tag{2.23}$$

Discussion 2.3 (Physical interpretation of permutation invariant distortion measure). Intuitively, a permutation invariant distortion measure is one where the distortion remains unchanged if both the input and the output are re-arranged. We would like to believe that physical distortion measures satisfy this requirement to some extent. For example, consider a voice signal (at the transmitter) and a corresponding reconstruction (which might not, as stated before, be a precise replica of the original voice signal) at the receiver. Suppose what the person spoke consisted of two sentences and the two sentences are interchanged. The reconstructed voice signal is rearranged in the same way at the receiver. We would like to believe that the receiver would be able to make out what the person spoke to the same extent in both of the above cases. The definition of a permutation invariant distortion measure is an abstraction of this. Note that this is a simplification in the sense that sentences of language have meaning and meaning is lost by permutation.

Definition 2.22 (Additive n letter distortion measure). n letter distortion metric d^n is said to be additive if

$$d^n(x^n, y^n) = \sum_{i=1}^{n} d(x^n(i), y^n(i)) \tag{2.24}$$

for some single letter distortion measure d

Note 2.14. Additive distortion measures are permutation invariant.

We would be interested in sequence of distortion measures for each block length n, $< d^n >_1^\infty$, and this is what we will call a distortion measure

Definition 2.23 (Permutation invariant distortion measure). A permutation invariant distortion measure is a sequence $< d^n >_1^\infty$ where d^n is a permutation invariant n letter distortion measure

Definition 2.24 (Additive distortion measure). An additive distortion measure $< d^n >_1^\infty$ is one for which each d^n is additive *with the same single letter distortion metric for each n*

■ 2.9 Universal communication of a random source over a partially known channel to within a certain distortion level

In this section, we define communication of a source over a partially known channel to within a certain distortion level.

Let $k \in \mathscr{A}$ be a partially known channel with input space $\mathscr{I}$ and output space $\mathscr{O}$ as described in Subsection 2.5.2.

First, we describe the point-to-point communication system which communicates i.i.d. X source over a channel $k \in \mathscr{A}$. Recall the action of a a point-to-point communication system in described in Subsection 2.6.6.

The input to the encoder is the i.i.d. X source. Thus, when the block length is n, the input is the i.i.d. X sequence of length n, X^n. The composition of the encoder, channel, and decoder, produce an output sequence Y^n. This results in a joint random variable X^nY^n on the input-output space $\mathscr{X}^n \times \mathscr{Y}^n$ and the corresponding probability distribution $p_{X^nY^n}$. Note that we are talking about a partially known channel and thus, $p_{X^nY^n}$ will vary depending on the particular $k \in \mathscr{A}$.

Definition 2.25 (A partially known channel which is *capable of* universally communicating i.i.d. X source to within a distortion level D). The partially known channel $k \in \mathscr{A}$ is said to be capable of universally communicating i.i.d. X source to within a distortion D if there exists a sequence $\omega =< \omega_n >_1^\infty$ such that $\omega_n \to 0$ as $n \to \infty$, and an encoder-decoder pair

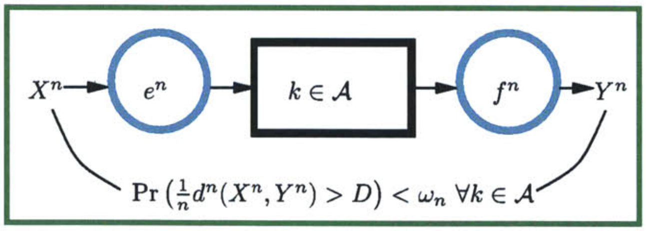


Figure 2.3. Universal communication to within a distortion D over a partially known channel $k \in \mathscr{A}$

$< e^n, f^n >_1^\infty$ independent of the particular $k \in \mathscr{A}$ such that under the joint distribution $p_{X^nY^n}$ as described above,

$$p_{X^nY^n}\left(\frac{1}{n}d^n(X^n, Y^n) > D\right) \leq \omega_n \; \forall \; k \in \mathscr{A} \tag{2.25}$$

See Figure 2.3. In the figure, we have not shown the common randomness input r to the encoder and the decoder. In the future, in this chapter, in the figures, we might not show the common randomness input. It will be assumed to be there.

Note 2.15 (Universal?). The word universal in the above definition refers to the fact that the same encoder-decoder work for all channels $k \in \mathscr{A}$.

Note 2.16 (Why ω?). The reason why the sequence $\omega = < \omega_n >_1^\infty$ is important is that it helps introduce *uniformity* in the rate at which error $\to 0$ as $n \to \infty$. If we made Definition 2.25 with

$$p_{X^nY^n}\left(\frac{1}{n}d^n(X^n, Y^n) > D\right) \to 0 \text{ as } n \to \infty \; \forall \; k \in \mathscr{A} \tag{2.26}$$

instead of (2.25), the rate of probability of distortion $> D$ will tend to 0 for all $k \in \mathscr{A}$, but this rate will not be independent of the particular channel $k \in \mathscr{A}$. In real scenarios, given a partially known channel, we would want to construct an encoder and decoder which will work and achieve a particular error criterion in the probability of distortion $> D$, and thus, we require a uniformity in the definition over all $k \in \mathscr{A}$.

Note 2.17 (Probability of excess distortion criterion).

$$p_{X^nY^n}\left(\frac{1}{n}d^n(X^n, Y^n) > D\right) \leq \omega_n \; \forall \; k \in \mathscr{A} \tag{2.27}$$

is called the probability of excess distortion criterion (for obvious reasons).

Note 2.18 (Intuitive action of a channel which is capable of communicating i.i.d. X source to within a certain distortion level). Intuitively, the composition of encoder, channel and

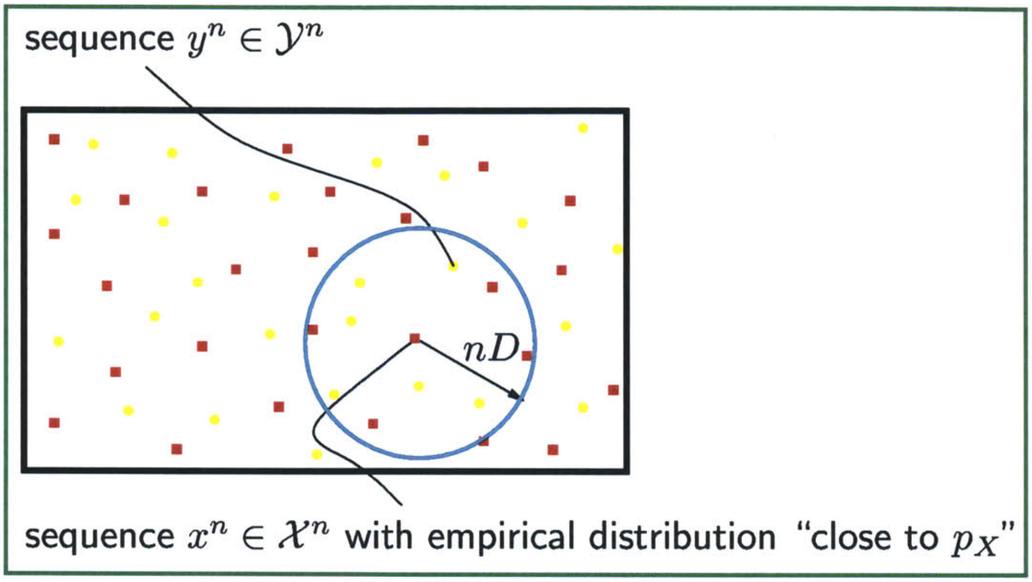


Figure 2.4. Pictorial action of a channel which is capable of communicating the i.i.d. X source to within a distortion level D

decoder in Definition 2.25 acts as follows: with high probability, a p_X typical sequence is distorted to within a ball of radius nD, and this probability $\to 1$ as the block length $n \to \infty$. See Figure 2.4. This figure superimposes the $\mathcal{X}^n$ and $\mathcal{Y}^n$ spaces. Red squares denote sequences in the $\mathcal{X}^n$ space with empirical distribution "close to p_X". Sequences x^n whose empirical distribution is "not close to p_X" do not affect the probability of excess distortion definition. For this reason, Figure 2.4 does not show these sequences. $y^n \in \mathcal{Y}^n$, however, can have any empirical distribution. Thus, the figure shows all points in $\mathcal{Y}^n$ and these are shown in gold circles.

When defining source codes which communicate sources to within particular distortion levels, in addition to the probability of excess distortion criterion, we will also consider the expected distortion criterion. Expected distortion criterion is the one more commonly found in literature, for example, in [Sha59].

■ 2.10 Source codes which code a source to within a particular distortion level, the rate-distortion source-coding problem, and the rate-distortion function

In this section, we define what it means for a source code to code a source to within a certain distortion level and the associated minimum rate at which this can be accomplished. This is followed by a discussion on the importance of the definition of a source code which codes a source to within a certain distortion level.

2.10.1 Source-coding or source compression?

We will use the terms source-coding and source-compression synonymously. This is because a source code is used to code a source to within a distortion level, and in effect, this is compressing the source.

2.10.2 Source codes which code (compress) a source to within a particular distortion level and the rate-distortion function

In this subsection, we define what it means for a source code to compress a source to within a certain distortion level and the associated minimum rate at which this can be accomplished.

Consider a source code $s =< s^n >_1^\infty$.

Let the block length be n. The composition of e_s^n and f_s^n is a transition probability $e_s^n \circ f_s^n(y^n|x^n)$. The action of $e_s^n \circ f_s^n$ on the n-length source X^n results in an output random variable Y^n on $\mathcal{Y}^n$, and thus, a joint random variable X^nY^n on $\mathcal{X}^n \times \mathcal{Y}^n$ with the corresponding probability distribution $p_{X^nY^n}$.

We consider two definitions of distortion: expected distortion and probability of excess distortion. These are defined below. The expected distortion definition is the one used usually in literature. This is the definition used by Shannon [Sha59]. The probability of excess distortion definition is used, for example, by Csiszar and Korner [CK97].

Definition 2.26 (Achievability of expected distortion D by source code s when encoding i.i.d. X Source). Distortion D is achievable in the expected sense (or that, distortion D is E-achievable, or that expected distortion D is achievable) by the source code s for the i.i.d. X source if under the joint distribution $p_{X^nY^n}$ as described above,

$$\limsup_{n\to\infty} \mathrm{E}_{X^nY^n}\left[\frac{1}{n}d^n(X^n,Y^n)\right] \leq D \tag{2.28}$$

See Figure 2.5. In this figure, the common randomness input r to e_s^n and f_s^n has been omitted.

Note 2.19. By definition, if distortion D is achievable in the expected sense (or that, distortion D is E-achievable) by the source code s for the i.i.d. X source, then distortion $D' > D$ is also achievable in the expected sense by the source code s for the i.i.d. X source.

Definition 2.27 (Rate-distortion function $R_X^E(D)$). Rate R is E-achievable corresponding to distortion level D for the i.i.d. X source if there exists a rate R source code s which achieves expected distortion D when encoding the i.i.d. X source. The infimum of all E-achievable rates for distortion level D is the rate-distortion function $R_X^E(D)$.

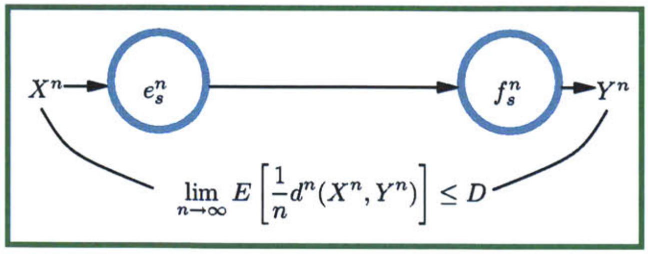


Figure 2.5. A source code which communicates i.i.d. X source to within an expected average distortion D

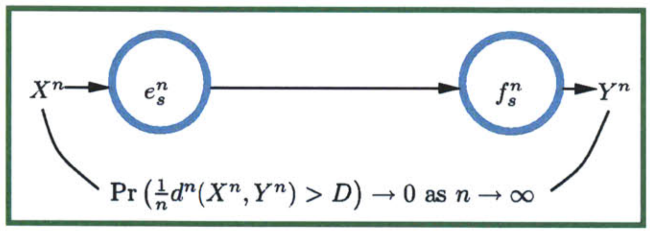


Figure 2.6. A source code which communicates i.i.d. X source to within a probability of excess distortion D

Definition 2.28 (Achievability of probability of excess distortion D by source code s when encoding i.i.d. X Source). Distortion D is achievable in the probability of excess distortion sense (or that, distortion D is P-achievable) by the source code s for the i.i.d. X source if

$$\lim_{n\to\infty} p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D\right) = 0 \tag{2.29}$$

See Figure 2.6. In this figure, the common randomness input r to e_s^n and f_s^n has been omitted.

Note 2.20. In the above definition, we use $\lim_{n\to\infty}$ and not $\limsup_{n\to\infty}$ because both definitions are the same. This is because, if $a_n \geq 0, 1 \leq n < \infty$, then, $\limsup_{n\to\infty} a_n = 0$ if and only if $\lim_{n\to\infty} a_n = 0$.

Note 2.21. By definition, if distortion D is achievable in the probability of excess distortion sense by the source code s for the i.i.d. X source, then distortion $D' > D$ is also achievable in the probability of excess distortion sense by the source code s for the i.i.d. X source.

Definition 2.29 (Rate-distortion function $R_X^P(D)$). Rate R is P-achievable corresponding to distortion level D for the i.i.d. X source if there exists a rate R source code s which achieves probability of excess distortion D for the i.i.d. X source. The infimum of all P-achievable rates for distortion level D is the rate-distortion function $R_X^P(D)$.

Note 2.22. The probability of excess distortion definition is local unlike the expected distortion definition which is global. In my opinion, the probability of excess distortion definition is also more intuitive, and makes more sense than the expected distortion definition. This is elaborated on, in Section 2.12.

The above definitions of $R_X^E(D)$ and $R_X^P(D)$ are based on the "physical" meaning of what it means to compress a source. The rate-distortion function can also be defined information-theoretically:

Definition 2.30 (The information-theoretic rate-distortion function $R_X^I(D)$).

$$R_X^I(D) \triangleq \inf_{\{p_{Y|X} \,:\, \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p_X(x) p_{Y|X}(y|x) \le D\}} I(X;Y) \tag{2.30}$$

The rate-distortion theorem of Shannon [Sha59] states that an expression for $R_X^E(D)$ is in fact, $R_X^I(D)$:

Theorem 2.1 (Rate-distortion theorem). *An expression for $R_X^E(D)$ is $R_X^I(D)$*

Note 2.23. *The source-coding problem that we consider is the above described problem of compressing the i.i.d. X source to within a distortion D under the expected and probability of excess distortion definitions.*

■ 2.10.3 The rate-distortion source-coding problem

The rate-distortion source-coding problem is to find the minimum rate needed to compress a source to within a certain distortion level. For the i.i.d. X source, these functions, as defined above, are denoted by $R_X^E(D)$ and $R_X^P(D)$.

■ 2.10.4 Discussion: Why are source codes which compress a source to within a certain distortion level, important?

In defining source codes which compress a source to within a certain distortion level, we have taken the composite transition probability corresponding to the composition of the source-encoder and the source-decoder. In practice, however, there is the channel encoder, channel and the channel decoder between the source encoder and the source decoder.

From discussion 2.1. it follows, on a high level, that the distortion introduced in a source (either under the expected distortion criterion or the probability of excess distortion criterion) after passing through the whole communication system consisting of the source encoder, the channel encoder, the channel, the channel decoder and the source decoder will be "close" to the distortion produced by a source code as defined in this section.

This will become clearer, after reliable communication has been defined rigorously. This is the subject of the next section.

■ 2.11 Universal capacity of a partially known channel

The way usual digital communication systems are constructed, the random binary sequence at the output of the source encoder is communicated with a very small error over the channel with the help of the channel encoder and the channel decoder. This motivates definition of reliable communication or the achievability of a rate R reliably, and the definition of channel capacity. The channel model we have is only partially known, and thus, the channel comes from a set $k \in \mathscr{A}$. Thus, we would be defining universal reliable achievability of rate R and universal capacity. The word "universal" refers to the fact that the same encoding-decoding scheme should work for all channels in the set.

Let $k \in \mathscr{A}$ be a partially known channel with input space $\mathscr{I}$ and output space $\mathscr{O}$ as described in Subsection 2.5.2. This is the partially known channel.

Recall the definition of a channel encoder $< e_c^n >_1^\infty$ and a channel decoder $< f_c^n >_1^\infty$ defined in Subsection 2.7.3.

When the block length is n, the input to the source encoder e_s^n is X^n. X^n is source encoded by e_s^n and the output is a random binary sequence which is a random variable M_R^n with corresponding distribution $p_{M_R^n}$. M_R^n is the input to the channel encoder e_c^n. The output of the channel encoder is I^n which is an input to the channel k^n. The channel produces output O^n which is an input to the channel decoder f_c^n, which reconstructs the random binary sequence as $\hat{M}_R^n$.

$\hat{M}_R^n$ is the input to the source decoder f_s^n which produces the reconstructed source output Y^n.

For now, we are only interested in the dynamics of M_R^n to $\hat{M}_R^n$.

If the source code has rate R, the output of the source encoder is a rate R random binary sequence M_R^n with the corresponding probability distribution on the set $\mathscr{M}_R^n = \{1,2,\ldots,2^{\lfloor nR \rfloor}\}$. $p_{M_R^n}$ can, in general, be quite arbitrary depending on e_s^n, and we will assume that it can be any possible probability distribution on $\{1,2,\ldots,2^{\lfloor nR \rfloor}\}$.

In a usual digital communication system, the channel encoder $< e_c^n >_1^\infty$ and the channel decoder $< f_c^n >_1^\infty$ are constructed in such a way that the random binary sequence should be communicated over the channel with a very small error. This is abstracted out by saying that the small error $\to 0$ as the block length $n \to \infty$.

Let the input to the channel encoder e_c^n be a rate R random binary sequence M_R^n. The output of the channel decoder is the reconstruction of the random binary sequence, $\hat{M}_R^n$. Note that $\hat{M}_R^n$ will depend on the particular $k \in \mathscr{A}$; however, we do not show this dependence. For reliable communication, we would require that

$$\Pr(\hat{M}_R^n \neq M_R^n) \to 0 \text{ as} n \to \infty \tag{2.31}$$

As we said above, the distribution $p_{M_R^n}$ to be arbitrary, the above should hold irrespective of the distribution $p_{M_R^n}$. One way of ensuring this is to say that this happens on a per message level with a uniformity over all messages

$$\Pr(\hat{M}_R^n \neq M_R^n | M_R^n = m^n) < \delta_n \ \forall m^n \in \mathcal{M}_R^n, \text{ for some } \delta_n \to 0 \text{ as } n \to \infty \tag{2.32}$$

Note that if

$$\Pr(\hat{M}_R^n \neq M_R^n | M_R^n = m^n) < \delta_n \ \forall m^n \in \mathcal{M}_R^n, \tag{2.33}$$

then it follows that

$$\Pr(\hat{M}_R^n \neq M_R^n) < \delta_n \text{ as } n \to \infty \tag{2.34}$$

for arbitrary distribution $p_{M_R^n}$, and note that δ_n is independent of the distribution $p_{M_R^n}$. The fact that δ_n is independent of the distribution $p_{M_R^n}$ can be important if we do not know the distribution apriori of the source. Then, after compression, the distribution $p_{M_R^n}$ might not be known. The system should be able to provide a certain error guarantee for the same block length (this is important when building the system) irrespective of the distribution of the source, and thus, it is important that δ_n be independent of the distribution $p_{M_R^n}$.

Also, we would want the rate of fall of error probability to zero with increasing block length at a uniform rate independent of which particular realization of the channel $k \in \mathcal{A}$ occurs. This is because, when building a real system, we will not know which particular channel $k \in \mathcal{A}$ will happen, and the communication guarantee should hold irrespective of this guarantee.

This motivates the definition of universal reliable achievability of rate R over a partially known channel.

Definition 2.31 (Universal reliable achievability of rate R over a partially known channel $k \in \mathcal{A}$). Rate R is said to be universally achievable over the channel set $\mathcal{A}$ if there exists a channel code $< e_c^n, f_c^n >_1^\infty$, independent of the particular $k \in \mathcal{A}$, and if there exists a sequence $< \delta_n >_1^\infty$, $\delta_n \to 0$ as $n \to \infty$ independent of the particular $k \in \mathcal{A}$ such that

$$\Pr(\hat{M}_R^n \neq M_R^n | M_R^n = m^n) < \delta_n, \forall m^n \in \mathcal{M}_R^n, \forall k =< k^n >_1^\infty \in \mathcal{A} \tag{2.35}$$

See Figure 2.7. The common randomness input r to the channel encoder and the channel decoder exists but has been omitted in the figure.

Note 2.24 (Universal?). Universality in the above definition refers to the fact that the same encoder-decoder work for the partially known channel $k \in \mathcal{A}$.

Definition 2.32 (Universal capacity of the partially known channel $k \in \mathcal{A}$, $C_{rc}(\mathcal{A})$). The supremum of all universally achievable rates over the partially known channel $k \in \mathcal{A}$ is the universal capacity of the partially known channel $k \in \mathcal{A}$, and is denoted by $C_{rc}(\mathcal{A})$.

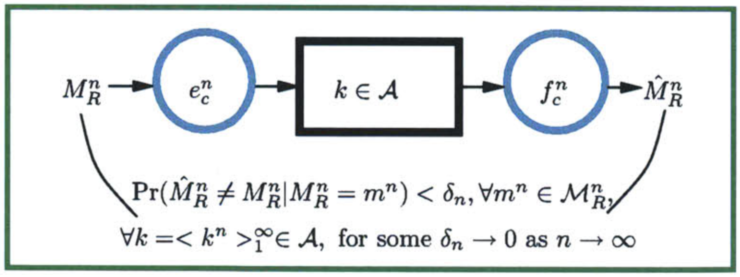


Figure 2.7. Universal reliable communication over partially known channel k

Note 2.25. The sub-script rc in $C_{rc}(\mathscr{A})$ stands for random-coding. We allow the encoder and decoder to be random, and thus, $C_{rc}(\mathscr{A})$ is the random-coding universal capacity of the partially known channel $k \in \mathscr{A}$.

Note 2.26 (Resource consumption). One would also like, in the definition of universal capacity, to have a dependence on the consumption of system resources. As we shall see, results concerning resource consumption that we need will automatically result from the encoding-decoding scheme that we will use and for that reason, we are not making the capacity depend on any resource constraints. In general, however, the capacity should be defined by restricting the encoder and decoder in such a way that only certain system resources are being consumed.

Note 2.27 (Universal capacity and compound capacity). The definition of universal capacity of a partially known channel is the same, in spirit, as the definition of compound capacity as defined in information theory literature, see, for example, [CK97]. [CK97] talks about a compound DMC and thus, the partially known channel comes from a set which consists of discrete memoryless channels (DMCs). We allow the set to consist of arbitrary channels *with the same input and output space.*

Note 2.28 (The universal capacity of a set of abstract channels). An abstract channel, that is, thinking of the composition of an encoder, channel, and a decoder, as a channel, was defined in Subsection 2.6.5. Universal capacity can analogously be defined for a set of abstract channels with input space $\mathscr{X}$ and output space $\mathscr{Y}$. We will use this view of universal capacity of a set of abstract channels in Sections 2.14 and 2.15, when proving the universal source-channel separation theorem for rate-distortion.

The channel-coding problem that we will consider is that of the universal capacity of a partially known channel which is capable of universally communicating i.i.d. X source to within a distortion D. We will relate this universal capacity to the rate-distortion function for the i.i.d. X source. This will be used to prove the universal source-channel separation theorem for rate-distortion.

2.12 A comparison of the expected distortion and the probability of excess distortion criterion and the reason why we use the probability of excess distortion criterion

In this section, we compare the expected distortion and the probability of excess distortion criteria. We also state the reason why we use the probability of excess distortion criterion and not the expected distortion criterion.

2.12.1 A comparison of the expected distortion and the probability of excess distortion criteria

We use the probability of excess distortion criterion (2.29) for communication to within a distortion level D and not the expected distortion criterion (2.28). The following is a comparison of the two criteria:

The probability of excess distortion criterion is local whereas the expected distortion criterion is not. This observation has been made by Csiszar and Korner [CK97]. The probability of excess distortion criterion is local in the sense that it says (roughly) that with high probability, a sequence $x^n \in \mathscr{X}^n$ whose empirical distribution is "close to p_X" is distorted to within a ball of radius nD with high probability and this probability $\rightarrow 1$ as the block length $n \rightarrow \infty$. Thus, with the probability of excess distortion criterion, statements can be made (with high probability) concerning individual x^n sequences. The expected distortion criterion, however, is an expectation condition, and hence, global (not local) because statements cannot be made about particular p_X typical sequences.

For this reason, the probability of excess distortion criterion is also more intuitive: it can be represented pictorially, see Figure 2.4 as opposed to the expected distortion criterion.

For the same reason, we believe that the probability of excess distortion criterion makes more sense than the expected distortion criterion. When communicating a source sequence, we would like to have most sequences be communicated with a certain guarantee, rather than a guarantee averaged over all the source sequences.

We should add that the probability of excess distortion criterion is stronger than the expected distortion criterion in the sense that if the probability of excess distortion criterion holds, the expected distortion criterion also holds (under minor technical assumptions).

2.12.2 Why do we use the probability of excess distortion criterion instead of the expected distortion criterion?

Most literature on information theory uses the expected distortion criterion. However, we use the probability of excess distortion criterion.

We use the probability of excess distortion criterion (2.29) for communication to within a distor-

tion level D instead of the expected distortion criterion (2.28) because we do not know how to prove universal results with the expected distortion criterion.

Also, as discussed in the previous sub-section, in my opinion, the probability of excess distortion criterion makes *more* sense than the expected distortion criterion. Finally, from a practical standpoint, none of them really makes sense: for example, if we want to put a distortion criterion on voice, neither the expected distortion criterion or the probability of excess distortion criterion make sense. However, in my opinion, the probability of excess distortion criterion does make "more" sense. I will add that the probability of excess distortion criterion making "more" sense than the expected distortion criterion is a very marginal reason for using the probability of excess distortion criterion: the main reason, as we said above, is that we do not know how to prove universal results with the expected distortion criterion.

■ 2.13 Important past literature

Shannon proved a source-channel separation theorem for rate-distortion in the point-to-point setting for the problem of reliable communication in [Sha48]. In this paper, Shannon also hinted at, but did not give any proofs of optimality of separation for communication with distortion. This was the subject of [Sha59].

As we have said before, Shannon assumed knowledge of the channel as a transition probability. We only assume partial knowledge of the channel as a transition probability, and prove a universal source-channel separation theorem for communication with distortion in the point-to-point setting. This is the main contribution of this chapter. In Chapter 3, these results are generalized to the multiuser setting.

Shannon assumed that the distortion measure is additive. Since then, results have been generalized to sub-additive distortion measures, see for example [Han10]. We prove our results with the assumption of permutation invariant distortion measures. There is no real relation, to the best of our understanding, between sub-additive distortion measures and permutation invariant distortion measures.

■ 2.14 The main ideas for why separation holds for universal communication with a fidelity criterion: separation for the uniform X source under a technical assumption on the rate-distortion function

This is the most important section in the whole thesis, and discusses the main idea for why separation or the optimality of digital communication holds for universal communication with a fidelity criterion.

We prove the universal source-channel separation theorem for rate-distortion for what we call, the uniform X source (which is defined below) under a technical assumption on behavior of the rate-distortion function. Throughout this section, the distortion metric is assumed to be permutation invariant.

2.14.1 The organization of this section

This section is organized as follows:

Subsection 2.14.2 discusses the uniform X source which we use throughout this section instead of the traditional i.i.d. X source. The uniform X source consists of all sequences with type precisely p_X. It thus has a single type class. We use the uniform X source because the proofs with a source which has a single type class avoids a lot of ϵs and δs in the proofs.

Subsection 2.14.3 discusses source codes and various rate-distortion functions when coding the uniform X source. These would be direct generalizations of the definitions for the i.i.d. X source. One new definition that we will introduce will be that of the inf rate-distortion function, and this would be related to the technical assumption that we will make.

Subsection 2.14.4 discusses encoders and decoders, and further discusses the capability of a partially known channel for universally communicating the uniform X source to within a certain distortion level.

Subsection 2.14.5 states the technical condition that we will need on the rate-distortion function in order to prove the universal source-channel separation theorem for rate-distortion.

Subsection 2.14.6 states the universal source-channel separation theorem for universal communication with a fidelity criterion for the uniform X source, Subsection 2.14.7 discusses the two steps in the proof, and Subsection 2.14.9 proves it. Before the proof, we make a small note on random-coding in Subsection 2.14.8.

This is followed by various discussions in Subsection 2.14.10 on the nature of the proof, why separation holds, and connections between source and channel coding.

Finally, we make a short note on the technical assumption that we make concerning the rate-distortion function in Subsection 2.14.11.

2.14.2 The uniform X source

Definition 2.33 (Uniform X source). Let X be a random variable on $\mathscr{X}$. *Let $p_X(x)$ be rational* $\forall x$. Let n_0 be the least positive integer for which $n_0 p_X(x)$ is an integer $\forall x \in \mathscr{X}$. Let $\mathscr{U}^n$ denote the set of sequences with (*exact*) empirical distribution (type) p_X. $\mathscr{U}^n$ is nonempty if and only if n_0 divides n. Let $n' \triangleq n_0 n$. Let $U^{n'}$ denote a random variable which is uniform on $\mathscr{U}^{n'}$ and zero elsewhere. Then, $< U^{n'} >_1^\infty$ is the uniform X source and is denoted by U. *Intuitively, the uniform X source is the source which puts uniform distribution on the set of all sequences whose empirical distribution is p_X.*

Note 2.29. The superscript n' in $\mathscr{U}^{n'}$ denotes that the block length is n'. It *does not* mean that $\mathcal{U}^{n'}$ is the cartesian product of some set $\mathcal{U}$ with itself n times. In fact, the set $\mathcal{U} = \mathcal{U}^1$ is empty unless $n_0 = 1$. Similarly, the superscript n' in $U^{n'}$ denotes block length. It *does not* mean that $U^{n'}$ is i.i.d. U source for some random variable U.

Definition 2.34 (n_0). n_0 is the least positive integer for which $n_0 p_X(x)$ is an integer $\forall x \in \mathcal{X}$.

Definition 2.35 (n'). $n' \triangleq n_0 n$.

Note 2.30 (Uniform X source makes sense only for block lengths divisible by n_0). *Uniform X source is defined only for those block lengths which are divisible by n_0.*

Note 2.31. If $p_X(x)$ is irrational for some $x \in \mathcal{X}$, $\mathcal{U}^n$ is empty $\forall n$. Thus, in order to define the uniform X source, the assumption that $p_X(x)$ be rational $\forall x \in \mathcal{X}$ is necessary.

Note 2.32. Let $p_X(x)$ be rational $\forall x \in \mathcal{X}$. The uniform X source and the i.i.d. X source are "close" to each other in the following sense. The uniform X source puts mass only on sequences with empirical distribution *exactly* p_X. For large n, i.i.d. X source puts "most of" its mass on sequences with empirical distribution "close to" p_X. We are interested in i.i.d. X source.

Note 2.33 (Why use the uniform X source). Uniform X Source has a single type class by definition. This helps avoid a lot of ϵs and δs in arguments.

2.14.3 Source codes for the uniform X source and rate-distortion functions for the uniform X source

Assume that X is such that $p_X(x)$ is rational $\forall x \in \mathcal{X}$.

The definitions of source codes for coding the uniform X and the transition probability corresponding to such a source code are analogous to the corresponding definitions, Definition 2.11, Definition 2.12 and Definition 2.13 of Subsection 2.7.2. There are two differences:

- Since the uniform X source is defined only for those block lengths of the form $n' = n_0 n$, source codes for the uniform X source are defined only for these block lengths
- The input space, when the block length is n' is $\mathcal{U}^{n'}$ and not $\mathcal{X}^{n'}$

Achievability of expected distortion D when coding the uniform X source is defined analogously to Definition 2.26, except that (2.28) is replaced by

$$\limsup_{n' \to \infty} \mathrm{E}_{U^{n'} Y^{n'}} \left[\frac{1}{n} d^{n'}(U^{n'}, Y^{n'}) \right] \leq D \tag{2.36}$$

The rate-distortion function $R^E_U(D)$ is defined analogously to (2.27).

Achievability of probability of excess distortion D when coding the uniform X source is defined analogously to Definition 2.28, except that (2.29) is replaced by

$$\lim_{n' \to \infty} p_{U^{n'}Y^{n'}}\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D\right) = 0 \tag{2.37}$$

The rate-distortion function $R_U^P(D)$ is defined analogously to (2.29).

We also need the definition of *inf achievability* of probability of excess distortion D when coding the uniform X source. This is the same as the definition of achievability of probability of excess distortion D when coding the uniform X source, except that (2.37) is replaced with

$$\liminf_{n' \to \infty} p_{U^{n'}Y^{n'}}\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D\right) = 0 \tag{2.38}$$

The above is called the inf probability of excess distortion criterion.

The rate-distortion function $R_U^P(D, \text{inf})$ can be defined analogously to $R_U^P(D)$ by using (2.38) instead of (2.37).

2.14.4 Encoders and decoders to communicate the uniform X source, and universal communication of the uniform X source over a channel to within a certain distortion level

In this subsection, we want to define, what it means for a partially known physical channel to be capable of communicating the uniform X source to within a certain distortion level.

A physical channel $k =< k^n >_1^\infty$ has been defined in Subsection 2.5.2 and a partially known physical channel $k \in \mathscr{A}$ has been defined in Subsection 2.5.3. The channel evolves for every integer time. However, what will matter, when we consider the interconnection of the encoder, channel and the decoder is the channel subsequence $< k^{n'} >_1^\infty$.

The view of an analog point-to-point communication system to communicate the uniform X source is the same as the view of Section 2.6 with the following differences:

- Encoders and decoders are sequences $< e^{n'} >_1^\infty$ and $< d^{n'} >_1^\infty$ defined only for those block lengths n' which are divisible by n_0
- The input space when the block length is n' is $\mathscr{U}^{n'}$ instead of $\mathscr{X}^{n'}$
- The interconnection will be made among encoder $< e^{n'} >_1^\infty$, channel $< k^{n'} >_1^\infty$ and decoder $< f^{n'} >_1^\infty$

It is important to note that encoders, channels and decoders are defined only for block lengths n' of the form $n_0 n$, but they evolve, as before, for each integer time.

The view of a point-to-point digital communication system to communicate the uniform X source is the same as that of Section 2.7 with the same differences:

- Source encoders, source decoders, channel encoders, channel decoders are defined only for those block lengths n' divislble by n_0
- The input space when the block length is n' is $\mathscr{U}^{n'}$ instead of $\mathscr{X}^{n'}$

The distortion function is defined in the same way as Section 2.8 except for the same reasons that the sequence is $< d^{n'} >_1^\infty$ and $d^{n'} : \mathscr{U}^{n'} \times \mathscr{Y}^{n'} \to [0,\infty)$ instead of $d^n : \mathscr{X}^n \times \mathscr{Y}^n \to [0,\infty)$. Definition of permutation invariance and additiveness is defined in the same way as in Section 2.8.

Consider a partially known channel $k \in \mathscr{A}$.

The input to the encoder is the uniform X source. Thus, then the block length is n', the input is the i.i.d. X sequence of length n', $U^{n'}$. The composition of the encoder, channel, and decoder, produce an output sequence $Y^{n'}$. This results in a joint random variable $U^{n'}Y^{n'}$ on the input-output space $U^{n'} \times \mathscr{Y}^{n'}$ and the corresponding probability distribution $p_{U^{n'}Y^{n'}}$. Note that we are talking about a partially known channel and thus, $p_{U^{n'}Y^{n'}}$ will vary depending on the particular $k \in \mathscr{A}$.

The partially known channel $k \in \mathscr{A}$ is said to be capable of universally communicating uniform X source to within a distortion D if there exists a sequence $\omega =< \omega_{n'} >_1^\infty$ such that $\omega_{n'} \to 0$ as $n \to \infty$, and an encoder-decoder pair $< e^{n'}, f^{n'} >_1^\infty$ independent of the particular $k \in \mathscr{A}$ such that under the joint distribution $p_{U^{n'}Y^{n'}}$ as described above,

$$p_{U^{n'}Y^{n'}}\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D\right) \leq \omega_{n'} \ \forall\, k \in \mathscr{A} \tag{2.39}$$

See Figure 2.3, except that all n are replaced with n' and X^n is replaced with $U^{n'}$. In the figure, we have not shown the common randomness input r to the encoder and the decoder.

2.14.5 The technical condition on the rate-distortion function that we will require in order to prove the universal source-channel separation theorem for rate-distortion for the uniform X source under a permutation invariant distortion metric

We will assume that $R_U^P(D) = R_U^P(D, \text{inf})$. We will prove this for an additive distortion metric in Chapter 5.

Another assumption which we require is that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. We require this assumption because otherwise, the uniform X source is not defined.

2.14.6 A statement of universal source-channel separation theorem for rate-distortion for the uniform X source

Theorem 2.2 (Universal source-channel separation theorem for rate-distortion in the point-to-point setting for the uniform X Source / Optimality of digital communication for universal communication of the uniform X source with a fidelity criterion). *Let $d =< d^{n'} >_1^\infty$ be a permutation invariant distortion metric under which $R_U^P(D) = R_U^P(D, \inf)$. Assuming random-coding is permitted, in order to communicate the uniform X source over a partially known channel to within a particular distortion level, it is sufficient to consider source-channel separation based architectures, that is, architectures which first compress the uniform X source to within the particular distortion level, followed universal reliable communication over the partially known channel. There is sufficiency in the sense if there exists some architecture to communicate the uniform X source to within a certain distortion universally over the partially known channel, and which consumes certain amount of system resources (for example, energy and bandwidth), then there exists a separation based scheme to universally communicate the uniform X source to within the same distortion universally over the partially known channel and which consumes the same or lesser system resources as the original scheme.*

2.14.7 Steps to prove Theorem 2.2

In this section, we state the steps in proving Theorem 2.2. The steps are:

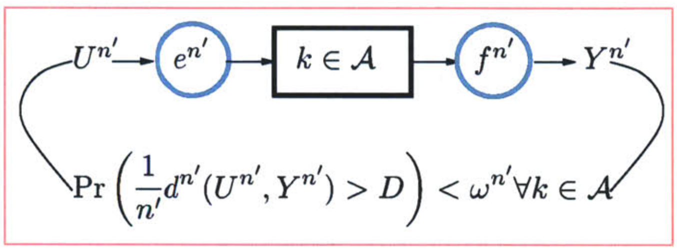
- *Step 1 for why Theorem 2.2 holds:* Given a partially known channel k which is capable of universally communicating uniform X source to within a distortion D, first prove that the universal capacity of the partially known channel k is larger than or equal to the rate-distortion function $R_U^P(D)$. Also, prove that the universal reliable communication at rates $< R_U^P(D)$ can be accomplished by using an encoder and a decoder such that the resulting architecture consisting of the composition of the encoder, channel and decoder, when used for universal reliable communication, consumes the same system resources irrespective of the distribution on the message set as the original architecture when used for universal communication to within a distortion level D of the uniform X source over the partially known channel.

  This step is illustrated in Figure 2.14.7.

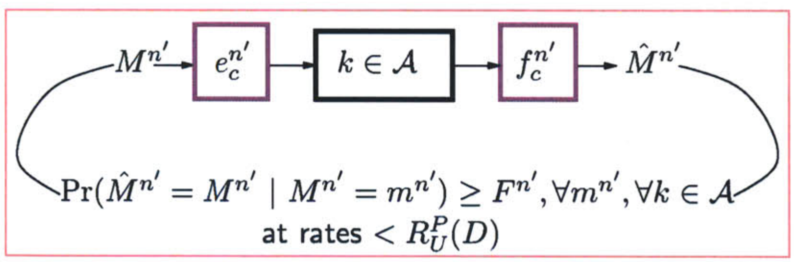
- *Step 2 for why, the Theorem 2.2 holds:* Thus, in fact, given a partially known channel which is capable of communicating uniform X source to within a probability of excess distortion D, and hence, from Step 1, its universal capacity is $\geq R_U^P(D)$, universal communication to within a distortion D over the partially known channel could actually be carried out by first compressing the uniform X source to within a distortion D under the probability of excess distortion criterion and then communicating the resulting rate $R_U^P(D)$ random binary sequence universally and reliably over the partially known channel. Since the reliable communication can be accomplished by using an encoder-decoder

Step 1:

Given, $\exists < e^{n'}, f^{n'} >_1^\infty$ and $< \omega^{n'} >_1^\infty$, $\omega^{n'} \to 0$ as $n' \to \infty$ such that



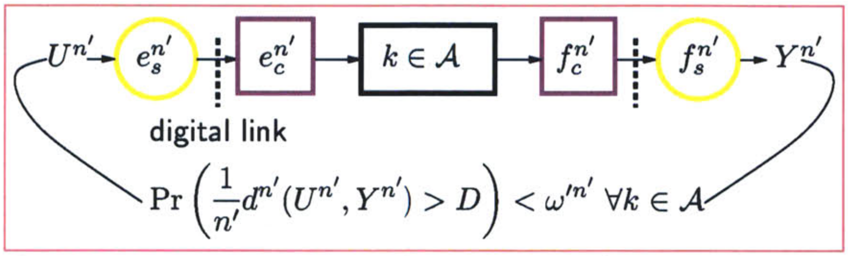
prove $\exists < e_c^{n'}, f_c^{n'} >_1^\infty$ and $< F^{n'} >_1^\infty$, $F^{n'} \to 1$ as $n' \to \infty$ such that



such that resource consumption remains unchanged.

Step 2:

Thus, $\exists$ source code $< e_s^{n'}, f_s^{n'} >_1^\infty$, and $< \omega'^{n'} >_1^\infty$, $\omega'^{n'} \to 0$ as $n' \to \infty$ such that



such that resource consumption is the same as in the original architecture of Step 1

such that the resulting architecture consisting of the composition of the digital encoder, channel and digital decoder consumes the same system resources as the original architecture to universally communicate the uniform X source to within a distortion D over the partially known channel k, the digital architecture to communicate the uniform X source to within a distortion D over the partially known channel k also consumes the same system resources.

This step is illustrated in Figure 2.14.7.

We argue these steps in Subsection 2.14.9. Before that, we make a note on random-coding in the next subsection.

2.14.8 Random codes

In part of the proof, we will generate codebooks uniformly from the set of all sequences which have type precisely X. As we have emphasized before, for us, random-coding is not just a proof technique. It is essential. This will be discussed further in Section 2.17.

2.14.9 The proof of Theorem 2.2

Proof. *Proof of Step 1 in order to prove Theorem 2.2*

Recall that we will denote $n' = n_0 n$.

Let $k =< k^{n'} >_1^\infty \in \mathscr{A}$ be a partially known channel which is capable of universally communicating the uniform X source U to within a distortion D. Thus, there exist an encoder-decoder $< e^{n'}, f^{n'} >_1^\infty$ and a sequence $\omega =< \omega_{n'} >_1^\infty$, $\omega_{n'} \to 0$ as $n' \to \infty$ such that with the composition of the encoder, channel and decoder, $< e^{n'} \circ k^{n'} \circ f^{n'} >_1^\infty$ with input $U =< U^{n'} >_1^\infty$, end to end,

$$\Pr\left(\frac{1}{n'} d^{n'}(U^{n'}, Y^{n'}) > D\right) < \omega_{n'}, \forall k \in \mathscr{A} \tag{2.40}$$

Consider the partially known abstract channel

$$c \in \{< e^{n'} \circ k^{n'} \circ f^{n'} >_1^\infty | k =< k^{n'} >_1^\infty \in \mathscr{A}\} \triangleq \mathscr{C}_{\mathscr{A}} \tag{2.41}$$

We will prove that the universal capacity of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$ is $\geq R_U^P(D)$, and this can be accomplished by an encoder-decoder $< E^{n'}, F^{n'} >_1^\infty$ such that for universal reliable communication at rates $R < R_X^P(D)$, the point-to-point communication system $< E^{n'} \circ c^{n'} \circ F^{n'} >_1^\infty$ consumes the same system resources (irrespective of the particular $c \in \mathscr{A}$) as the original point-to-point communication system $< e^{n'} \circ k^{n'} \circ f^{n'} >_1^\infty$ when used to communicate the uniform X source universally to within a distortion D.

From this it will follow that the universal capacity of the partially known channel k is $\geq R_U^P(D)$, and this universal reliable communication at rates $< R_X^P(D)$ can be accomplished with the help of encoder $< e_c^{n'} >_1^\infty =< E^{n'} \circ e^{n'} >_1^\infty$ and decoder $< f_c^{n'} >_1^\infty =< f^{n'} \circ F^{n'} >_1^\infty$. The point-to-point communication system $< e_c^{n'} \circ k^{n'} \circ f_c^{n'} >_1^\infty$ when used for reliable communication at rates $< R_X^P(D)$ consumes the same system resources as the original point-to-point communication system $< e^{n'} \circ k^{n'} \circ f^{n'} >_1^\infty$ when used to communicate the uniform X source to within a distortion D.

We proceed to prove that the universal capacity of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$ is $\geq R_U^P(D)$. Note that we have assumed that $R_U^P(D) = R_U^P(D, \text{inf})$. Thus, it is sufficient to prove that the universal capacity of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$ is $\geq R_U^P(D, \text{inf})$. This is what we proceed to prove.

This is done via parallel random-coding arguments for

- the universal capacity of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$, and
- the rate-distortion source-coding problem of finding the minimum rate needed to compress the uniform X source to within a distortion D under the inf probability of excess distortion criterion.

The random-coding arguments are similar, yet different from the ones used in the information theory literature. We want to derive a connection between the above two problems in order to prove the desired result, and we are not interested in simplified functional expressions for the universal capacity of the partially known channel $c \in \mathscr{A}$ or simplified expressions for the rate-distortion function $R_U^P(D, \inf)$.

The two problems:

- *The channel-coding problem:* The channel-coding problem is that of computing the universal capacity of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$
- *The source-coding problem:* The source-coding problem that we consider is to derive an upper bound on $R_U^P(D, \inf)$, the minimum rate needed to compress the uniform X source to within a distortion level D under the inf probability of excess distortion criterion

Block length: For both the channel coding and the source coding problems, let the block length be n'. Towards the end of the argument we will take the limit $n' \to \infty$. Recall that $n' = n_0 n$ is the set of all integers for which the uniform X source makes sense.

Codebook generation:

- *Codebook for the channel-coding problem:* Let communication be desired at rate R. Generate $2^{\lfloor n'R \rfloor}$ sequences independently and uniformly from the set $\mathscr{U}^{n'}$, the set of all sequences $\in \mathscr{X}^n$ which have empirical distribution *precisely* p_X.

  This is the code book $\mathscr{K}^{n'}$. Note that the codewords $\in \mathscr{U}^{n'}$. The encoder is denoted by $< E^{n'} >_1^\infty$. Note that the encoder is random.

- *Codebook for the source-coding problem:* Let q be an empirical distribution (type) on $\mathscr{Y}$, that is $q \in \mathscr{P}(\mathscr{Y})$. Let q be an achievable type when the block length is n'. In other words, $n'q(y)$ is an integer $\forall y \in \mathscr{Y}$. Let $\mathscr{U}_q^{n'} \subset \mathscr{Y}^{n'}$ denote the set of all sequences with empirical distribution, precisely q. Generate $2^{\lfloor n'R \rfloor}$ codewords independently and uniformly from the set $\mathscr{U}_q^{n'}$

  This is the code book $\mathscr{L}^{n'}$. Note that the codewords $\in \mathscr{U}_q^{n'} \subset \mathscr{Y}^{n'}$. Note that the codebook is random.

Joint typicality:

- *Joint typicality for the channel coding problem:* Sequences $(u^{n'}, y^{n'}) \in$ the channel input-output space $\mathscr{U}^{n'} \times \mathscr{Y}^{n'}$ are said to be jointly typical if

$$\frac{1}{n'} d^{n'}(u^{n'}, y^{n'}) \le D \tag{2.42}$$

- *Joint typicality for the source coding problem:* Sequences $(u^{n'}, y^{n'}) \in$ the source input - source reconstruction space $\mathscr{U}^{n'} \times \mathscr{Y}^{n'}$ are said to be jointly typical if

$$\frac{1}{n'} d^{n'}(u^{n'}, y^{n'}) \leq D \tag{2.43}$$

Note that the definition of joint-typicality for both the channel-coding and the source-coding problems is the same.

Note 2.34 (A note on the definition of joint typicality). Jointly typical sequences are defined in the information theory literature in a way so that the set of all jointly typical sequences occurs with high probability and this probability $\to 1$ as the block length $n' \to \infty$. In our framework, from the codebook generation and from the action of the channel, all we know is that

$$\frac{1}{n'} d^{n'}(u^{n'}, y^{n'}) \leq D \tag{2.44}$$

with high probability , and this is, thus, our definition of jointly typical sequences. In usual information theory frameworks, the channel action is known as a transition probability and thus, when defining jointly typical sequences, there is usually a requirement on the conditional type of the output sequence given the input sequence. However, our description of the channel is not in terms of a transition probability. Our description of the channel is in terms of the distortion that it produces on sequences with type precisely X and hence, the above definition of joint typicality.

Decoding:

- *Decoding for the channel coding problem:* Let the sequence $y^{n'}$ be received. If there exists *unique* codeword $u^{n'}$ in the code book $\mathscr{K}^{n'}$ for which $(u^{n'}, y^{n'})$ are jointly typical, declare that $u^{n'}$ is transmitted, else declare error. The decoder is denoted by $F^{n'}$. Note that the encoder-decoder $E^{n'}, F^{n'}$ is random

  Note 2.35. This decoding rule can be thought of as a variant of minimum distance decoding

  Note 2.36 (Are the encoder-decoder random-codes in the sense of Definition 2.3). We defined random codes in Definition 2.3. In our coding scheme, we are generating codewords uniformly from the set $\mathscr{U}^{n'}$ and the decoder is, as defined above, a variant of minimum distance decoding . This encoder-decoder be thought of as a random-code in the sense of Definition 2.3. We leave out an elaboration as to why that is the case. As disussed in Note 2.6, our encoder-decoder can thus be generated using common randomness.

- *Encoding for the source coding problem:* Let the sequence $u^{n'} \in \mathscr{U}^{n'}$ needs to be source coded. If there exists some sequence $y^{n'}$ in the code book $\mathscr{L}^{n'}$ for which $(u^{n'}, y^{n'})$ are jointly typical, encode $u^{n'}$ to one such $y^{n'}$, else declare error. Note that the encoder-decoder is random.

Note 2.37. Note that "unique" in the channel coding problem gets converted to "some" in the source coding problem

Some notation:

- *Notation for the channel coding problem:* We will do the analysis assuming that a particular message is transmitted. The message set is

$$\mathcal{M}_R^{n'} = \{1, 2, \ldots, 2^{\lfloor n'R \rfloor}\} \tag{2.45}$$

  Assume that message $m_i^{n'} \in \mathcal{M}_R^{n'}$ is transmitted.

  Let the codeword corresponding to message $m_i^{n'}$ be denoted by $u_c^{n'}$. Let the nontransmitted codewords be denoted by $u'^{n'}_1, u'^{n'}_2, u'^{n'}_{2^{\lfloor n'R \rfloor}-1}$.

  $u_c^{n'}$ is a realization of $U_c^{n'}$. By the random code book generation, $U_c^{n'}$ has uniform distribution on $\mathcal{U}^{n'}$.

  $u'^{n'}_i$ is a realization of $U'^{n'}_i$, $1 \le i \le 2^{\lfloor n'R \rfloor} - 1$. By the random code book generation, $U'^{n'}_i$, $1 \le i \le 2^{\lfloor n'R \rfloor} - 1$, has uniform distribution on $\mathcal{U}^{n'}$.

  By the random code book generation, the codewords are generated independently of each other, and thus, $U_c^{n'}$, $U'^{n'}_i$, $1 \le i \le 2^{\lfloor n'R \rfloor} - 1$ are all independent of each other as random variables.

  The action of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$ on the transmitted codeword $u_c^{n'}$ produces an output $y^{n'}$.

  $y^{n'}$ is the realization of some random variable $Y^{n'}$ which is got by the action of the channel c on $U_c^{n'}$. Note that $Y^{n'}$ will be different for different $c \in \mathscr{C}_{\mathscr{A}}$. Assume that some particular $c \in \mathscr{C}_{\mathscr{A}}$ happens, and $Y^{n'}$ is the corresponding channel output random variable. Our argument will hold for all $c \in \mathscr{C}_{\mathscr{A}}$.

  $y^{n'}$ depends on $u_c^{n'}$.

  By the codebook generation, the codewords are generated independently of each other, and there is no dependence between $y^{n'}$ and $u'^{n'}_1, u'^{n'}_2, u'^{n'}_{2^{\lfloor n'R \rfloor}-1}$. That is, $y^{n'}$, and $Y^{n'}$ are independent of $U'^{n'}_i$, $1 \le i \le 2^{\lfloor n'R \rfloor} - 1$.

- *Notation for the source coding problem:* We will do the analysis assuming that a particular $u^{n'} \in \mathcal{U}^{n'}$ needs to be coded.

  The source is the uniform X source. Thus, $u^{n'}$ is a realization of $U^{n'}$ where $U^{n'}$ has uniform distribution on $\mathcal{U}^{n'}$.

The codebook is

$$\mathscr{L}^{n'} = \{y_1^{n'}, y_2^{n'}, \ldots, y_{2^{\lfloor n'R \rfloor}}^{n'}\} \tag{2.46}$$

For all i, $y_i^{n'}$ is a realization of the random variable $V_i^{n'}$. By the random codebook generation, $V_i^{n'}$ is the uniform distribution on the set $\mathscr{U}_q^{n'} \subset \mathscr{Y}^{n'}$ of all sequences with precise type q.

By the random code book generation, the codewords are generated independently of each other, and thus, $V_i^{n'}$, $1 \le i \le 2^{\lfloor n'R \rfloor}$ are independent of each other as random variables.

Also, the codewords are of course, independent of the source sequence, and thus, $u^{n'}$ and $U^{n'}$ are independent of $V_i^{n'}$, $1 \le i \le 2^{\lfloor n'R \rfloor}$.

Analysis:

- *Error analysis for the channel coding problem:* We analyze the probability of *correct* decoding.

  We analyze the probability that a message is correctly received given that a particular message is transmitted. Think of some probability distribution $M^{n'}$ on the message set $\mathscr{M}_R^{n'}$. This probability distribution will *not* matter for the calculation. In fact, the calculation that we do can be done even if there is no probability distribution on the set of messages. We calculate

$$\Pr(\hat{M}_R^{n'} = M_R^{n'} | M^{n'} = m_i^{n'}) \text{ where } m_i^{n'} \in \mathscr{M}_R^{n'} \tag{2.47}$$

  The code book generation is symmetric. For this reason, the above probability will be independent of the particular message $m_i^{n'} \in \mathscr{M}_R^{n'}$.

  Also, $M^{n'}$ will depend on the particular $k \in \mathscr{A}$. We will get a bound for

$$\Pr(\hat{M}_R^{n'} = M_R^{n'} | M^{n'} = m_i^{n'}) \tag{2.48}$$

  which is independent of the particular $k \in \mathscr{A}$.

  From the decoding rule, it follows that for correct decoding, the following should happen:

  -
$$\frac{1}{n'} d^{n'}(u_c^{n'}, y^{n'}) \le D \tag{2.49}$$

  -
$$\frac{1}{n'} d^{n'}(u_i'^{n'}, y^{n'}) > D, 1 \le i \le 2^{\lfloor n'R \rfloor} - 1 \tag{2.50}$$

Thus, the event of correct decoding is:

$$\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \leq D\right\} \cap \cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\} \tag{2.51}$$

- *Error analysis for the source coding problem:* We analyze the probability of *error*.

  The analysis is done assuming that a particular sequence $u^{n'} \in \mathcal{U}^{n'}$ needs to be source coded. As we shall see, this error is independent of the particular source sequence because of the same empirical distribution of the source sequences, the symmetric nature of the code book construction, and permutation invariant distortion measure.

  An error happens if there exists no $y^{n'}$ in the code book $\mathcal{L}^{n'}$ such that $(u^{n'},y^{n'})$ are jointly typical, that is, an error happens if

$$\frac{1}{n'}d^{n'}(u^{n'},y^{n'}) > D \forall y^{n'} \in \mathcal{L}^{n'} \tag{2.52}$$

  The event of error is

$$\cap_{i=1}^{2^{\lfloor n'R\rfloor}}\left\{\frac{1}{n'}d^{n'}(u^{n'},V_i^{n'}) > D\right\} \tag{2.53}$$

  Note 2.38. Note that in the channel coding problem, we analyze the probability of correct decoding and in the source coding problem we analyze the probability of error

  Calculation:

- *Calculation of probability of correct decoding for the channel coding problem:*

  The correct decoding event is:

$$\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \leq D\right\} \cap \cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\} \tag{2.54}$$

  We wish to calculate the probability of the above event.

$$\begin{aligned}
&\Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \leq D\right\} \cap \cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)\\
=&\Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \leq D\right\}\right) + \Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right) -\\
&\Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'}) \leq D\right\} \cup \cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)\\
\geq&(1-\omega_{n'}) + \Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right) - 1\\
=&-\omega_{n'} + \Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i,Y^{n'}) > D\right\}\right)
\end{aligned}$$

$$= -\omega_{n'} + \prod_{i=1}^{2^{\lfloor n'R \rfloor}-1} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U'^{n'}_i, Y^{n'}) > D\right\}\right)$$

(since $U'^{n'}_i, 1 \le i \le 2^{\lfloor n'R \rfloor} - 1$, $Y^{n'}$ are independent random variables)

$$= -\omega_{n'} + \left[\Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, Y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

(where $U^{n'}$ has the same distribution as $U'^{n'}_i$ and is independent of $Y^{n'}$)

$$= -\omega_{n'} + \left[\sum_{y^{n'} \in \mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'}) \Pr\left(\frac{1}{n'} d^{n'}(U^{n'}, Y^{n'}) > D \mid Y^{n'} = y^{n'}\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

$$= -\omega_{n'} + \left[\sum_{y^{n'} \in \mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'}) \Pr\left(\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D \mid Y^{n'} = y^{n'}\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

$$= -\omega_{n'} + \left[\sum_{y^{n'} \in \mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'}) \Pr\left(\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

(since $U^{n'}$ and $Y^{n'}$ are independent)

$$\ge -\omega_{n'} + \left[\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1} \tag{2.55}$$

Rate R is achievable if

$$-\omega_{n'} + \left[\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1} \to 1 \text{ as } n' \to \infty \tag{2.56}$$

It is known that $\omega_{n'} \to 0$ as $n' \to \infty$. It follows that rate R is achievable if

$$\left[\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1} \to 1 \text{ as } n' \to \infty \tag{2.57}$$

- *Calculation of probability of error for the source coding problem:*

  The error event is:

$$\cap_{i=1}^{2^{\lfloor n'R \rfloor}} \left\{\frac{1}{n'} d^{n'}(u^{n'}, V_i^{n'}) > D\right\} \tag{2.58}$$

We wish to calculate the probability of this event.

$$\Pr\left(\cap_{i=1}^{2^{\lfloor n'R \rfloor}} \left\{\frac{1}{n'} d^{n'}(u^{n'}, V_i^{n'}) > D\right\}\right) \tag{2.59}$$

$$= \prod_{i=1}^{2^{\lfloor n'R \rfloor}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V_i^{n'}) > D\right\}\right) = \left[\Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}} \tag{2.60}$$

where $V^{n'}$ is a random variable which is uniformly distributed on $\mathcal{U}_q^{n'}$ and is independent of $u^{n'}$ for all $u^{n'} \in \mathcal{U}^{n'}$.

The type q with which the codewords are generated can be chosen by us. For block length n', we can choose the best possible achievable q for which the above error probability is the minimum. Let the set of all possible achievable types q for block length n' be denoted by $\mathcal{G}^{n'}$. The least possible error probability is given by

$$\left[\inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}} \tag{2.61}$$

To show the above dependence of the distribution of $V^{n'}$ on q, we denote it by $V_q^{n'}$. Thus, the least possible error probability is

$$\left[\inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}} \tag{2.62}$$

Since we are using the inf probability of excess distortion criterion, it t follows that rate R is achievable if

$$\left[\inf_{q \in \mathcal{G}^{n'_i}} \Pr\left(\left\{\frac{1}{n'_i} d^{n'_i}(u^{n'_i}, V_q^{n'_i}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}} \to 0 \text{ for some } n'_i = n_0 n_i \text{ for some } n_i \to \infty \tag{2.63}$$

Connection between channel coding and source coding:

It turns out that the main calculation we need to do in the channel coding problem is

$$\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) \tag{2.64}$$

and the main calculation we need to do in the source coding problem is

$$\inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) \tag{2.65}$$

We will prove that the above two expressions are equal.

We will prove more generally, that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'})>D\right\}\right),\ \text{if } y^{n'} \text{ has type } q \tag{2.66}$$

Let $y^{n'}$ have type q.

First we prove for the channel coding problem that if $y^{n'}$ and $y'^{n'}$ have the same type q, then

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y'^{n'})>D\right\}\right) \tag{2.67}$$

Since $U^{n'}$ is the uniform distribution on $\mathcal{U}^{n'}$, it follows that it is sufficient to prove that the cardinalities of the sets

$$\left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y^{n'})>D\right\} \text{ and } \left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y'^{n'})>D\right\} \tag{2.68}$$

are equal

Since $y^{n'}$ and $y'^{n'}$ have the same type, $y'^{n'}$ is a permutation of $y^{n'}$. Let $y'^{n'}=\pi^{n'}y^{n'}$.

Denote the sets

$$\mathcal{B}_{y^{n'}} \triangleq \left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y^{n'})>D\right\} \tag{2.69}$$

and

$$\mathcal{B}_{y'^{n'}} \triangleq \left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y'^{n'})>D\right\} \tag{2.70}$$

Let $u^{n'}\in\mathcal{B}_{y^{n'}}$. Since the distortion measure is permutation invariant, $d^{n'}(\pi^{n'}u^{n'},\pi^{n'}y^{n'})=d^{n'}(u^{n'},y^{n'})$. Thus, $\pi^{n'}u^{n'}\in\mathcal{B}_{y'^{n'}}$. If $u^{n'}\neq u'^{n'}$, $\pi^{n'}u^{n'}\neq\pi^{n'}u'^{n'}$. It follows that $|\mathcal{B}_{y'^{n'}}|\geq|\mathcal{B}_{y^{n'}}|$. $y^{n'}$ and $y'^{n'}$ in the above argument can be interchanged. Thus, $|\mathcal{B}_{y^{n'}}|\geq|\mathcal{B}_{y'^{n'}}|$. It follows that $|\mathcal{B}_{y^{n'}}|=|\mathcal{B}_{y'^{n'}}|$. Thus, it follows that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y'^{n'})>D\right\}\right) \tag{2.71}$$

$V_q^{n'}$ denotes the uniform random variable on the set of all sequences of all type q. Let V_q^n be independent of $U^{n'}$. It follows, by use of 2.71 that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},V_q^{n'})>D\right\}\right) \tag{2.72}$$

Next, we prove for the source-coding problem that if $u^{n'}, u'^{n'} \in \mathcal{U}^{n'}$ (in particular, they have the same type), then

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u'^{n'}, V_q^{n'}) > D\right\}\right) \tag{2.73}$$

Since $V_q^{n'}$ is the uniform distribution on the set of sequences $\mathcal{U}_q^{n'}$ of type q, it follows that it is sufficient to prove that the cardinalities of the sets

$$\left\{y^{n'} : \frac{1}{n'}d^{n'}(u^{n'}, y^{n'}) > D\right\} \text{ and } \left\{y^{n'} : \frac{1}{n'}d^{n'}(u'^{n'}, y^{n'}) > D\right\} \tag{2.74}$$

are equal.

Since $u^{n'}$ and $u'^{n'}$ belong to the set $\mathcal{U}^{n'}$, $u'^{n'}$ is a permutation of $u^{n'}$. Let $u'^{n'} = \pi^{n'} u^{n'}$.

Denote the sets

$$\mathcal{D}_{u^{n'}} \triangleq \left\{y^{n'} : \frac{1}{n'}d^{n'}(u^{n'}, y^{n'}) > D\right\} \tag{2.75}$$

and

$$\mathcal{D}_{u'^{n'}} \triangleq \left\{y^{n'} : \frac{1}{n'}d^{n'}(u'^{n'}, y^{n'}) > D\right\} \tag{2.76}$$

Let $y^{n'} \in \mathcal{D}_{y^{n'}}$. Since the distortion measure is permutation invariant, $d^{n'}(\pi^{n'} u^{n'}, \pi^{n'} y^{n'}) = d^{n'}(u^{n'}, y^{n'})$. Thus, $pi^{n'} y^{n'} \in \mathcal{D}_{u'^{n'}}$. If $y^{n'} \neq y'^{n'}$, $\pi^{n'} y^{n'} \neq \pi^{n'} y'^{n'}$. It follows that $|\mathcal{D}_{u'^{n'}}| \geq |\mathcal{D}_{u^{n'}}|$. $u^{n'}$ and $u'^{n'}$ in the above argument can be interchanged. Thus, $|\mathcal{D}_{u^{n'}}| \geq |\mathcal{D}_{u'^{n'}}|$. It follows that $|\mathcal{D}_{u'^{n'}}| = |\mathcal{D}_{u^{n'}}|$. Thus, it follows that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u'^{n'}, V_q^{n'}) > D\right\}\right) \tag{2.77}$$

$U^{n'}$ denotes the uniform random variable on $\mathcal{U}^{n'}$. Let $U^{n'}$ be independent of $V_q^{n'}$. It follows from 2.77 that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, V_q^{n'}) > D\right\}\right) \tag{2.78}$$

From (2.73) and (2.78), it follows that if $y^{n'}$ has type q,

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) \tag{2.79}$$

It follows that

$$\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) = \inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) \quad (2.80)$$

This proves what we had set out to prove in the connection between source and channel coding.

Denote

$$F^{n'} \triangleq \inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) = \inf_{q \in \mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'} d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) \quad (2.81)$$

Relation between the universal capacity of the partially known channel $c \in \mathcal{C}_{\mathcal{A}}$ and the rate-distortion function $R_U^P(D, \inf)$

- *Channel coding problem:* From (2.57), it follows that rate R is achievable if

$$[F^{n'}]^{2^{\lfloor n'R \rfloor} - 1} \to 1 \text{ as } n' \to \infty \quad (2.82)$$

- *Source coding problem:* From (2.63), it follows that rate R is achievable if

$$[F^{n_i'}]^{2^{\lfloor n_i' R \rfloor}} \to 0 \text{ as } n_i' \to \infty \text{ for some } n_i' = n_0 n_i \text{ for some } n_i \to \infty \quad (2.83)$$

If rate R is achievable for the channel-coding problem, so is any rate $< R$. Define:

$$\alpha \triangleq \sup\{R | \text{rate } R \text{ is achievable for the channel coding problem by use of the above random-coding method}\} \quad (2.84)$$

Then,

$$\lim_{n_i \to \infty} (F^{n_i})^{2^{\lfloor n_i R' \rfloor} - 1} < 1 \ \forall R' > \alpha \text{ for some sequence } n_i \to \infty \quad (2.85)$$

Thus,

$$\lim_{n_i \to \infty} (F^{n_i})^{2^{\lfloor n_i R'' \rfloor} - 1} = 0 \text{ for } R'' > R' \quad (2.86)$$

Note that $R'' > R' > \alpha$, but other than that, R' and R'' are arbitrary. It follows that rates $\leq \alpha$ are achievable for the source coding problem.

Note that the above random-coding method is just one possible method to generate codes for the channel coding problem. In general, it is possible that there exists another coding method which performs better than the above random-coding method, that is, for which rates $> \alpha$ are achievable for the channel coding problem. Thus, what we can claim from the above argument is that rates $< \alpha$ are achievable for the channel-coding problem. Thus, $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq \alpha$. Similarly, the above random-coding method is just one possible method to generate codes for the source coding problem. In general, it is possible that there exists another coding method which performs better than the above random-coding method, that is, for which rates $< \alpha$ are achievable for the source-coding problem when we use the probability of excess distortion criterion with the inf definition. That is, $R_U^P(D, \inf) \leq \alpha$. Thus, $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq \alpha$ and $R_U^P(D, \inf) \leq \alpha$. In particular, $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq R_U^P(D, \inf)$.

We have assumed that $R_U^P(D, \inf) = R_U^P(D)$, and thus, $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq R_U^P(D)$.

Thus, by use of encoder-decoder $< E^{n'}, F^{n'} >_1^\infty$, rate R is universally and reliably achievable over the unknown channel $c \in \mathscr{C}_{\mathscr{A}}$ if $R < R_U^P(D)$. Recall that $c \in \{< e^{n'} \circ k^{n'} \circ f^{n'} >_1^\infty \mid < k^{n'} >_1^\infty \in \mathscr{A}\}$. It follows that by use of encoder $< E^{n'} \circ e^{n'} >_1^\infty$ and decoder $< f^{n'} \circ F^{n'} >_1^\infty$, rates $R < R_U^P(D)$ are universally and reliably achievable over the partially known channel $k \in \mathscr{A}$.

Next, we want to see the resource consumption of the architecture for reliable communication.

Let the block length be n'. The original architecture consists of the encoder $e^{n'}$, the partially known channel k and decoder $f^{n'}$. With the input, the uniform X source $U^{n'}$, in the limit, the uniform X source is communicated to within a distortion D universally over the partially known channel k. With input $U^{n'}$ to the encoder $e^{n'}$, let the distribution of the channel input be denoted by $I^{n'}$. In the new architecture, encoder $E^{n'}$ and decoder $F^{n'}$ are built "on top of" the already existing architecture in order to communicate the message source $M^{n'}$ universally and reliably over the channel. The encoder $E^{n'}$ generates codewords with the same distribution as the source $U^{n'}$. This is because, the codewords are generated independently and uniformly from the set $\mathscr{U}^{n'}$. Let this random variable be denoted by $U^{s,n'}$. The superscript "s" should be thought of as "simulated". It follows that in this new architecture for reliable communication, the input to the channel will be some random variable $I^{s,n'}$ which has *the same distribution* as $I^{n'}$. From comments in Subsections 2.6.7 and 2.7.6, it follows that the new architecture, consisting of the encoder $< E^{n'} \circ e^{n'} >_1^\infty$, partially known channel k, and decoder $< f^{n'} \circ F^{n'} >_1^\infty$, when used to communicate the message source $< M^{n'} >_1^\infty$, consumes the same system resources as when the original architecture consisting of the encoder $< e^{n'} >_1^\infty$, the partially known channel k, and decoder $< f^{n'} >_1^\infty$ is used to communicate the uniform X source. See Figure 2.8. The common randomness input exists but has been omitted in the figure.

This finishes Step 1. We use this to prove Step 2.

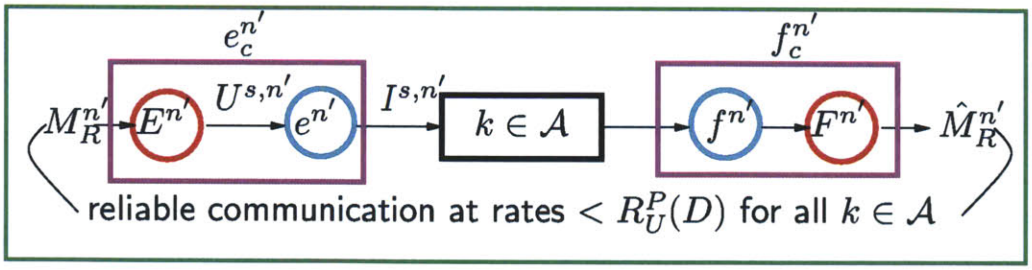


Figure 2.8. Universal reliable communication over the partially known channel $k \in \mathcal{A}$

Proof of Step 2 in order to prove Theorem 2.2

Let the partially known channel k be capable of universally communicating the uniform X source to within a distortion D. This is accomplished with the help of an encoder-decoder $< e^{n'}, f^{n'} >_1^\infty$. From the above argument of Step 1, it follows that with the help of encoder $< e_c^{n'} >_1^\infty = < E^{n'} \circ e^{n'} >_1^\infty$ and decoder $< f_c^{n'} >_1^\infty = < f^{n'} \circ F^{n'} >_1^\infty$, universal reliable communication can be accomplished over the channel k at rates $< R_U^P(D)$ by use of the same system resources as in the original architecture. In other words, the universal capacity of the partially known channel k is $\geq R_U^P(D)$.

Assume that the universal capacity of k is strictly greater than $R_U^P(D)$. It now follows that by source-compression followed by universal reliable communication, the uniform X source can be communicated universally over the partially known channel k to within a distortion D. A rough argument is the following: Take the uniform X source. Compress it using a source-encoder $< e_s^{n'} >_1^\infty$ to within a probability of excess distortion D. The output is a rate $R_U^P(D)$ message source (this is not entirely precise and we are omitting some ϵs and δs). This rate $R_U^P(D)$ message source can now be communicated universally and reliably over the partially known channel k with the help of channel encoder $< e_c^{n'} >_1^\infty = < E^{n'} \circ e^{n'} >_1^\infty$ and channel decoder $< f_c^{n'} >_1^\infty = < F^{n'} \circ f^{n'} >_1^\infty$. Finally, the output of the channel-decoder is source-decoded using decoder $< f_s^{n'} >_1^\infty$. End to end, the uniform X source is universally communicated to within a distortion D over the partially known channel k, digitally. The input to the channel has distribution $I^{s,n'}$ when block length is n' as described in Step 1 and thus, this source-channel based scheme consumes the same system resources.

A precise argument is the following:

We said above that the universal capacity of the partially known channel k is $\geq R_U^P(D)$. Assume that the universal capacity is strictly $> R_U^P(D)$. Let the universal capacity be $R_U^P(D) + \delta, \delta > 0$.

Let $\epsilon = \frac{\delta}{2}$. By the definition of $R_U^P(D)$, it follows that there exists a rate $R_U^P(D) + \epsilon$ source code $< e_s^{n'}, f_s^{n'} >_1^\infty$ which compresses the uniform X source to within a probability of excess distortion D.

Let the block length be n'.

The action of $e_s^{n'}$ on $U^{n'}$ produces an output random variable $M_R^{n'}$ on the set $\mathscr{M}_R^{n'}$. The set $\mathscr{M}_R^{n'}$ is

$$\mathscr{M}_R^{n'} = \{1,2,\ldots,2^{\lfloor n'(R_U^P(D)+\epsilon)\rfloor}\} \tag{2.87}$$

Since the universal capacity of $k \in \mathscr{A}$ is strictly greater than $R_U^P(D)+\epsilon$ by assumption, the message $M_R^{n'}$ can be universally and reliably communicated over the partially known channel k in the limit as $n \to \infty$. Finally, the source is re-constructed by using the source decoder $f_s^{n'}$.

See Figure 2.9. The common randomness input exists but it has been omitted in the figure, but it exists.

For every $\epsilon > 0$, $\exists n'_\epsilon$ such that when the block length is $> n'_{\epsilon_{n'}}$, for all $k \in \mathscr{A}$, $\Pr(\hat{M}_R^{n'} \neq M_R^{n'}) \leq \epsilon$. It follows that

$$\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D\right) < \omega_{n'} + \epsilon, \text{ if } n' > n'_\epsilon, \forall k \in \mathscr{A} \tag{2.88}$$

$\epsilon > 0$ is arbitrary, and thus, it follows that end-to-end, in this separation based architecture, the uniform X source is communicated universally and reliably to within a distortion level D over the partially known channel k.

The input to the channel has distribution $I^{s,n}$ when block length is n as described in Step 1 and thus, this source-channel based scheme consumes the same system resources like energy and bandwidth.

This completes the argument.

Note that we assumed that the universal capacity of the partially known channel k is strictly $> R_U^P(D)$, whereas from Step 1, it only follows that the universal capacity of the partially known channel k is $\geq R_U^P(D)$. It is unclear what will happen if the capacity of the partially known channel k is precisely $R_U^P(D)$. This "tension" of what happens if the capacity is precisely $R_U^P(D)$ is usual in information theory.

□

Note 2.39 (Time delays). Note that the definition of capability of universal communication of the uniform X source over a set of channels $\mathscr{A}$ requires the existence of *some* sequence $\omega =< \omega_{n'} >_1^\infty$, $\omega_{n'} \to 0$ as $n' \to \infty$, such that end-to-end, over the composition of the encoder, channel and decoder,

$$\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D\right) \leq \omega_{n'} \forall k \in \mathscr{A} \tag{2.89}$$

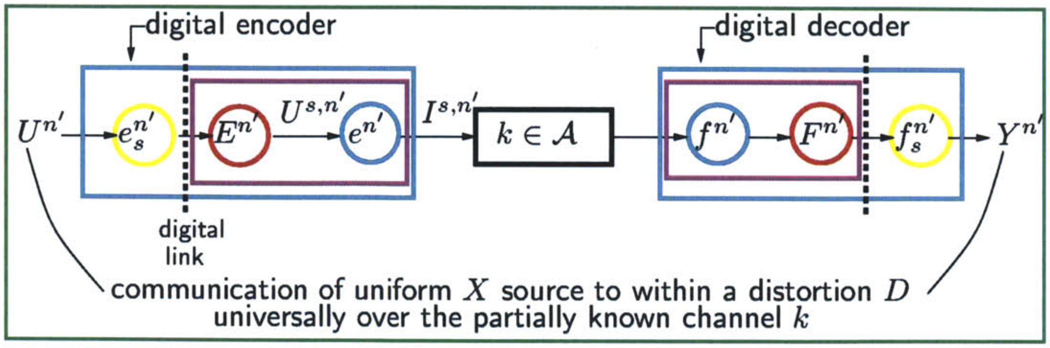


Figure 2.9. Converting an arbitrary architecture for communicating the uniform X source to within a distortion D universally over the partially known k, into a digital architecture

In the separation based scheme for communication of i.i.d. X source to within a distortion level D over the set of channels $\mathscr{A}$,

$$\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'})>D\right)\leq \omega_{n'}+\epsilon\forall n'>n'_\epsilon \tag{2.90}$$

Note that $\epsilon > 0$ can be chosen arbitrarily. Thus, the error sequence is some other $\omega' =< \omega'^{n'} >_1^\infty$. On a physical level, this translates into saying that for a particular probability of excess distortion requirement, the delay (the block length) required might be different, and in particular, larger in the digital separation architecture compared to the analog architecture. For this reason, we require the assumption stated in Section 2.2, that delays do not matter. As we said in Section 2.2. if delays did concern us, separation *does not* hold.

2.14.10 Discussions

A note on the proof technique and a comparison with Shannon's proof: using achievability techniques to prove a converse

Our proof, as the usual proofs of source-channel separation go, consist of two steps which have been stated before and are stated incompletely here:

1. If there exists some scheme in order to communicate the uniform X source universally over the partially known channel k to within a distortion D under the probability of excess distortion criterion, then the universal capacity of k is $\geq R_U^P(D)$
2. If the universal capacity of a partially known channel k is $> R_U^P(D)$, then universal communication of the uniform X source to within a probability of excess distortion D can be accomplished over the channel k by source compression followed universal reliable communication

We need to make sure that the system resource consumption is the same in both steps. From these two steps, the universal source-channel separation theorem follows.

These two steps traditionally are called converse and achievability, respectively. Step 1 is called converse because its traditional proof due to Shannon [Sha59] uses the usual converse techniques of equalities and inequalities related entropies and mutual informations. This proof of Shannon is discussed in brief in Subsection 5.7.4. Step 2 is usually achievability. Also, usually, Step 2 is Step 1, and Step 1 is Step 2. However, we will stick to our ordering.

For us, both Step 1 and Step 2 are achievability. Step 1 is achievability for us because we demonstrate a coding scheme with which communication at rates $< R_U^P(D)$ is possible over the partially known channel k. We use a traditional random-coding argument for Step 1. *We are thus using achievability methods to prove a result which is traditionally viewed as converse. In my opinion, this lends more insight into the nature of separation.*

Further comparison of Shannon's and our proof is made in Subsection 5.7.4.

A further note on the proof technique: layering "on top of" the original architecture

Note that given an original analog scheme consisting of encoder $< e^{n'} >_1^\infty$ and decoder $< f^{n'} >_1^\infty$ for universal communication to within a distortion D over the partially known channel k, the digital scheme that we construct is built "on top of" this scheme. The digital encoder consists of the source encoder $< e_s^{n'} >_1^\infty$, the channel encoder $< e_c^{n'} >_1^\infty = < E^{n'} \circ e^{n'} >_1^\infty$. Note that the channel encoder $< E^{n'} \circ e^{n'} >_1^\infty$ is layered on top of the original encoder $< e^{n'} >_1^\infty$. The digital decoder consists of the channel decoder $< f_c^{n'} >_1^\infty = < f^{n'} \circ F^{n'} >_1^\infty$ and the source decoder $< f_s^{n'} >_1^\infty$. Note that the channel decoder $< f^{n'} \circ F^{n'} >_1^\infty$ is layered on top of the original decoder $< f^{n'} >_1^\infty$. The original scheme is thus, converted into a digital scheme by introduction of a digital link. This also lends insight into how separation holds

Also, this should also be thought of as a proof technique, and we are illustrating just one possible digital scheme. There might be other ways of constructing digital schemes which accomplish the same goal which might not be layering on top of the original scheme.

Note, further the layered input-output view we have taken. The composite abstract channel $< e^{n'} \circ k \circ f^{n'} >_1^\infty$ *can be thought of as a* black-box. *In the first step, we are building additional encoders to accomplish universal reliable communication over this black-box at rates* $< R_U^P(D)$. *In this sense, we see a relationship between two major constructs of information, namely, capacity and rate-distortion function, and to how they are related in this black-box sense.*

A note on the connection between source coding and channel coding

The main step, in our opinion, in proving Steps 1 and 2 is noting that equality (2.66) is true. This equality is re-stated below:

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'})>D\right\}\right),\ \text{if } y^{n'} \text{ has type } q \quad (2.91)$$

This illustrates the mathematical duality between source and channel coding which finally leads to separation being true. We believe that this duality can be interpreted as a covering-packing duality; however, we are unsure. On a more intuitive level, the duality can also be seen in the parallel random-coding argument for the source-coding and the channel-coding problems. We discuss this further in Subsection 5.8.1

Why does separation hold?

In my opinion, fundamentally why separation holds boils down to (2.66), and is reproduced below:

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'})>D\right\}\right),\ \text{if } y^{n'} \text{ has type } q \quad (2.92)$$

This equation illustrates the connection between the problems of rate of reliable communication and the rate-distortion function, and why reliable communication at rates $< R_U^P(D)$ is possible over a partially known channel which is capable of communicating the uniform X source to within a distortion D. This helps in converting the original scheme into a digital scheme.

An operational perspective on the optimality of digital communication

The perspective on the optimality of digital communication in this section is operational: we use only the definition of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a certain distortion level. Unlike traditional proofs, for example, the one in [Sha59], we do not use simplified mathematical expressions, for example, mutual-information expressions for the channel capacity or the rate-distortion function. As we said, the proof in this section is not entirely precise. A precise proof, both for the uniform X source and the i.i.d. X source, are the subject of Chapter 5. We believe that our operational proof lends more insight into the nature of separation compared to the traditional proofs.

The operational nature of our proof is discussed in much more detail in Subsection 5.7.4: Chapter 5 is in fact devoted to an operational perspective on the optimality of digital communication for communication with a fidelity criterion.

2.14.11 A note on the technical assumption $R^P_U(D) = R^P_U(D, \inf)$

We have assumed that the distortion metric is permutation invariant and that, $R^P_U(D) = R^P_U(D, \inf)$. We do not know if $R^P_U(D) = R^P_U(D, \inf)$ is true for an arbitrary permutation invariant distortion metric. We would like to believe that it is *not* true for an arbitrary permutation invariant distortion metric. However, we would also like to believe that for most "well behaved" permutation invariant distortion metrics, $R^P_U(D) = R^P_U(D, \inf)$ should hold.

One kind of permutation invariant distortion metrics for which $R^P_U(D) = R^P_U(D, \inf)$ holds is additive distortion metrics, and this proof is carried out in Chapter 5.

2.15 A rigorous proof of the universal source-channel separation theorem for rate-distortion for i.i.d. X source and additive distortion measure

In this section, we provide a rigorous proof of the source-channel separation theorem for rate-distortion for the i.i.d. and the distortion measure is additive.

2.15.1 A statement of the universal source-channel separation theorem for rate-distortion for i.i.d. X source and additive distortion metric

Theorem 2.3 (Universal source-channel separation theorem for rate-distortion in the point-to-point setting for the i.i.d. X Source / optimality of digital communication for universal communication of the i.i.d. X source with a fidelity criterion). *Assuming random-coding is permitted, in order to communicate the i.i.d. X source over a partially known channel to within a particular distortion level under a an additive distortion metric, it is sufficient to consider source-channel separation based architectures, that is, architectures which first compress the i.i.d. X source to within the particular distortion level, followed universal reliable communication over the partially known channel. There is sufficiency in the sense if there exists some architecture to communicate the i.i.d. X source to within a certain distortion universally over the partially known channel, and which consumes certain amount of system resources (for example, energy and bandwidth), then there exists a separation based scheme to universally communicate the i.i.d. X to within the same distortion universally over the partially known channel and which consumes the same or lesser system resources as the original scheme.*

2.15.2 Steps to prove Theorem 2.3

There are two steps in the proof of Theorem 2.3 They are the same as the steps in the proof of Theorem 2.2 stated in Subsection 2.14.7 with the following changes:

- Replace the uniform X source with the i.i.d. X source
- Replace permutation invariant distortion metric with additive distortion metric

■ 2.15.3 The proof of Theorem 2.3

Proof. *Proof of Step 1 in order to prove Theorem 2.3*

Let $k =< k^n >_1^\infty \in \mathscr{A}$ be a partially known channel which is capable of universally communicating the i.i.d. X source to within a distortion D. Thus, there exist an encoder-decoder $< e^n, f^n >_1^\infty$ and a sequence $\omega =< \omega_n >_1^\infty$, $\omega_n \to 0$ as $n \to \infty$ such that the if the input to the composition of the encoder, channel and decoder $< e^n \circ k^n \circ f^n >_1^\infty$ is i.i.d. X source $X =< X^n >_1^\infty$, the output is $Y =< Y^n >_1^i nfty$ such that end to end,

$$\Pr\left(\frac{1}{n} d^n(X^n, Y^n) > D\right) < \omega_n \forall k \in \mathscr{A} \tag{2.93}$$

Consider the partially known abstract channel

$$c \in \{< e^n \circ k^n \circ f^n >_1^\infty | k =< k^n >_1^\infty \in \mathscr{A}\} \triangleq \mathscr{C}_{\mathscr{A}} \tag{2.94}$$

We will prove that the universal capacity of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$ is $\geq R_X^P(D)$, and this can be accomplished by an encoder-decoder $< E^n, F^n >_1^\infty$ such that for universal reliable communication at rates $R < R_X^P(D)$, the point-to-point communication system $< E^n \circ c^n \circ F^n >_1^\infty$ consumes the same system resources (irrespective of the particular $c \in \mathscr{A}$) as the original point-to-point communication system $< e^n \circ k^n \circ f^n >_1^\infty$ when used to communicate the i.i.d. X source universally to within a distortion D.

From this it will follow that the universal capacity of the partially known channel k is $\geq R_X^P(D)$, and this universal reliable communication at rates $< R_X^P(D)$: in fact, this communication can be accomplished with the help of encoder $< e_c^n >_1^\infty =< E^n \circ e^n >_1^\infty$ and decoder $< f_c^n >_1^\infty =< f^n \circ F^n >_1^\infty$. The point-to-point communication system $< e_c^n \circ k^n \circ f_c^n >_1^\infty$ when used for reliable communication at rates $< R_X^P(D)$ consumes the same system resources as the original point-to-point communication system $< e^n \circ k^n \circ f^n >_1^\infty$ when used to communicate the i.i.d. X source to within a distortion D.

We proceed to prove that the universal capacity of the partially known channel $c \in \mathscr{C}_{\mathscr{A}}$ is $\geq R_X^P(D)$.

We do this by use of a random-coding argument. First, we prove that $C_{rc}(\mathscr{C}_{\mathscr{A}})$ is $\geq R_X^I(D)$, where $R_X^I(D)$ is defined in Definition 2.30.

First, we recall some notation concerning the method of types.

Notation 2.9 (The type of x^n, p_{x^n}). Let $x^n \in \mathscr{X}^n$. p_{x^n} denotes the empirical distribution of x^n and is called the type of x^n. That is, for $x \in \mathscr{X}$,

$$p_{x^n}(x) \triangleq \frac{\text{number of } x \text{ in the sequence } x^n}{n} \tag{2.95}$$

Notation 2.10 (Typical set, $\mathcal{T}(p,\epsilon)$). Let p be a probability distribution on $\mathcal{X}$. Let $\epsilon \geq 0$. The sequence x^n is said to belong to $\mathcal{T}(p,\epsilon)$, and such an x^n is said to be ϵ p typical if

$$\sum_{x\in\mathcal{X}} |p_{x^n}(x) - p(x)| \leq \epsilon \tag{2.96}$$

Codebook generation: Generate $2^{\lfloor nR \rfloor}$ codewords i.i.d. p_X. This is the code book $\mathcal{K}^n$. As usual, the superscript n in $\mathcal{K}^n$ denotes the block length. This is the encoder E^n. Note that this is a random encoder

Note 2.40 (Note on code book generation method). For a channel $\in \mathcal{C}_{\mathcal{A}}$, a p_X typical sequence is distorted, with high probability, to within an average distortion D. In general, the behavior of the channel on a nonp_X typical sequence can be arbitrary. In other words, the belonging of a channel to $\mathcal{C}_{\mathcal{A}}$ is independent of the behavior of the channel on nonp_X typical sequences. When thinking of the action of $\mathcal{C}_{\mathcal{A}}$ as an attacker, the attacker can act on nonp_X typical sequences arbibrarily: for example, produce a random output sequence. For this reason, the codebook should contain sequences which are p_X typical, else, the attacker will *destroy* the sequence. For this reason, we use i.i.d. p_X source generation. Another way of thinking about i.i.d. p_X code book generation is the following: we only have knowledge of the behavior of the attacker when the source is i.i.d. X. Thus, it makes sense to use a code book which is i.i.d. X. The encoder can be thought of as *simulating* the i.i.d. X source.

Definition 2.36 (Joint typicality). (x^n, y^n), $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$ are said to be ϵ jointly typical if

1. x^n is ϵ p_X typical, that is, $x^n \in \mathcal{T}(p_X, \epsilon)$
2. $\frac{1}{n} d^n(x^n, y^n) \leq D$

Note 2.41 (A note on the definition of joint typicality). In information theory, typical set are defined in a way so that the set of typical sequences have high probability. Since the input to the channel is i.i.d. X, the definition of joint typicality requires that $p_{x^n} \in \mathcal{T}(p_X, \epsilon)$. When the input to the channel is i.i.d. X, with high probability, the average distortion between the channel input and output $\leq D$. For this reason, we require that $\frac{1}{n} d^n(x^n, y^n) \leq D$. We do not have any other information about the action of the channel. In the usual information theory literature, the channel is a discrete memoryless channel. In that case, the definition of joint typicality requires that the conditional type of the output y^n given the input x^n be *close to* the channel transition probability. However, our description of the channel is not in terms of its transition probability. Our description of the channel is in terms of the distortion that it produces on p_X typical sequences. Hence, the above definition of joint typicality.

Decoding: Let y^n be received as the output of the channel. If $\exists$ unique $x^n \in$ the code book $\mathcal{K}^n$ such that (x^n, y^n) ϵ-jointly typical, declare that x^n is transmitted, else declare error. This is the decoder F^n. Note that the encoder-decoder E^n, F^n is random.

Note 2.42 (A note on the decoding process). The decoding rule is the usual joint typicality decoding rule. It can potentially be thought of as a variant of minimum distance decoding.

Note that the encoding and decoding scheme are independent of the particular channel $\in \mathscr{C}_{\mathscr{A}}$.

In what follows,

- $x^n \in \mathscr{X}^n$ denotes the transmitted codeword. x^n is a realization of the random variable X^n.
- y^n denotes the received sequence (output of the channel). y^n is a realization of the random variable $\mathscr{Y}^n$. Y^n is the output of the channel when the input is X^n.
- $z^n \in \mathscr{X}^n$ denotes a codeword which is *not* transmitted. z^n is a realization of the random variable Z^n. Z^n is i.i.d. X and by the codebook construction, Z^n is independent of X^n. Since, Y^n is the output of the channel when the input is X^n, Z^n and Y^n are also independent.

Next, error analysis is carried out.

The error analysis is carried out given a particular message m^n needs to be communicated. That is, we calculate

$$\Pr(\text{error} \mid \text{message } m^n \text{ needs to be communicated}) = \Pr(\hat{M}_R^n \neq M_R^n \mid M_R^n = m^n) \tag{2.97}$$

Since the random code book is symmetric over all messages, this error probability is independent of the particular message m^n. It follows that

$$\Pr(\text{error}) = \Pr(\hat{M}_R^n \neq M_R^n) \tag{2.98}$$

is in fact equal to

$$\Pr(\hat{M}_R^n \neq M_R^n \mid M_R^n = m^n) \tag{2.99}$$

By the notation described above, x^n denotes the encoding of m^n. Since the code book is random, x^n is a realization of X^n. z^n denotes the encoding corresponding to another message m'^n.

The error events *given that the message m^n is transmitted* can be decomposed into two (not necessarily disjoint) events:

1. $\mathscr{E}_1^n$: (X^n, Y^n) is not ϵ jointly typical
2. $\mathscr{E}_2^n$: $\exists z^n$ such that (Z^n, Y^n) is ϵ jointly typical. That is,
   (a) $\exists z^n$ such that $p_{z^n} \in \mathscr{T}(p_X, \epsilon)$ and

(b) $\frac{1}{n}d^n(Z^n, Y^n) \leq D$

$$\Pr(\hat{M}_R^n \neq M_R^n \mid M_R^n = m^n) = \Pr(\mathcal{E}_1^n \cup \mathcal{E}_2^n) \leq \Pr(\mathcal{E}_1^n) + \Pr(\mathcal{E}_2^n) \tag{2.100}$$

Since, as discussed before,

$$\Pr(\hat{M}_R^n \neq M_R^n) = \Pr(\hat{M}_R^n \neq M_R^n \mid M_R^n = m^n), \tag{2.101}$$

it follows that

$$\Pr(\hat{M}_R^n \neq M_R^n) \leq \Pr(\mathcal{E}_1^n) + \Pr(\mathcal{E}_2^n) \tag{2.102}$$

Note that the encoding and decoding scheme is independent of the particular $c \in \mathscr{C}_{\mathscr{A}}$. However, the probabilities are not. We would like to make statements concerning probabilities which hold irrespective of the $c \in \mathscr{A}$. Thus, rate R is universally achievable over the partially known abstract channel $c \in \mathscr{C}_{\mathscr{A}}$ if

$$\Pr(\hat{M}_R^n \neq M_R^n) \to 0 \text{ as} n \to \infty \tag{2.103}$$

at a uniform rate over all channels $\in \mathscr{C}_{\mathscr{A}}$.

It follows that rate R is achievable if

$$\Pr(\mathcal{E}_1^n) \to 0 \text{ as } n \to \infty \tag{2.104}$$

at a uniform rate over all channels $\in \mathscr{C}_{\mathscr{A}}$, and

$$\Pr(\mathcal{E}_2^n) \to 0 \text{ as } n \to \infty \tag{2.105}$$

at a uniform rate over all channels $\in \mathscr{C}_{\mathscr{A}}$.

We now proceed to bound $\Pr(\mathcal{E}_1^n)$ and $\Pr(\mathcal{E}_2^n)$

By definition of a channel $\in \mathscr{C}_{\mathscr{A}}$ and by the way the code book is generated,

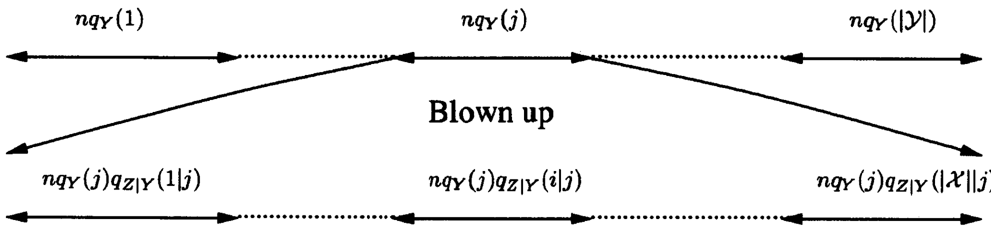
$$\Pr(\mathcal{E}_1^n) \leq \omega_n, \forall c \in \mathscr{C}_{\mathscr{A}} \tag{2.106}$$

$\Pr(\mathcal{E}_2^n)$ requires a more elaborate calculation. A bound on $\Pr(\mathcal{E}_2^n)$ is calculated, next. This is done by a method of types calculation, a-la [CK97].

For simplifying notation, let the sets $\mathscr{X}$ and $\mathscr{Y}$ be

$$\mathscr{X} = \{1, 2, \ldots, |\mathscr{X}|\}, \text{and}$$

Figure 2.10. The sorted received sequence y^n and the correspondingly shuffled codeword z^n illustrating the relevant types



$$\mathscr{Y} = \{1,2,\ldots,|\mathscr{Y}|\} \tag{2.107}$$

In what follows, it will be helpful to remember that $q(\cdot)$ and $q_{\cdot|\cdot}(\cdot|\cdot)$ will denote probability measures and transition probability kernels respectively of *observed* types (empirical distributions), and $p(\cdot)$ will denote probability measures of *transmitted* types.

Recall that the received sequence is y^n. Let the type of y^n be q_Y. That is, $\forall j \in \mathscr{Y}$, the number of j occurring in y^n is $nq_Y(j)$.

Sort the output y^n to place all the $j \in \mathscr{Y}$ together, and correspondingly shuffle the positions in the code book's codewords. This leads to no change in distortion between shuffled codewords and the sorted received sequence y^n, and thus, will not effect the analysis of $\Pr(\mathscr{E}_2^n)$. The sorting and shuffling is done, only to aid the reader in this error calculation: the shuffling will make it easier to give a pictorial representation, Figure 2.10, which is described, and alluded to, in the next paragraph.

Recall that z^n is a nontransmitted codeword. Over the chunk of length $nq_Y(j)$, let the type of the corresponding entries of z^n be $q_{X|Y}(\cdot|j)$. In other words, over the chunk of length $nq_Y(j)$, the number of i in z^n is $nq_Y(j)q_{Z|Y}(i|j)$. See Figure 2.10.

For now, the type of y^n, p_{y^n} is assumed to be q_Y. Later, we will take a bound over all possible q_Y.

For error event $\mathscr{E}_2^n$, (z^n,y^n) are ϵ jointly typical. Mathematically,

1. Denote $\sum_{j\in\mathscr{Y}} q_Y(j)q_{X|Y}(i|j)$ as $q_Z(i)$. q_Z is a probability distribution on $\mathscr{X}$. By the definition of joint typicality, z^n is ϵ p_X typical if

$$q_Z \in \mathscr{T}(p_X,\epsilon) \tag{2.108}$$

2. Denote $q_{ZY}(i,j) \triangleq q_Y(j)q_{Z|Y}(i|j)$. q_{ZY} is a probability distribution on $\mathscr{X}\times\mathscr{Y}$ resulting from the probability distribution $q_Y \in \mathscr{P}(\mathscr{Y})$ and the transition probability kernel

$q_{Z|Y} : \mathscr{Y} \to \mathscr{P}(\mathscr{X})$. Thus, $\frac{1}{n} d^n(z^n, y^n) \leq D$ can be restated as

$$\sum_{i \in \mathscr{X}, j \in \mathscr{Y}} q_{ZY}(i,j) d(i,j) \leq D \tag{2.109}$$

Recall that Z^n is generated i.i.d. p_X, and is independent of y^n. The probability that over the chunk of length $nq_Y(j)$, the corresponding entries of Z^n have type $q_{Z|Y}(\cdot|j)$ is

$$\leq 2^{-nq_Y(j) D(q_{Z|Y}(\cdot|j) || p_X)} \tag{2.110}$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence and defined for probability distributions p and q where $p, q \in \mathscr{P}(\mathscr{X})$ as

$$D(p||q) \triangleq \sum_{x \in \mathscr{X}} p(x) \log \frac{p(x)}{q(x)} \tag{2.111}$$

Note that we are using the same alphabet D for the distortion D and the Kullback-Leibler divergence $D(\cdot||\cdot)$. It will be clear from context, which one is being referred to.

Thus, the probability that over the whole block of length n, in the chunks $nq_Y(j)$, the corresponding entries of z^n have type $q_{X|Y}(\cdot|j)$ for all j

$$\leq \prod_{j \in \mathscr{Y}} 2^{-nq_Y(j) D(q_{Z|Y}(\cdot|j) || p_X)} \tag{2.112}$$

$$= 2^{-n \sum_{j \in \mathscr{Y}} q_Y(j) D(q_{Z|Y}(\cdot|j || p_X)} \tag{2.113}$$

$$= 2^{-nD(q_{ZY} || p_X q_Y)} \tag{2.114}$$

It would be helpful to note the positions of where p occur and where q occur, in the above expression.

To bound the probability that z^n is at a distortion $\leq D$ from y^n, the above probability needs to be summed over all possible types $q_{Z|Y}(\cdot|j), 1 \leq j \leq |\mathscr{Y}|$ such that (2.108) and (2.109) are satisfied.

The number of conditional types $q_{Z|Y}(\cdot|j)$ is $\leq (n+1)^{|\mathscr{X}||\mathscr{Y}|}$. Recall that number of nontransmitted codewords $< 2^{\lfloor nR \rfloor}$. Putting all this together and using the union bound,

$$\Pr(\mathscr{E}_2^n | p_{y^n} = q_Y) \tag{2.115}$$

$$\leq \sum_{q_{ZY} \in \mathscr{S}} 2^{\lfloor nR \rfloor} 2^{-nD(q_{ZY} || p_X q_Y)} \tag{2.116}$$

$$\leq (n+1)^{|\mathscr{X}||\mathscr{Y}|} 2^{\lfloor nR \rfloor} 2^{-n \inf_{q_{ZY} \in \mathscr{S}} D(q_{ZY} || p_X q_Y)} \tag{2.117}$$

where $\mathscr{S}$ denotes the set of types satisfying (2.108) and (2.109), along with the type of y^n, which, for now, has been fixed to q_Y:

$$\mathscr{S} = \left\{ q_{ZY} : \begin{array}{l} q_Z \in p_X \pm \epsilon \\ \sum_{i \in \mathscr{X}, j \in \mathscr{Y}} q_{ZY}(i,j) d(i,j) \le D \\ q_Y \text{ fixed} \end{array} \right\} \tag{2.118}$$

The type of the received sequence y^n is arbitrary. In other words, q_Y is arbitrary. Thus,

$$\Pr(\mathscr{E}_2^n) \le (n+1)^{|\mathscr{X}||\mathscr{Y}|} 2^{\lfloor nR \rfloor} 2^{-n \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} || p_X q_Y)} \tag{2.119}$$

where the set $\mathscr{T}$ is

$$\mathscr{T} = \left\{ q_{ZY} : \begin{array}{l} q_Z \in p_X \pm \epsilon \\ \sum_{i \in \mathscr{X}, j \in \mathscr{Y}} q_{ZY}(i,j) d(i,j) \le D \end{array} \right\} \tag{2.120}$$

The set $\mathscr{T}$ is the union of the $\mathscr{S}$ over all possible q_Y. Thus, in in the bound (2.119), q_Y is allowed to be arbitrary.

It follows that

$$\Pr(\mathscr{E}_1^n) + \Pr(\mathscr{E}_2^n) \le \omega_n + (n+1)^{|\mathscr{X}||\mathscr{Y}|} 2^{\lfloor nR \rfloor} 2^{-n \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} || p_X q_Y)} \tag{2.121}$$

From previous discussions, it follows that

$$\begin{aligned} \Pr(\text{error} = \Pr(\hat{M}_R^n \neq M_R^n) \le \Pr(\mathscr{E}_1^n) + \Pr(\mathscr{E}_2^n) \le \\ \omega_n + (n+1)^{|\mathscr{X}||\mathscr{Y}|} 2^{\lfloor nR \rfloor} 2^{-n \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} || p_X q_Y)} \end{aligned} \tag{2.122}$$

By definition, $\Pr(\mathscr{E}_1^n) = \omega_n \to 0$ as $n \to \infty$ and ω_n is independent of the particular channel $\in \mathscr{C}_{\mathscr{A}}$.

Note that the bound

$$(n+1)^{|\mathscr{X}||\mathscr{Y}|} 2^{\lfloor nR \rfloor} 2^{-n \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} || p_X q_Y)}, \tag{2.123}$$

on $\Pr(\mathscr{E}_2^n)$ is independent of the particular channel $\in \mathscr{C}_{\mathscr{A}}$.

Also, note that $(n+1)^{|\mathscr{X}|(|\mathscr{Y}|)}$ is a polynomial. Thus, $\Pr(\mathscr{E}_2^n) \to 0$ as $n \to \infty$ at a rate independent of the particular channel $\in \mathscr{C}_{\mathscr{A}}$ if

$$R < \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} || p_X q_Y) \tag{2.124}$$

It follows that Pr(error) $\to 0$ as $n \to \infty$ at a rate which is independent of the particular channel $\in \mathscr{C}_{\mathscr{A}}$ for rates

$$R < \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} || p_X q_Y) \tag{2.125}$$

(2.126)

Thus, rates $R < \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} \| p_X q_Y)$ are universally achievable over the set of channels $\mathscr{C}_{\mathscr{A}}$.

Note that

$$D(q_{ZY} \| p_X q_Y) = D(q_Z \| p_X) + D(q_{ZY} \| q_Z q_Y) \geq D(q_{ZY} \| q_Z q_Y) \tag{2.127}$$

Thus, all rates R for which

$$R < \inf_{q_{ZY} \in \mathscr{T}} D(q_{ZY} \| q_Z q_Y) = \inf_{q_{ZY} \in \mathscr{T}} I(Z;Y) \tag{2.128}$$

are universally achievable.

Recall (2.30), the definition of the information theoretic rate-distortion function, $R_X^I(D)$. It follows that

$$\inf_{q_{ZY} \in \mathscr{T}} I(Z;Y) = \inf_{Z \in \mathscr{T}(p_X, \epsilon)} R_Z^I(D) \tag{2.129}$$

Thus, rates

$$R < \inf_{Z \in \mathscr{T}(p_X, \epsilon)} R_Z^I(D) \tag{2.130}$$

are universally achievable.

$\epsilon > 0$ is arbitrary and the information theoretic rate-distortion function is continuous. It follows that rates $R < R_X^I(D)$ are achievable. In general, it is potentially possible that with the use of other encoding-decoding schemes, rates $\geq R_X^I(D)$ are pseudo universally achievable. Thus,

$$C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq R_X^I(D) \tag{2.131}$$

The goal is to prove that $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq R_X^P(D)$. To this end, we prove that $R_X^P(D) = R_X^I(D)$. In fact, we will also end up proving that $R_X^P(D) = R_X^E(D)$ and thus, in effect, we will prove that $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq R_X^P(D) = R_X^E(D)$.

We proceed to prove that $R_X^P(D) = R_X^E(D) = R_X^I(D)$

Proof that $R_X^E(D) = R_X^I(D)$ is there in [Sha59].

Next, note that $R_X^E(D) \leq R_X^P(D)$. This is because of the following:

The idea is that the probability of excess distortion criterion is "stronger" than the expected probability of error criterion. That is, if a particular probability of excess distortion level is achievable for some source, the same expected distortion is also achievable by the same source code for the same source. A rigorous argument is the following:

Define:

$$D_{\max} \triangleq \max_{x in \mathscr{X}, y \in \mathscr{Y}} d(x,y) \tag{2.132}$$

Let probability of excess distortion D be achievable with source code $s = < s^n >_1^\infty$. Then, for the source code s,

$$\Pr\left[\frac{1}{n} d^n(X^n, Y^n) > D\right] = \epsilon^n \to 0 \text{ as } n \to \infty \tag{2.133}$$

It follows from the above equation that

$$E\left[\frac{1}{n} d^n(X^n, Y^n)\right] \leq (1-\epsilon^n)D + \epsilon^n D_{\max} \to D \text{ as } n \to \infty \tag{2.134}$$

Thus, the expected distortion D is achievable by use of the same source code s, and in particular, by a source code of the same rate. It follows that $R_X^E(D) \leq R_X^P(D)$.

By interpreting the definitions in Chapter 2.2 in [CK97] properly, one can see that they are in fact using the probability of excess distortion criterion. From their proofs, and by some additional arguments, it follows that $R_X^P(D) \leq R_X^I(D)$. These additional arguments are related to taking lim inf instead of lim. We omit these arguments. They are similar in spirit to the ones in Chapter 5 .

We have proved or cited references where the following are proved:

1. $R_X^E(D) = R_X^I(D)$
2. $R_X^E(D) \leq R_X^P(D)$
3. $R_X^P(D) \leq R_X^I(D)$

It thus follows that $R_X^E(D) = R_X^P(D) = R_X^I(D)$.

Back to finishing Step 1, it follows that $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq R_X^I(D) = R_X^P(D) = R_X^E(D)$, and in particular, $C_{rc}(\mathscr{C}_{\mathscr{A}}) \geq R_X^P(D)$.

Thus, by use of encoder-decoder $< E^n, F^n >_1^\infty$, rate R is universally and reliably achievable over the unknown channel $c \in \mathscr{C}_{\mathscr{A}}$ if $R < R_U^P(D)$. Recall that $c \in \{< e^n \circ k^n \circ f^n >_1^\infty \mid < k^n >_1^\infty \in \mathscr{A}\}$. It follows that by use of encoder $< E^n \circ e^n >_1^\infty$ and decoder $< f^n \circ F^n >_1^\infty$, rates $R < R_X^P(D)$ are universally and reliably achievable over the partially known channel $k \in \mathscr{A}$.

Next, we want to see the resource consumption of the architecture for reliable communication.

Let the block length be n. The original architecture consists of the encoder e^n, the partially known channel k and decoder f^n. With the input, the uniform X source U^n, in the limit, the uniform X source is communicated to within a distortion D universally over the partially known channel k. With input X^n to the encoder e^n, let the distribution of the channel input be denoted by I^n. In the new architecture, encoder E^n and decoder F^n are built "on top of" the already existing architecture in order to communicate the message source M^n universally and reliably over the channel. The encoder E^n generates codewords with the same distribution as the source X^n. This is because, the codewords are generated independently and uniformly from the set $\mathscr{X}^n$. Let this random variable be denoted by $X^{s,n}$. The superscript "s" should be thought of as "simulated". It follows that in this new architecture for reliable communication, the input to the channel will be some random variable $I^{s,n}$ which has *the same distribution* as I^n. From the discussion in Subsection 2.6.7, it follows that the new architecture, consisting of the encoder $< E^n \circ e^n >_1^\infty$, partially known channel k, and decoder $< f^n \circ F^n >_1^\infty$, when used to communicate the message source $< M_R^n >_1^\infty$, consumes the same system resources as when the original architecture consisting of the encoder $< e^n >_1^\infty$, the partially known channel k, and decoder $< f^n >_1^\infty$ is used to communicate the uniform X source.

See Figure 2.8 with the uniform X source replaced with the i.i.d X source and the rate-distortion function $R_U^P(D)$ replaced with the rate-distortion function $R_X^P(D)$ for the i.i.d. X source.

This finishes Step 1. We use this to prove Step 2.

Proof of Step 2 to prove Theorem 2.3

Let the partially known channel k be capable of universally communicating the i.i.d. X source to within a distortion D. This is accomplished with the help of an encoder-decoder $< e^n, f^n >_1^\infty$. From the above argument of Step 1, it follows that with the help of encoder-decoder $< E^n \circ e^n, f^n \circ F^n >_1^\infty$, universal reliable communication can be accomplished over the partially known channel k at rates $< R_U^P(D)$ by use of the same system resources as in the original architecture. In other words, the universal capacity of the partially known channel $k \in \mathscr{A}$ is $\geq R_X^P(D)$.

It now follows that by source-compression followed by universal reliable communication, the i.i.d. X source can be communicated universally over the partially known channel k to within a distortion D. A rough argument is the following: Take the i.i.d. X source. Compress it using a source-encoder $< e_s^n >_1^\infty$ to within a probability of excess distortion D. The output is a rate $R_X^P(D)$ message source. This rate $R_X^P(D)$ message source can now be communicated universally and reliably over the partially known channel k with the help of channel encoder $< e_c^n >_1^\infty \triangleq < E^n \circ e^n >_1^\infty$ and channel decoder $< F^n \circ f^n >_1^\infty$. Finally, the output of the channel-decoder is source-decoded using decoder $< f_s^n >_1^\infty$. End to end, the i.i.d. X source is universally and communicated to within a distortion D over the partially known channel k, digitally.

A rigorous argument for the above source-coding followed by channel coding is the following:

We said above that the universal capacity of the partially known channel k is $\geq R_X^P(D)$. Assume that the universal capacity is strictly $> R_X^P(D)$. Let the universal capacity be $R_X^P(D) + \delta, \delta > 0$.

Let $\epsilon = \frac{\delta}{2}$. By the definition of $R_X^P(D)$, it follows that there exists a rate $R_X^P(D) + \epsilon$ source code $< e_s^n, f_s^n >_1^\infty$ which compresses the i.i.d. X source to within a probability of excess distortion D.

Let the block length be n.

The action of e_s^n on X^n produces an output random variable M_R^n on the set $\mathcal{M}_R^n$. The set $\mathcal{M}_R^n$ is

$$\mathcal{M}_R^n = \{1, 2, \ldots, 2^{\lfloor n(R_X^P(D)+\epsilon) \rfloor}\} \tag{2.135}$$

Since the universal capacity of $k \in \mathcal{A}$ is strictly greater than $R_X^P(D) + \epsilon$ by assumption, the message M_R^n can be universally and reliably communicated over the partially known channel k in the limit as $n \to \infty$. Finally, the source is re-constructed by using the source decoder f_s^n.

See Figure 2.9 with the uniform X source replaced with the i.i.d X source.

For every $\epsilon > 0$, $\exists n_\epsilon$ such that when the block length is $> n_{\epsilon_n}$, for all $k \in \mathcal{A}$, $\Pr(\hat{M}_R^n \neq M_R^n) \leq \epsilon$. It follows that

$$\Pr\left(\frac{1}{n} d^n(X^n, Y^n) > D\right) < \omega_n + \epsilon, \text{ if } n > n_\epsilon, \forall k \in \mathcal{A} \tag{2.136}$$

$\epsilon > 0$ is arbitrary, and thus, it follows that end-to-end, in this separation based architecture, the i.i.d. X source is communicated universally to within a distortion level D over the partially known channel k.

The input to the channel has distribution $I^{s,n}$ when block length is n as described in Step 1 and thus, this source-channel based scheme consumes the same system resources.

Note that we assumed that the universal capacity of the partially known channel k is strictly $> R_X^P(D)$, whereas from Step 1, it only follows that the universal capacity of the partially known channel k is $\geq R_X^P(D)$. It is unclear what will happen if the capacity of the partially known channel k is precisely $R_X^P(D)$. This "tension" of what happens if the capacity is precisely $R_X^P(D)$ is usual in information theory.

This completes the argument, and thus, rigorously proves the universal source-channel separation theorem for rate-distortion when the source is i.i.d. and the distortion metric is additive.

□

2.16 A universal source-channel separation theorem for rate-distortion for permutation invariant distortion measures: discussion and high-level view

We do not know how to generalize our results to arbitrary distortion measures.

However, we do know, on a high level, how to generalize our results to certain permutation invariant distortion measures. This follows because the proof of Section 2.14 holds for permutation invariant distortion measures, though it required a technical condition on the rate-distortion function .

Also, the proof calculations of Section 2.15 for the i.i.d. X source, to a large extent, hold for permutation invariant distortion measures. In particular, the method of types calculations works exactly, for permutation invariant distortion measures. This is because for permutation invariant distortion measures, rearranging the sequences x^n and y^n by the same rearrangement does not change the distortion. Still, some technical conditions might be required on the distortion function. We have not worked out these details.

I do not think the results can be generalized to arbitrary distortion measures (which are not necessarily permutation invariant).

High level ideas for generalization to stationary ergodic sources and permutation invariant distortion measures are discussed in Section 2.18.

2.17 Discussion: are random codes needed? And if yes, can random-coding be practically realized?

2.17.1 Are random codes needed?

In Shannon's random-coding argument, [Sha48], random-coding is a proof technique. Given a random code to achieve a particular rate of reliable communication, there exists a deterministic code to. For us, random-coding is *not* just a proof technique: random codes are in fact needed. The difference in our situation and the situation in [Sha48] is that we have a partially known channel whereas Shannon assumed a fully known channel.

On a high level, with the average probability of error criterion this happens for the following reason: given a random code which achieves a particular error, one of the deterministic codes that "make up" the random code will have an error which is less than or equal to that produced by the random code. It follows that restriction to deterministic codes is sufficient. For the maximum probability of error criterion that we use, the usual argument which we omit has to go through throwing away half the codewords (note that throwing away half the codewords does not change the rate).

An example can be provided for the case of partially known channel $k \in \mathscr{A}$ which is capable of communicating i.i.d. X source to within a distortion D, for which universal capacity with random-coding is $R_X^p(D) > 0$ but universal capacity without random-coding is zero. We omit this example here; but one can be constructed based on ideas in [AKM]. *From this example, it in fact follows that a universal source-channel separation theorem for rate-distortion does not hold if random-coding is not permitted.*

2.17.2 How can random codes be generated in practice?

In practice, perfect randomness is not needed. What is needed is pseudo randomness. Pseudo randomness is used, for example, in PN sequences in CDMA. Pseudo randomness can thus be generated with the help of a seed.

We believe that the requirement of random codes (or pseudo-random codes) is *not* a hinderance in practical implementation.

2.18 Discussion: Continuous time sources

The whole discussion in this chapter has rested on sources and channels evolving in discrete time. In our framework, continuous time evolution of channels is easy to deal with and this is one of the points discussed in the next section.

In this section, we discuss the problem of how does one deal with continuous time source evolution in this framework. This view also generalizes to general stationary ergodic sources (in fact, i.i.d. sources in continuous time do not make physical sense: the only way to model them would be as some kind of white noise).

One incorrect way which of thinking about generalizing to continuous time sources is the following:

Assume that the source is band-limited. The source can then be sampled by the sampling theorem or by using some other orthonormal expansion as discussed, for example, in Gallager's book [Gal08], the corresponding course notes [Gala] or the video lectures [Galb]. This makes the source discrete.

This approach works for the problem of reliable communication. However, this approach does not work for communication with a fidelity criterion. This is because if the distortion measure for the continuous source is additive (instead of a summation, this would be an integral) or permutation invariant, it is not necessary that the corresponding distortion measure for the sampled source is additive or permutation invariant. Unless the distortion measure is, for example, mean-squared distortion in which case, by Parseval's theorem, the distortion measure for the discretized source will also be additive. In general, this is not the case.

We believe that discretization procedures of the form of sampling will not work. For continuous time sources, we need to carry out arguments either in the continuous time domain, or discretize the time very finely and then take a limit as the discretization blocks become smaller and smaller. [PG77], for example, proves the source-coding theorem for compression with a fidelity criterion for quite general continuous time sources, and it might be possible, then, to use this result to prove a source-channel separation theorem for communication with a fidelity criterion.

We offer another approach for generalization to continuous time source evolution. We outline this approach, below.

The fundamental reason, why source-channel separation holds is, as we discussed before in Section 2.14, is that (2.66) holds, and this equation is reproduced below.

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'})>D\right\}\right),\ \text{if } y^{n'} \text{ has type } q \tag{2.137}$$

We want to see, how this will generalize to stationary ergodic sources, both in discrete time or continuous time. The discussion will be rough.

Assume that time runs from $-\infty$ to $+\infty$.

As Shannon says in [Sha48], "If an ensemble is ergodic we may say roughly that each function in the set is typical of the ensemble." Thus, all realizations of the ensemble look the same under time permutations. Whatever behavior happens in one realization in certain time interval, a similar behavior will be observed in all other realizations in different time intervals. In other words, after permutation, all realizations of the ensemble look the same. *This is not entirely true, but in spirit, it is.*

This is precisely the property needed for (2.137) to be true: the sequences $u^{n'} \in \mathcal{U}^{n'}$, are precisely the same under time permutations, and similarly, the sequences $y^{n'} \in \mathcal{U}_q^{n'}$ are also precisely the same under time permutations.

In the continuous time, stationary ergodic source case, still, the codewords can be generated from a stationary ergodic process. Thus, the equivalent of (2.137) will be true, the entire argument. Of course, this needs to be made precise, and things get complicated when the time is finite. We have *not* worked out the details ourselves. However, we do believe that this high level idea can be made precise.

■ 2.19 A discussion of the assumptions described in Section 2.2

In the light of our proof of the universal source-channel separation theorem for rate-distortion, we want to further comment on the assumptions of Section 2.2.

The assumptions in Section 2.2 were divided into various categories:

The basic assumptions of communications theory:

- *The assumption that delays do not matter:* We use this assumption crucially. In the digital scheme, the delay (or the block length) required for a particular error for the probability of excess distortion beyond D might be larger than in the original architecture. This is elaborated on, in Discussion 2.39. As stated in this discussion, if delays matter, digital architectures are not optimal

- *The assumptions that the source can be modeled as a stationary ergodic random process:* This is the usual assumption in information theory, and it is difficult to prove results for sources which are not stationary ergodic. Such results exist, however, for example, see [VVS95]. However, we do require this assumption. Further comments were made in Subsection 2.4.3, on this assumption

- *The assumption that the channel can be modeled as a partially known transition probability:* This was commented on, in Section 2.5.4. What is interesting, however, is that we do not require any further assumptions on the channel. The usual assumption made on the channel is memorylessness, Markoff nature, or more generally, in Shannon's words, channels for which "historical influences die away", a concept made precise by Gallager and called indecomposability in [Gal68]. We do not require any further assumptions on the channel because we use the probability of excess distortion criterion instead of the expected distortion criterion. The probability of excess distortion criterion can be thought of as forcefully enforcing a weak law of large numbers of condition, and that ends up, on an intuitive level, being the reason that we do not require any further assumptions on the nature of the transition probability of the channel

- *The source can incur distortion, and the distortion can be modeled as a distortion metric:* This is the usual assumption made in information theory.

The following crucial assumptions has been made on the nature of the distortion metric and the allowability of random-coding:

- *The assumption that the distortion metric can be modeled as a permutation invariant distortion metric, and that, the distortion criterion is the "probability of excess distortion" criterion:* Many comments have been made on the permutation invariant distortion metric in the previous sections. We will not repeat here. However, we wonder if our results can be generalized to more general distortion metrics, for example, sub-additive distortion metrics as is the case in some of the information theory literature, see for example, [Han10]. We do not know how to do this. Our proof technique of Section 2.14 seems to rely crucially on the use of a permutation invariant distortion metric. Also, to the best of our knowledge, there is no relation between permutation invariant distortion metrics and sub-additive distortion metrics; however, it would be good to understand this further.

- The use of the probability of excess distortion criterion is crucial in the sense that the universal source-channel separation theorem is not true if we use the expected distortion criterion instead of the probability of excess distortion criterion. An example can be constructed similar to the example in [VH94] on the first page of this paper where it says, "Consider a binary channel where the output codeword is equal to the transmitted codeword with probability $\frac{1}{2}$ ..."

Note that this is an example of a "highly nonergodic channel," and the only examples that we know of, belong to this category. If we impose ergodicity assumptions on the channel, we know that at least in certain cases, even under the expected distortion criterion, a universal source-channel separation theorem for rate-distortion holds. Some of these ideas are due to Amos Lapidoth.

This approach of Amos Lapidoth to proving a universal source-channel separation theorem for rate-distortion uses the traditional information theory tools of entropy and mutual information. We conjecture another possible way of attacking the same problem which is the following: given a partially known channel k which satisfy some ergodicity assumptions and is capable of communicating i.i.d. X source to within a distortion level D *under the expected distortion criterion*. Does this imply that the partially known channel k is also capable (with a possibly different encoder-decoder) of communicating the i.i.d. X source to within a distortion D under the probability of excess distortion criterion. We would like to believe that this is true. However, we are unsure of this. However, if this were true, then based on our results, it would follow that even if we used the expected distortion criterion and imposed some ergodicity assumptions on the channel, we could convert the problem into a problem with the probability of excess distortion criterion, and thus, universal source-channel separation would hold even under the expected distortion criterion. This is discussed further in Chapter 6.

- The assumption that random-coding is permitted: As we have stated before, if random-coding is not permitted, universal source-channel separation for rate-distortion is not true. This has been commented on, in Section 2.17.

The following assumption has been made on the knowledge of source statistics but we conjecture that this assumption can be removed:

- *The assumption that the source statistics are known*: As stated in Section 2.2, we conjecture that this assumption can be removed, but we are not sure. A possible way to remove this assumption is the following: Ziv proved in [Ziv72] that under certain alphabet assumptions, there exist universal algorithms for the class of all stationary sources that perform as well for each source as an optimum source code custom designed for each source, that is, reaches the rate-distortion limit for each source. The distortion definition used in this paper is the expected distortion criterion, and we believe the result should hold even with the probability of excess distortion criterion. Also, this is a source-coding result, and we are looking more for a separation result for communication over a channel. With the above source-coding result, though, we believe that we should be able to prove a universal source-channel separation theorem for rate-distortion where universality is both over the source and the channel. Possibly, [Ziv80] might be useful. We have not carried this argument out.

The following assumptions have been made to prove results rigorously and to avoid mathematical complications, and we are quite sure that modulo some technical assumptions, they can be removed:

- *The assumption that the source alphabet and source reproduction alphabet is finite:* We believe that argument can be carried out for some source alphabet which is not finite. For example, let the source alphabet be a finite length interval of $\mathscr{R}$. Then, a rigorous argument can be carried out by discretizing the alphabet into small intervals of size Δ and then, taking the limit as $\Delta \to 0$. If the source alphabet is the whole real line $\mathscr{R}$, and assuming that the source has sufficiently light tails in that probability outside finite length intervals falls off sufficiently fast, we can truncate the real line and then take limit to the whole real line. We believe that this argument can also be made rigorous though we have not carried out the steps.

- *The assumption that the source is i.i.d:* We believe that the simulation argument of Section 2.15 can be carried out for many non i .i.d. sources also. We have not carried out this argument. In particular, we believe that the simulation argument can be carried out for many stationary ergodic sources.

  For sources in which memory dies out with time, another way of carrying out a rigorous argument is the following: consider large blocks of the source k interspersed with smaller blocks of size k'. Over the block of size k', the source memory will die out so that the source will be "almost" independent over the blocks of length k. This argument needs to be carried out rigorously, and we believe it can be carried out. However, we have not done it.

  Another view of this was provided in the previous section where we justified, on a high-level, why the result should hold for stationary ergodic sources.

- *The assumption that the distortion metric is additive*: This has been commented on, in Sections 2.14 and 2.16.

- *The assumption that the channel evolves discretely in time:* We have made this assumption only for simplicity of presenting the physical channel model of Subsection 2.5.2. This model can be assumed to be continuous time and none of the proof change. This is because of the nature of our proofs. As stated in Subsection 2.14.7, the first step is to prove that the universal capacity of the partially known channel $k \in \mathscr{A}$ which is capable of universally communicating i.i.d. X source to within a distortion level D is $\geq R_X^P(D)$. We claim that this argument does not depend on whether the channel evolves in time discretely or continuously. This is because the first step in the argument is to consider the composite channel $\mathscr{C}_{\mathscr{A}}$ defined in argument in Section 2.14.9. $c \in \mathscr{C}_{\mathscr{A}}$

evolves in discrete time even if $k \in \mathscr{A}$ evolved in continuous time. The proof nowhere uses the exact description of the set of channel $k \in \mathscr{A}$; it only uses the description of the channel $c \in \mathscr{C}_{\mathscr{A}}$. It follows that proving that the universal capacity of the partially known channel $k \in \mathscr{A}$ which is capable of universally communicating i.i.d. X source to within a distortion level D is $\geq R_X^P(D)$ does not require the assumption that the partially known channel channel $k \in \mathscr{A}$ evolve in discrete time. The proof of the universal source-channel separation theorem for rate-distortion uses source-compression followed by universal reliable communication over $c \in \mathscr{C}_{\mathscr{A}}$, and thus, again, uses the description of the partially known channel $k \in \mathscr{A}$ only through the partially known abstract channel $c \in \mathscr{C}_{\mathscr{A}}$. It follows that the universal source-channel separation theorem for rate-distortion holds even if the channel evolved in time continuously.

The only thing that one needs to care about is, how to rigorously model channels which evolve in time continuously. Once this is done, the proofs are automatic because of the reason described above.

- *The assumption that the source evolves in discrete time:* The ideas for how the universal source-channel separation theorem for rate-distortion can be generalized to a source evolving in continuous time is discussed in Section 2.18.
- *The assumption that the source and the channel evolve on the same time scale:* For simplicity of presentation, we have presented the results only for the source and the channel evolving on the same time scale. The results can be generalized to the case when the source and the channel evolve on different time scales.

2.20 Recapitulation

In this chapter, we proved a universal source-channel separation theorem for rate-distortion. The universality is over the channel and not the source, and we conjecture that the result can be made universal over the source.

The source-channel separation theorem, on a high-level says the following:

Assuming random-coding is permitted, in order to communicate a random source universally over a partially known channel to within a particular distortion level, it is sufficient to consider source-channel separation based architectures, that is, architectures which first code (compress) the random-source to within the particular distortion level, followed by universal reliable communication over the partially known channel. There is sufficiency in the sense if there exists some architecture to communicate the random source to within a certain distortion universally over the partially known channel, and which consumes certain amount of system resources (like energy and bandwidth), then there exists a separation based architecture to universally communicate the random source to within the same distortion universally over the partially known channel, and which consumes the same or lesser system resources as the original architecture.

By a partially known channel, we mean a channel whose behavior as a transition probability is only partially known. That is, the channel transition probability might belong to a family of channels. Other than that, the channel model is very general in that the present output of the channel can depend on some initial channel state, all past channel inputs and all past channel outputs

The set up is information theoretic, and thus, we assume that delays do not matter. In fact, if delays did matter, separation based architectures are not optimal from the point of view of Reason 1c stated in Section 1.4.1 of Chapter 1.

We prove why universal source-channel separation should hold when all sets (source space, channel input space, channel output space and source reproduction space) are finite, the source is the uniform X source and the distortion metric is permutation invariant. This requires a technical condition on the distortion function. *This section is the most important section of this thesis and should be thought of as the main idea for why separation holds for universal communication with a fidelity criterion.*

We rigorously prove the optimality of point-to-point setting when all sets (source space, channel input space, channel output space and source reproduction space) are finite, the source is i.i.d. and the distortion metric is additive.

We discuss, on a high level, how to generalize the results to infinite sets, stationary ergodic sources, and continuous time source and channel evolution. We also comment in brief that if random-coding were not permitted, universal source-channel for rate-distortion *does not* hold.

■ 2.21 In the next chapter ...

In the next chapter, we generalize the results of this chapter to the multiuser setting. We prove that if random-coding is permitted, and if the sources that various users want to communicate to each other are independent of each other (unicast setting), then separation based architectures are optimal for universal multiuser communication with fidelity criteria in the sense of reason 1c.

Chapter 3

Optimality of digital communication for communication with fidelity criteria: universal, unicast multi-user setting

It is, therefore, with a certain amount of hesitation that the present paper has been written. Its purpose is to present the formalization of the picture of Fig. 1 as the paradigm of a dynamical system in the hope of showing its usefulness in mathematics, engineering and physics alike.

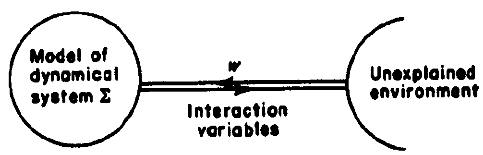


Fig. 1

A *dynamical system* Σ is defined as a triple

$$\Sigma = (T, W, \mathscr{B}) \tag{3.1}$$

with $T \subset \mathscr{R}$ the *time axis*; W an abstract set, called the signal alphabet; and $\mathscr{B} \subset W^T$ the *behaviour*.

-*Jan Willems*

■ 3.1 In this chapter ...

■ 3.1.1 Introduction

In this chapter, we generalize the results of Chapter 2 to the multi-user setting. That is, we prove that digital communication is optimal for rate-distortion communication in the sense of reason 1c stated in Section 1.4.1 in Chapter 1, in the multi-user setting.

Mathematically, the question is the following: Given a partially known medium m over which N users want to communicate. N might be unknown. For $1 \leq i, j \leq N, i \neq j$, user i wants to communicate a random source X_{ij} to user j to within a distortion level D_{ij} under the distortion metric d_{ij}. If such communication is possible with certain consumption of system resources (like energy and bandwidth) at each user, is the same communication possible with digital schemes with the same or lesser consumption of system resources at each user? See Figure 3.1.

Note that since we are assuming the medium to be partially known, we are asking the universal question, where universality is over the medium of communication.

We will answer the above question in the affirmative under the following assumptions:

- The sources X_{ij} are independent of each other. In the information-theory literature, the technical term for this is that the setting is unicast. This assumption is crucial in the sense that if the sources are correlated, digital communication is not optimal in the sense of reason 1c. This is discussed further in Section 3.2.
- Random-coding is permitted. This is the same assumption required in the point-to-point setting discussed in Chapter 2, and for precisely the same reason, as discussed in Section 2.17.
- For rigorous results, we assume that the sources are i.i.d. and the distortion measures are additive. As in Section 2.14, results will hold for permutation invariant distortion measures for uniform sources under certain technical assumptions on the rate-distortion functions of the various sources. As in Section 2.16, results should be rigorously generalizable for certain permutation invariant measures, for both i.i.d. and uniform sources without enforcing technical assumptions on the rate-distortion functions.

In this chapter, we will outline the proof for the reduction of the multi-user problem to the point-to-point problem. The proof is complete, though it is written in discursive style: the reader can fill in the minor missing details.

We do not provide any answers to the problem of reliable communication of bits over a medium. This is the classical problem of network information theory. Our view is a reduction view. *We reduce the problem of rate-distortion communication of various sources over a medium to the classical network information theory problem of reliable communication by showing the optimality of digital communication/source-channel separation architectures.* Indeed, we do this in the universal setting.

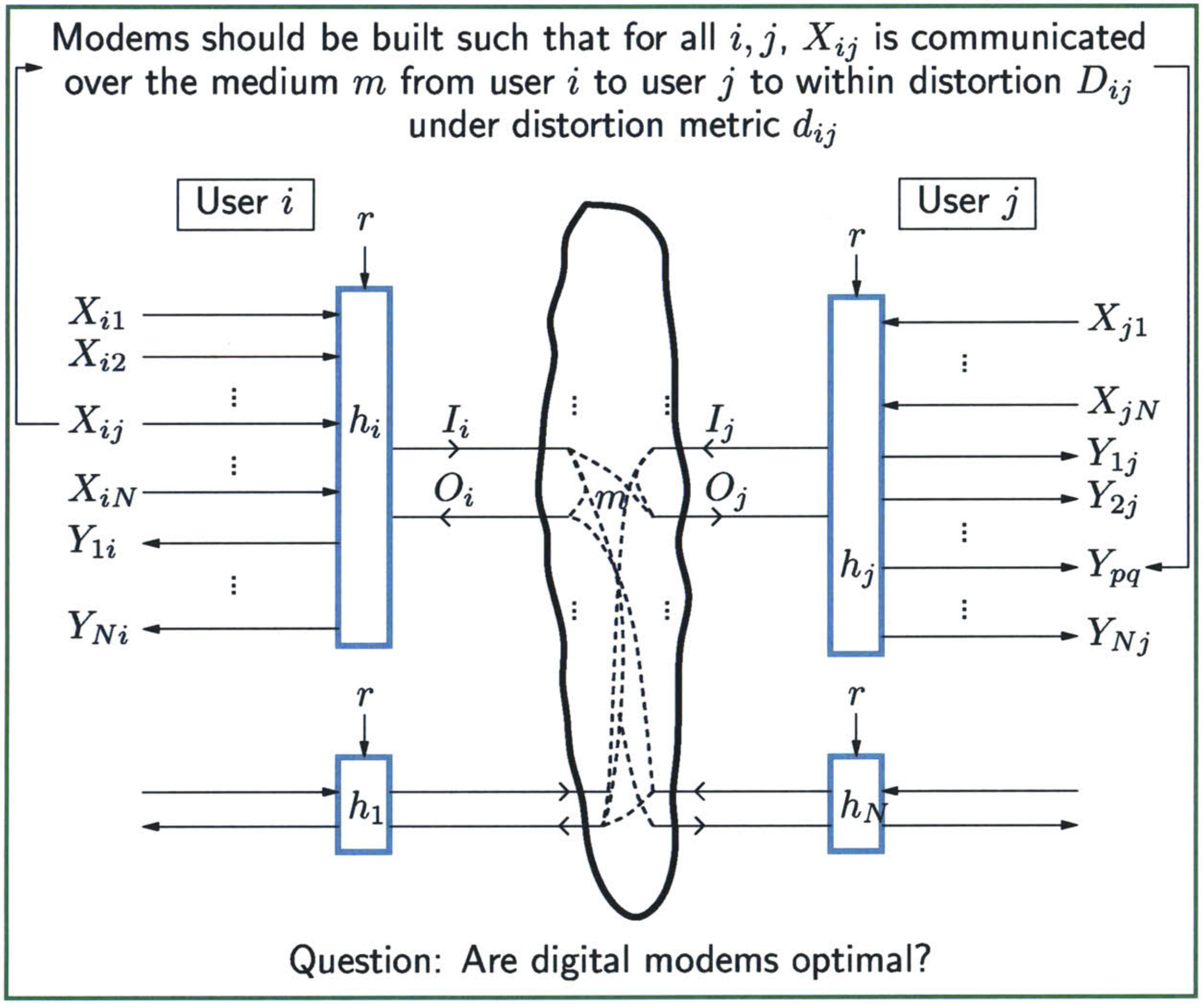


Figure 3.1. Are digital modems optimal for communication of independent random sources between various users to within certain distortion levels over a medium

3.1.2 A high-level statement of universal source-channel separation for rate-distortion in the multi-user setting

High level statement 3.1 (Universal source channel separation or the optimality of digital communication for universal multi-user communication with fidelity criteria in the unicast setting). *Let there be N users which want to communicate with each other* universally *over a partially known medium m. User i wants to communicate source X_{ij} over the medium m to user j to within a distortion level D_{ij} under distortion metric d_{ij}. Assuming that sources X_{ij} are independent and assuming that random-coding is permitted, it is sufficient to consider separation architectures: each user i, $1 \leq i \leq N$ first compresses the sources X_{ij}, $1 \leq j \leq N$, to within distortion levels D_{ij} under the distortion metric d_{ij}, followed by the universal reliable communication of all the compressed at various users over the partially known medium. There is sufficiency in the sense if there exists some architecture to communicate the random sources to within the required distortion levels universally over the partially known medium, and which consumes certain system resources (like energy and bandwidth) at each user, then there exists a separation architecture to universally communicate the random sources to within the same distortion levels universally over the partially known medium, and which consumes the same or lesser system resources at each user as in the original architecture.*

3.1.3 Chapter outline

Section 3.2 discusses the important past literature on the problem of communication with fidelity criteria in multi-user scenarios and the associated reduction to the traditional network information theory problem reliable communication over a network.

Section 3.3 discusses multi-user communication systems. A multi-user communication system consists of a physical medium interconnected with modems. We first we discuss multi-user communication systems on a high level, and then a rigorous view of these systems. As in the point-to-point framework, we assume that time evolves discretely in the medium though the results can be directly generalized to the case when the medium evolves in continuous time for the medium. This is followed by a discussion of resource consumption, for example, energy and bandwidth consumption in the system. This is followed by a short note on digital communication systems.

Section 3.5 is probably the most important section of this chapter. This section discusses the methodology that we will use to prove the optimality of digital communication for multi-user communication with fidelity criteria. This section will describe the methodology to reduce this problem to the problem of point-to-point communication discussed in Chapter 2.

Section 3.1 gives a precise statement of the universal source-channel separation for rate distortion in the unicast, multi-user setting, and Section 3.7 uses the methodology of Section 3.5 to prove this theorem.

This is followed by a recapitulation of this chapter in Section 3.8.

3.2 Important past literature

In Chapter 2, we discussed previous research for the point-to-ponit setting. In this section, we discuss previous important work for the multi-user setting.

The model of the medium that we will use is very general: the present medium outputs (the number of medium outputs is some number N) may depend, in genral, on all past medium inputs and all past medium outputs. The only research on multi-user problems with fidelity criteria which hold for general media (and not just for 3 user multiple-access or broadcast or for very specific media) is [TCDS]. [TCDS] proves the optimality of separation based architectures in multi-user, rate-distortion context, assuming the sources are independent of each other (that is the setting is unicast). In that sense, [TCDS], like us, also use the unicast setting. Our work is more general in the sense that we assume the medium to be only partially known whereas the medium in [TCDS] is assumed to be fully known: in other words, we are solving the universal problem over the medium, whereas [TCDS] does not. There is one other minor difference: [TCDS] uses the expected distortion criterion whereas we use the probability of excess distortion criterion. As a result, [TCDS] requires finite memory assumptions on the medium whereas we do not: this is a minor difference however, the major difference is that we are solving the universal problem over the medium whereas [TCDS] is not.

Both [TCDS] and our work use the unicast setting to prove separation for communication with a fidelity criterion. One wonders if this is necessary. The answer is that there exist counter examples when separation does not hold if the sources are correlated. Two examples where separation does not hold are discussed in [Gas02] on Pages 27 and 28: reliable communication over the multiple access channel with correlated sources and communication of a gaussian source over a broadcast channel with fidelity. In both these examples, it is shown that uncoded transmission achieves better performance than the performance achievable by separation architectures. *Thus, it follows that in order to prove general results, the unicast setting is the extent to which we can go in order to prove exact separation results.*

One question that arises is the following: can we say anything about the nature of separation when the sources are not independent (that is, they are correlated). [TCDS] proves approximate optimality results for separation under various restrictions on the distortion measure, in the multi-cast setting, that is, when a user wants to communicate the same source to within distortion levels to different users: in certain cases, they prove that the performance of separation architectures is to within $\frac{1}{2}$ bit of the performance of general architectures.

We have *not* proved any approximate optimality results in this thesis when the sources are correlated.

The network coding literature also talks about another sense of separation: separation between channel-coding and network-coding. See for example, [KEM]. Separation of channel-coding and network-coding means the following: consider a network which consists of noisy channels/links. Channel/network coding separation holds for a link in a network if a channel of capacity C in the network can be replaced with an error free link with throughput C without changing the rate region. For networks for which this happens, the network information theory problem of reliable communication can be reduced to the network-coding problem. As we have stated before, reliable communication and how to accomplish it is not the focus of this thesis. The focus of this thesis is the reduction of rate-distortion communication problems in multi-user settings to the problem of reliable communication of bits in multi-user settings; thus, we implicitly assume that we know how to solve the multi-user reliable communication problem. In this thesis, when we mention separation, we always mean the separation of source-coding and channel-coding, and not the separation of channel-coding and network-coding.

■ 3.3 A multi-user communication system

In this section, we describe the model of a multi-user communication system and a digital multi-user communication system. We also comment on the resource consumption in a multi-user communication system.

In a point-to-point communication system, communication happens over a channel. In a multi-user communication system, communication happens over what we call, a medium. In a point-to-point communication system, encoders and decoders aid communication. In a multi-user communication system, the communication happens with the help of modulators-demodulators or modems: this is because each user is both a sender and a receiver.

■ 3.3.1 High level view of a multi-user point-to-point communication system

The following is the high-level view of a multi-user point-to-point communication system. This is a more detailed description of the description can be found in the next sub-section.

There are various users. The users communicate sources among each other. The system consists of "architecture boxes" interconnected to a medium. The architecture boxes aid communication. The architecture boxes can be thought of as system protocol or modulators-demodulators. Architecture boxes, now onwards, will be refered to as modems. See Figure 3.2.

More concretely:

There are N users. N might change with time. User i communicates source $X_{ij}(t)$ to user j over the system. $i \neq j$: a user does not transmit anything to itself. The reproduction of $X_{ij}(t)$ at user j is $Y_{ij}(t)$. The system potentially provides some guarantees on how close the source reproduction $Y_{ij}(t)$ is to the source $X_{ij}(t)$. One example of a guarantee, and the one we will use is the following: source $X_{ij}(t)$ is communicated to within some distortion level.

Note the ordering of i and j in $Y_{ij}(t)$.

m denotes the medium. $h_1, h_2, \ldots, h_i, \ldots h_N$ are various modems constructed over the medium. h_i is the modem used by user i, $1 \le i \le N$.

Modem h_i at user i takes source inputs $X_{i1}(t), X_{i2}(t), \ldots, X_{ij}(t), \ldots, X_{iN}(t)$. h_i takes input $I_i(t)$ from the medium m. In wireless systems, $I_i(t)$ is an electromagnetic wave. Modem h_i produces an output $O_i(t)$ into the medium m. In wireless systems, $O_i(t)$ is an electromagnetic wave. The modems also have a common randomness input r which they use the generate random codes. Modem h_i produces outputs source reproductions $Y_{1i}(t), Y_{2i}(t), \ldots, Y_{ji}(t), \ldots, Y_{Ni}(t)$. I_i is an input to the medium m but output to the modem h_i. $O_i(t)$ is an output of the medium m but an input to the modem h_i.

The medium takes inputs $I_1(t)$, $I_2(t)$, ..., $I_N(t)$ and produces outputs $O_1(t)$, $O_2(t)$, ..., $O_N(t)$.

The time evolution in $X_{ij}(t)$, $I_i(t)$, $O_i(t)$, $Y_{ij}(t)$ will be sometimes supressed. They will be denoted by X_{ij}, I_i, O_i, Y_{ij}.

The modem h_i encodes information into input I_i. I_i contains information about

1. Sources $X_{ij}, 1 \le j \le N$ that user i wants to communicate to other users.
2. Sources $X_{i'j'}, i' \neq i$. Modem h_i has knowledge of other other sources $X_{i'j'}$ which are not inputs at user i through the medium output O_i. In this case, information about $X_{i'j'}$ is being relayed through user i.

The behavior of medium m may be complex. For example, wireless medium. The interaction of medium m and modems h_i and the resulting flow of information may be complex. The users may be co-operating. There may be multi-hopping and feedback.

The exact medium and system behavior does not concern us. In our formulation, the end-to-end system behavior that X_{ij} is communicated from user i to user j and received as Y_{ij} is what will matter. The precise model that we will use is that the medium behavior as a transition probability is only partially known in the sense that it might come from a set of transition probabilities. We will call this a partially known medium. The precise mathematical model is the subject of Subsection 3.3.2.

The sources X_{ij} should be thought of as primitive in the sense that system behavior, that is, the behavior of modems h_i and medium m do not affect the sources. Sources evolve in time, independently of "everything else."

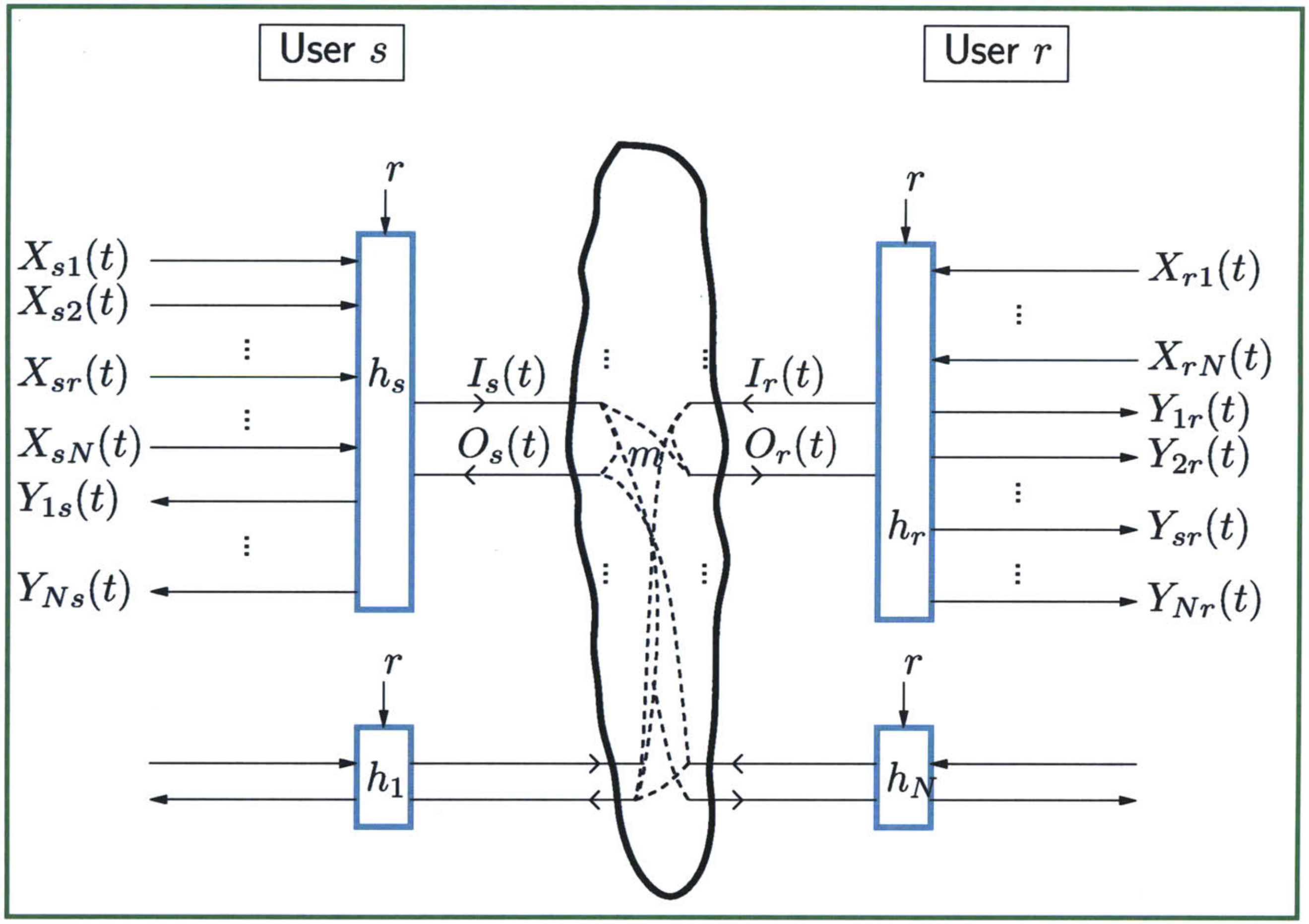


Figure 3.2. High level model of the system

3.3.2 Rigorous, mathematical view of a point-to-point communication system

This subsection states the precise mathematical model of the given system which was described imprecisely in the previous subsection. This model illustrates, how a "physical" model of interconnection of a *partially known* medium with modems can be abstracted mathematically.

Some basic definitions

There are N users. N might change with time.

The system consists of

1. a medium m, and
2. a modem h_i at each user i, $1 \leq i \leq N$,

interconnected to each other.

Medium and modems are modeled stochastically. Sources are modeled as random processes.

We assume that medium and modems evolve only at integer times. They would start evolving at some time, and we assume that this time is 1 (in general, it can be any integer time). This assumption of evolution at integer times is made only for simplicity of presentation. All our results hold even if medium and modems evolved in time, continuously.

In what follows, $i \neq j$. This just means that a user is not communicating to itself. This will not be stated again.

A fully known medium

First, we describe a fully known medium. This discussion will parallel the discussion of a fully known physical channel in Section 2.5.2.

A fully known physical medium is one whose action as a transition probability is known. This is mathematically abstracted as follows:

We want to use a very general medium model. The output of the medium can depend on all past medium inputs and all past medium outputs. The medium input at user i is denoted by ι_i and the medium output at user i is denoted by o_i. Let the block-length be n. The medium transition probability at time k is denoted by m_k:

$$\begin{aligned} &m_k(o_i^n(k), 1 \leq i \leq N \mid \iota_i^n(1..k-1), 1 \leq i \leq N, o^n(1..k-1), 1 \leq i \leq N) \text{ if } if\, k \leq n \\ &m_k(o_i^n(k), 1 \leq i \leq N \mid \iota_i^n, 1 \leq i \leq N, o^n(1..k-1), 1 \leq i \leq N) \text{ if } k > n \end{aligned} \tag{3.2}$$

is the probability that the medium output at time k at user i is $o_i^n(k)$, $1 \le i \le N$, given that past medium inputs at various users are $\iota_i^n(1..k-1)$ (or ι_i^n, $1 \le i \le N$ if $k > n$) and the past medium outputs at the various users are $o_i^n(1..k-1), 1 \le i \le N$.

See Figure 3.3.

For each k, the medium input $\iota_i^n(k)$ is assumed to belong to some finite set $\mathscr{I}_i$ and medium output $o_i^n(k)$ is assumed to belong to some finite set $\mathscr{O}_i$.

Note that m_k is independent of the block-length n. When the block-length is n, the medium evolves until some time $t_n \ge n$ where t_n is an increasing function of n. The medium is $m^n = (m_1, m_2, \ldots, m_{t_n})$.

Note that the medium model is nested.

Over various block-lengths, the medium evolution is $m =< m^n >_1^\infty$.

In general, there is a dependence of m_k on the initial channel state. However, we do not show this dependence. This is because, the model of the medium that we will use is a partially known medium, in that, the medium can belong to a family. For that reason, we will treat the same medium with different initial states as different media and assume that all these media belong to the family which make up the partially known medium. A partially known medium is discussed next.

A partially known medium

Since the medium initial state might not be entirely known, and also, the exact action of the medium m^n as a transition probability might not be entirely known, we will model the medium as a partially known medium:

A partially known medium is one which belongs to a family of transition probabilities $\mathscr{A}$. We will denote this by $m \in \mathscr{A}$.

Modems

There is a modem h_i at each user i. Each modem takes various inputs and outputs. For block-length n, the modem acts as h_i^n. As the encoders and decoders in the point-to-point setting, the modems h_i^n need not be nested. Thus, $h_i =< h_i^n >_1^\infty$. The action of a modem is described below. See also, Figure 3.4.

When the block-length is n, the modem h_i^n at user i takes various inputs:

1. The source input sequence x_{ij}^n, $1 \le j \le N$, that user i wants to communicate to user j. x_{ij}^n is a realization of X_{ij}^n. $x_{ij}^n(k)$ belongs to some set $\mathscr{X}_{ij}$ for all k
2. The input o_i^n from the medium m. The notation is o_i because o_i is an *output* of the

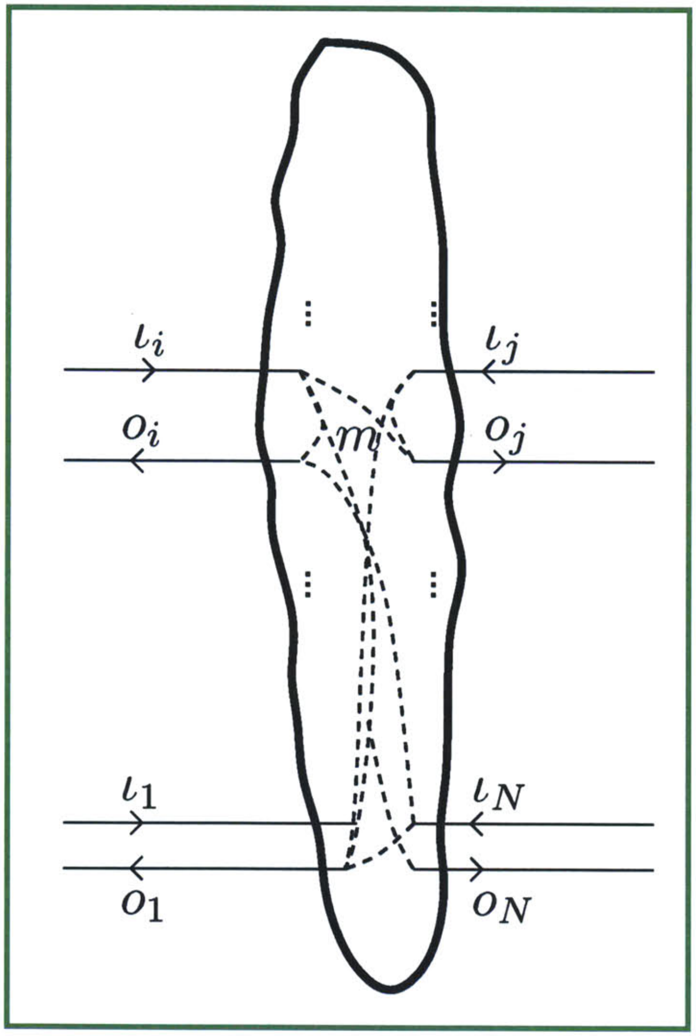


Figure 3.3. The medium

medium.

3. Common randomness input r. r is input to all $h_i, 1 \leq i \leq N$. r is a realization of a continuous-valued random variable R.

The modem h_i produces various outputs:

1. The source reproduction output y^n_{ji} $1 \leq j \leq N$, that user j wants to communicate to

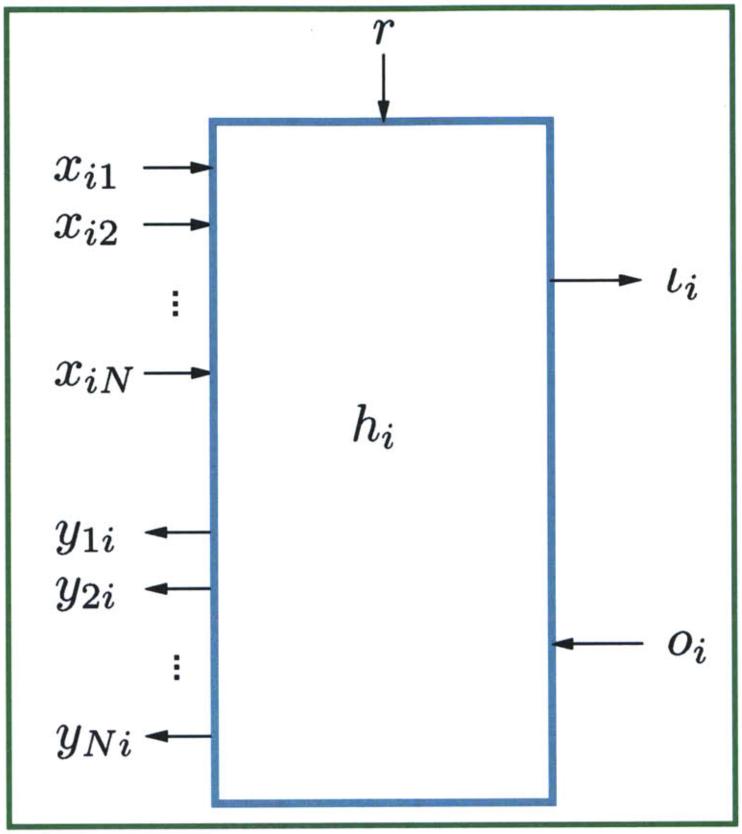


Figure 3.4. The modem at user i

user i. $y^n_{ji}(k)$ belongs to some set $\mathscr{Y}_{ji}$ $\forall k$.

Note the ordering of i and j in x_{ij} and y_{ji}. The modems and the medium act causally. We should think of y_{ij} as a reproduction of x_{ij}, but with a time delay. The n^{th} source symbol of the source x_{ij} may be reproduced at some time $n' > n$. Thus, $y^{n'}_{ij}(n')$ might be the reproduction of $x^n_{ij}(n)$ for some $n' > n$.

2. The output ι^n_i into the medium from the modem. The notation is ι_i because ι_i is an input to the medium.

Note 3.1. Sources, modem inputs and outputs, and source reproduction outputs are all random processes. The corresponding notation is described later.

The modem h^n_i is a transition probability. At time k, the transition probability is $h^n_i(k)$ and it acts as:

$$h_i^n(k)(y_{ji}^n(k), 1 \le j \le N, \iota_i^n(k) \mid \\ x_{ij}^n, 1 \le j \le N, o_i^n(1..k-1), \iota_i^n(1..k-1), y_{ji}^n(1..k-1), 1 \le j \le N, r), \quad (3.3)$$

is the probability that when the block-length is n, the modem outputs at time k are

- source reproduction outputs $y_{ji}^n(k), 1 \le j \le N$ and
- modem output $\iota_i^n(k)$ which is an input to the medium

given that

- source inputs are x_{ij}^n,
- medium outputs (into the modem) until time $k-1$ are $o_i^n(1..k-1)$,
- the common randomness input is r, and
- the past modem outputs until time $k-1$ are $y_{ji}^n(1..k-1), 1 \le j \le N$ and $\iota_i^n(1..k-1)$

The modems evolve until time t_n. t_n is the same time which we discussed when talking about media. The reproduction of the source input $x_{ij}^n(k)$ will happen at some time $t_{n,k} \ge k$. Thus $x_{ij}^n(k)$ is reproduced as $y_{ij}^n(t_{n,k})$. When the block-length is n, the modem

$$h_i^n = (h_i^n(1), h_i^n(2), \ldots, h_i^n(t_n)) \quad (3.4)$$

h_i^n need not be nested.

In the point-to-point case, when the encoders and decoders were separate, the model was less complex. We could just define a block-length n encoder as $e^n(i^n \mid x^n)$ and block-length n decoder as $f^n(y^n \mid o^n)$. Since, the modem acts as both a modulator and a demodulator, we need to take into account the causality of the interaction of the medium inputs (which are outputs of the modems) and medium outputs (which are inputs to the modems), and this leads to the above model where we need to describe $h_{ij}^n(k)$ separately for each k.

Interconnection of medium and modems

The medium m and modems $h_i, 1 \le i \le N$ are interconnected as shown in Figure 3.2. Note, in this figure, that the common randomness input r is the same for all modems.

The sources x_{ij}, as stated before, are realizations of random processes X_{ij}. With these inputs, and the common randomness input variable R, the interconnected system evolves stochastically. After marginalizing out the common randomness variable R, for each initial state s, this leads to a joint random variable on the space of source, medium input, medium output and source reproduction random variables:

$$X^n_{ij} I^n_i O^n_i Y^n_{ij}, 1 \le i, j \le N \tag{3.5}$$

Note that for all i, j, X^n_{ij} is a vector of length n, whereas I^n_i, O^n_i, Y^n_{ij} are vectors of length t_n. $x^n_{ij}(k)$ is reproduced at a time $t_{n,k} > k$ as $y^n_{ij}(t_{n,k})$.

We will re-label $y^n_{ij}(t_{n,k})$ as $y^n_{ij}(k)$. This way, $y^n_{ij}(k)$ will be the reproduction of $x^n_{ij}(k)$ and this would lead to simpler notation. The rest of the $y^n_{ij}(t)$ such that t is not equal to $t_{n,k}$ for any k are not represented in this notation. That is fine because these $y^n_{ij}(t)$ do not serve any purpose anyway.

Thus, now, in (3.5), X^n_{ij} and Y^n_{ij} are vectors of length n whereas I^n_i and O^n_i are vectors of length t_n.

See Figure 3.2, except that in the more rigorous notation, $X_{ij}(t)$ would be denoted by X^n_{ij} when the block-length is n and similarly for other inputs and outputs, and the modem h_i is denoted by h^n_i, and similarly for other modems.

The end-to-end or the abstract system

With the interconnection of the medium and the modems, we can think of the end-to-end system consisting only of source inputs and and source reproduction outputs and not looking at the medium inputs and outputs. When the block-length is n, at time k, the source input $x^n_{ij}(k)$ is reproduced as $y^n_{ij}(k)$.

■ 3.3.3 Resource consumption in the multi-user communication system

Th following discussion of resource consumption builds on top of the corresponding discussion for the point-to-point setting in Subsection 2.6.7.

Consider two multi-user communication systems s_1 and s_2. Let the various random-variables of (3.5) for system s_1 when the block-length is n be $X^n_{ij} I^n_{i,1} O^n_{i,1} Y^n_{ij,1}$, $1 \le i, j \le N$, and let these random-variables for system s_2 be $X^n_{ij} I^n_{i,2} O^n_{i,2} Y^n_{ij,2}$. Note that as in the point-to-point setting of Subsection 2.6.7, the inputs X_{ij}, the same for both the systems.

A sufficient condition for the two systems s_1 and s_2 to consume the same system resources is that for each block-length n, the joint distribution of $I^n_{i,1}, 1 \le i \le N$ is the same as the joint distribution of $I^n_{i,2}, 1 \le i \le N$. Note the parallel with the definition in the point-to-point setting. Also, as in the point-to-point setting, this is a sufficient condition for two systems to consume the same system resources, not a necessary condition. This sufficient condition for equality for consumption of system resources makes sense for reasons similar to the point-to-point setting, discussed in Subsection 2.6.7; a discussion is omitted here. For our purposes, when making digital architectures corresponding to general analog architectures, we will maintain the medium input distributions, and for that reason, this is the only condition which we will need.

The following discussion of the consumption of "lesser" system resources is precisely the same as in the point-to-point setting:

Suppose a communication system needs to be built to meet certain communication guarantees. Suppose this can be done with certain consumption of system resources. Then, we will say, abstractly, that the same guarantee can be met by consumption of the same or "lesser" system resources. This, again, is an abstract definition because we have not defined the consumption of a system resource; we have only stated a sufficient condition for the equality of consumption of the same system resources by two systems. However, the reason for this abstract use of the word "lesser" is done because for physical systems where resource consumption can in fact be defined, this would be the right usage of "the same guarantee can be met by consumption of the same or lesser system resources."

■ 3.4 A multi-user digital communication system

A multi-user communication system is one where each modem is digital. We only provide a high-level description, which is the same as the description of Section 1.6.

A digital modem h_i is portrayed in Figure 1.4. At user i, the sources $X_{i1}, X_{i2}, \ldots, X_{iN}$ are first converted to random binary sequences by the source encoders. These binary sequences are communicated reliably over the medium with the help of medium modems at the various users. Finally, at user i, the source decoders help produce the reproductions $Y_{1i}, \ldots, Y_{Ni}$, of $X_{1i}, \ldots, X_{Ni}$, respectively. The story is the same at each modem h_i.

Resource consumption in a digital communication system is defined in exactly the same way as in a general communication system: this is because the sufficient condition that we have stated for equality of consumption of system resources depends only on the medium inputs.

A rigorous description is omitted.

■ 3.5 Spirit of the question: the idea that we will use to reduce proving optimality of digital communication in multi-user setting to proving the optimality of digital communication in the point-to-point setting

In this section, we describe the idea that we will use to reduce proving the optimality of digital communication for the multi-user setting to the point-to-point setting. This section just describes the idea: the high-level methodology for proving the optimality of digital communication in multi-user communication problems is discussed in the next section.

Given a multi-user communication system which is known to communicate random sources $X_{ij}(t)$ from user i to user j, $1 \leq i, j \leq N$ over a medium m with the help of modems h_i at user i. Further, let s and r be two particular users. It is known that source $X_{sr}(t)$ is communicated from user s to user r over the system to within some guarantee. Denote the guarantee by G. See Figure 3.5.. $X_{sr}(t)$ is received as $Y_{sr}(t)$. An example of a guarantee and the one we will use is that $X_{sr}(t)$ is communicated to within some distortion level.

We ask a question about the communication of another random source $X'_{sr}(t)$ evolving in time from user s to user r, *in place of* the source $X_{sr}(t)$. The source $X'_{sr}(t)$ should be received with some other guarantee G'. An example of gaurantee G' and the one we will use is that $X'_{sr}(t)$ needs to be communicated to within some distortion level. By "$X'_{sr}(t)$ should be communicated *in place of* $X_{sr}(t)$ with guarantee G," we mean that the source $X_{sr}(t)$ need not be communicated to with guarantee G any more in the new communication system which communicates $X'_{sr}(t)$ with guarantee G.

We will assume that the sources X_{ij} are independent of each other $\forall i, j$. This assumption is crucial.

We will also assume that the source $X'_{sr}(t)$ is independent of sources $X_{ij}(t) \forall i, j$. We are uncertain about the cruciality of this assumption for our results. In order to prove the result concerning optimality of digital communication in the multi-user setting that we do in the next section, this assumption is okay to make; in fact, we will make $X'_{sr}(t)$ have the same distribution as $X_{sr}(t)$ but independent of all $X_{ij}(t) \forall i, j$, in particular, independent of $X_{sr}(t)$. $X'_{sr}(t)$ is primitive in the sense that it evolves in time, independently of the rest of the system.

We require that changes made in the system for the desired communication of $X'_{sr}(t)$ from user s to user r should not change the communication of $X_{ij}(t)$ from user i to user j for $(i, j) \neq (s, r)$. Mathematically, this means that $X_{ij}(t)$ should be received precisely as $Y_{ij}(t)$ *in distribution* for $(i, j) \neq (s, r)$. Of course, as stated and emphasized before, instead of $X_{sr}(t)$, $X'_{sr}(t)$ now, needs to be communicated from user s to user r.

Each user only has local knowledge. At time τ, user i has knowledge of the source realization $x_{ij}(t), -\infty \leq t < \tau, 1 \leq j \leq N$, the modem h_i, medium input realization $\iota_i(t), -\infty < t < \tau$, medium output realization $o_i(t), -\infty < t < \tau$, the realization of reproduction of sources from various users destined for user i, $y_{ji}(t), -\infty < t < \tau, 1 \leq j \leq N$ and the common randomness input r. User i also has knowledge of any guarantees associated with sources at user i, that is, sources $X_{ij}, 1 \leq j \leq N$. There is knowledge at each user that the sources $X_{ij}, 1 \leq i, j \leq N, X'_{sr}$ are all independent of each other.

Users do not have knowledge of the action of the medium as a transition probability. In particular, the setting is universal.

System architecture can be changed, only locally. h_s and h_r can be changed in order to communicate the source X'_{sr}. All other modems should remain the same: that is, for $i \neq s, r$, h_i should remain unchanged.

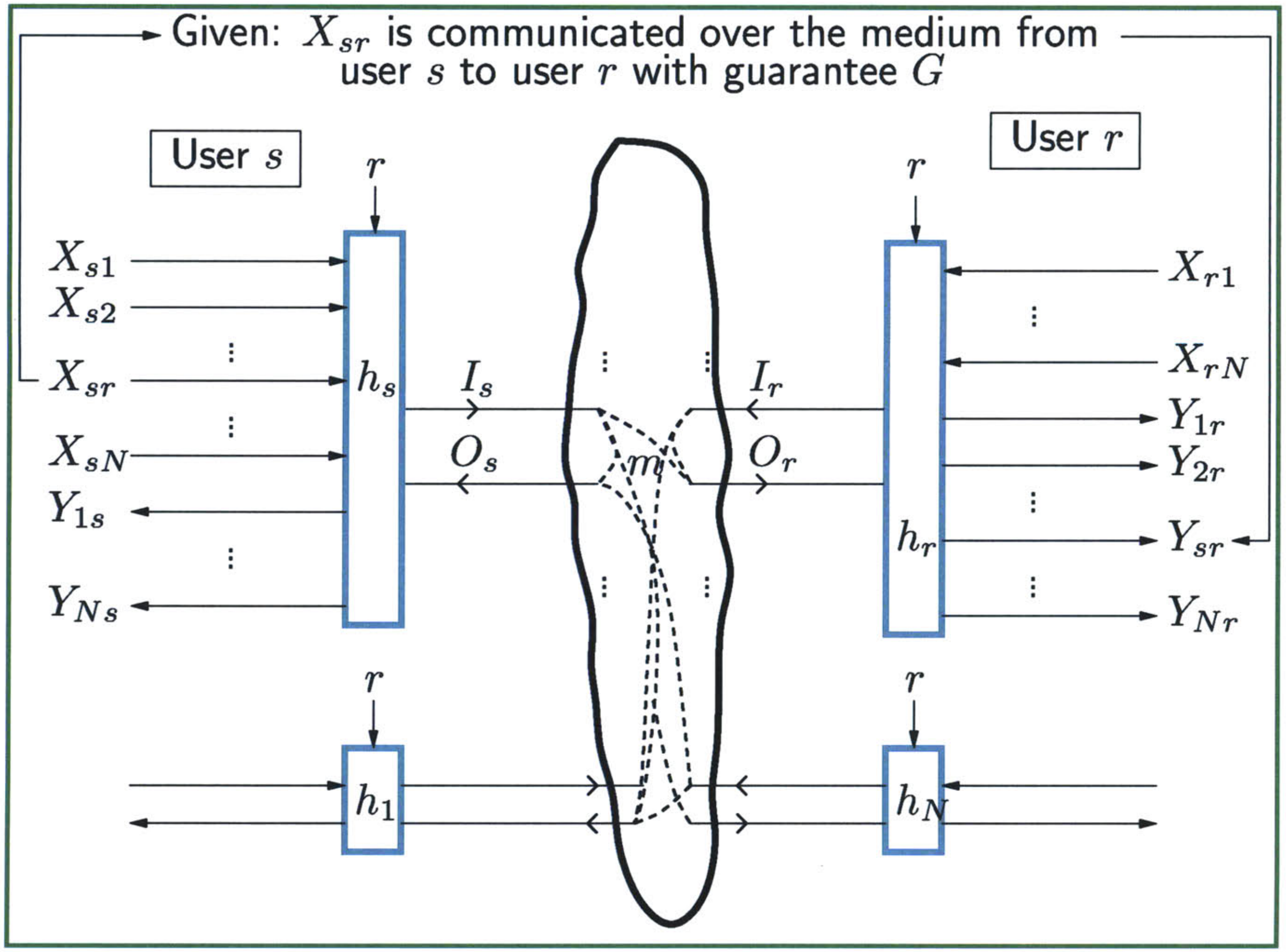


Figure 3.5. Spirit of the question: given that source X_{sr} is communicated with guarantee G over the medium from user s to user r

Question: When can $X'_{sr}(t)$ be communicated to within the required guarantee G' in place of $X_{sr}(t)$ which is known to be communicated with guarantee G, and how?

We have not stated the definitions of guarantees G and G'. In this section, we only state the view that we are going to take.

The communication of $X'_{sr}(t)$ *in place of* $X_{sr}(t)$ will be accomplished in the following way:

Since there is no knowledge of the medium transition probability, we would like to mantain the input-output behavior of the medium. We want to use this method because if the joint input distribution of the medium inputs $I_i(t), 1 \leq i \leq N$ is changed, in the absence of the knowledge of medium transition probability, it is impossible to know the evolution of the medium outputs. In order to mantain the medium joint input distribution, we would mantain the distribution $X_{sr}(t)$. We would build an encoder e which would map the source $X'_{sr}(t)$ into an encoded input whose distribution is precisely the same as the source process $X_{sr}(t)$. *We will thus simulate* $X_{sr}(t)$. *Denote this simulated source by* $X^s_{sr}(t)$. The guarantee G will be satisfied between the simulated source X^s_{sr}(t) and output which we denote by $Y^s_{sr}(t)$. We will then use this output $Y^s_{sr}(t)$ to make a decoding $Y'_{sr}(t)$ with the use of a decoder f. See Figure 3.6.

Notation 3.1 (Simulated source). The simulated source at the input to modem h_s at the point where the input was X_{sr} is X^s_{sr}. The corresponding "simulated output" is Y^s_{sr}.

This encoding procedure can be thought of as embedding information about X'_{sr} *into* X_{sr}.

Interconnecting the encoder e to the modem h_s by maintaining the distribution can also be thought of in terms of interconnections and maintaining behaviors, as in the stochastic counterpart of the behavioral view of Willems [Wil89] and [Wil07].

Note that with this encoding-decoding procedure, we will not be "breaking" the modems h_s and h_r. The new modem h'_s at user s is the composition of h_s and e. The new modem h'_r at user r is the composition of d and h_r. In other words, we are building "on top of" the existing architecture to accomplish the required communication. Note that e is a random code. Note that the encoder-decoder (e, f) is a random code.

By requirement, the modem h'_i is the same as h_i for $i \neq s, r$.

The joint distribution of the inputs to modems h_i has been mantained. This is because $X_{ij}(t)$ is unchanged for $(i, j) \neq (s, r)$. For $(i, j) = (s, r)$, the input, now is $X^s_{sr}(t)$ instead of $X_{sr}(t)$. $X^s_{sr}(t)$ has the same distribution as $X_{sr}(t)$. $X^s_{sr}(t)$ is independent of $X_{ij}(t), (i, j) \neq (s, r)$ by construction and because of the assumption that $X'_{sr}(t)$ is independent of $X_{ij}(t)$. Thus, the joint distribution at the inputs to modems h_i has been mantained. As a result, $X_{ij}(t)$ is received precisely as $Y_{ij}(t)$ for $(i, j) \neq (s, r)$ (note that this is a statement about random processes and what we mean is that in distribution, $X_{ij}(t)$ is received precisely as $Y_{ij}(t)$ for $(i, j) \neq (s, r)$). Of course, $X_{sr}(t)$ is not received as $Y_{sr}(t)$ because $X_{sr}(t)$ does not need to be communicated any more. The goal is to communicate $X'_{sr}(t)$ instead of $X_{sr}(t)$. Instead of

$X_{sr}(t)$, thus, its simulated version $X^s_{sr}(t)$ is transmitted.

We stated before that we would like the joint medium input and output distributions to be mantained. By mantaining the distribution of $X_{sr}(t)$, this has automatically happened.

Note 3.2. We are using this way of simulating $X_{sr}(t)$ and "building on top" of the already existing architecture in order to communicate $X'_{sr}(t)$ from user s to user r. Other ways may exist. This is the view and method that we use.

The assumption of independence of sources X_{ij} is required in the above construction for the following reason:

Let $X_{ij}(t)$ and $X_{sr}(t)$, $(i,j) \neq (s,r)$ be dependent. In order to communicate $X'_{sr}(t)$, we simulate $X_{sr}(t)$ as described above. This would mean that $X_{ij}(t)$ would also need to be, atleast partially simulated in order to respect the joint distribution of $X_{sr}(t)$ and $X_{ij}(t)$. This would mean that the system behavior would change for the transmission of $X_{ij}(t)$ from user i to user j. This is not permitted.

By construction, the joint distribution of the medium inputs has been maintained. By the discussion in Subsection 3.3.3. consumption of systems resources at each modem and the consumption of system resources in total is unchanged.

Note that it is not yet unclear if such a system for communication of X'_{sr} with guarantee G' *in place of* X_{sr} exists. We are just describing the view that we will take. A similar procedure can potentially be followed for communication of other sources $X'_{ij}(t)$ from a user i to user j, $1 \leq i,j \leq N$ in place of source X_{ij}, with some other guarantee. This results in a decentralized system for communication of various sources between various users over a medium.

Note further that in order to construct e and f, we did not require the knowledge of the medium transition probability. Thus, the medium might only be partially known. We are thus, solving the universal problem, where universality is over the medium.

In Section 3.7, we will use the reasoning described in this section to prove universal source-channel separation for rate-distortion in the multi-user setting by making the source $X'_{sr}(t)$ have the same distribution as the source $X_{sr}(t)$. First, we make a precise statement in the next section.

■ 3.6 A precise statement of the optimality of digital communication for universal multi-user communication with fidelity criteria

Theorem 3.1 (Universal source channel separation or the optimality of digital communication for universal multi-user communication with fidelity criteria in the unicast setting). *Let there be N users which want to communicate with each other over a partially known medium m. User i wants to universally communicate source X_{ij} over the medium m to user j to within a distortion level D_{ij} under an additive distortion metric d_{ij}. Assuming that sources X_{ij} are*

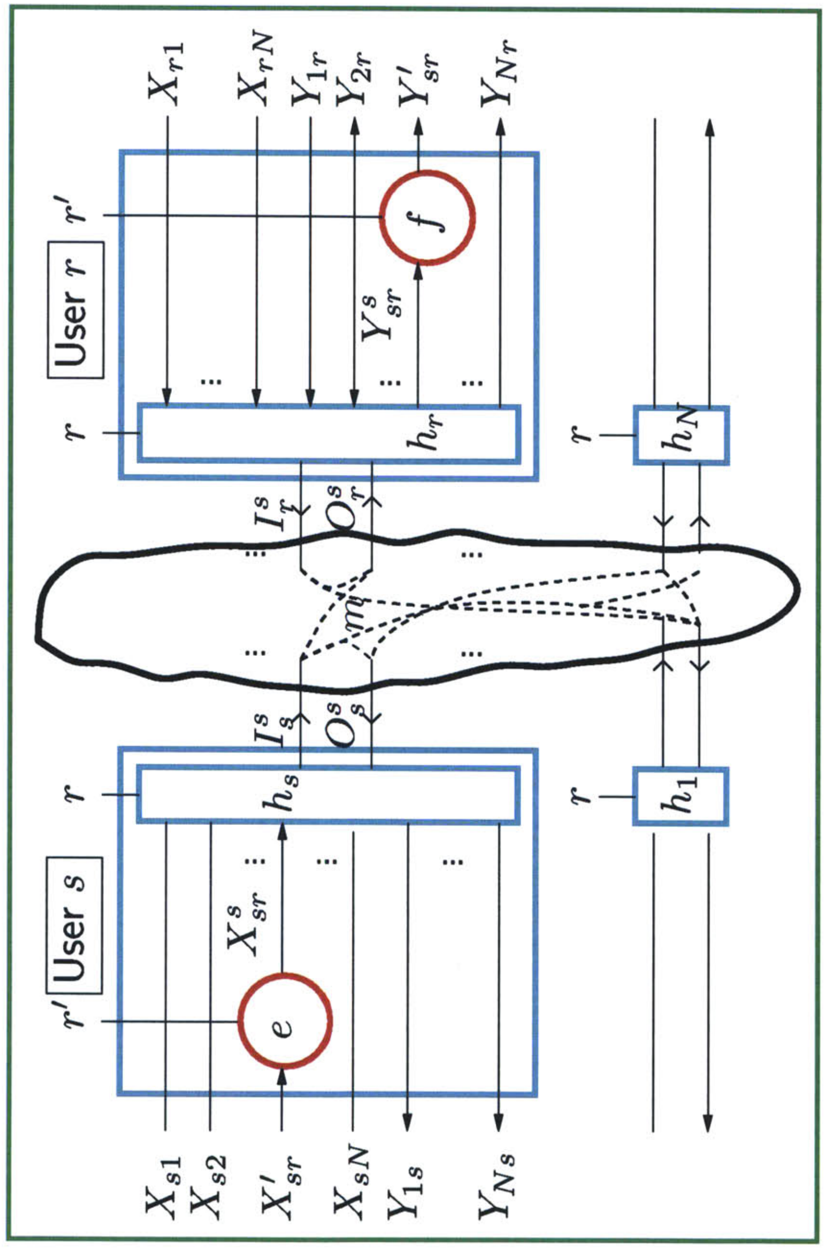


Figure 3.6. Methodology that we will use: need to build encoder-decoder $e - f$ such that source X'_{sr} is communicated with guarantee G from user s to user r *in place of* X_{sr} in such a way that the marginal input into h_s, X^s_{sr} has the same distribution as X_{sr}

independent and assuming that random-coding is permitted, it is sufficient to consider separation architectures: each user i, $1 \le i \le N$ first compresses the sources X_{ij}, $1 \le j \le N$, to within distortion levels D_{ij} under the distortion metric d_{ij}, followed by the universal reliable communication of all the compressed sources at various users over the partially known medium. There is sufficiency in the sense if there exists some architecture to communicate the random sources to within the required distortion levels universally over the partially known medium, and which consumes certain system resources (like energy and bandwidth) at each user, then there exists a separation architecture to universally communicate the random sources to within the same distortion levels universally over the partially known medium, and which consumes the same or lesser system resources at each user as in the original architecture.

■ 3.7 The proof of Theorem 3.1

Using the point-to-point formalism of Chapters 2 and the methodology described in the previous section, we provide an outline of how to generalize universal source-channel separation from the point-to-point setting to the multi-user setting. The proof is written in discursive style and is essentially complete.

Let $X_{ij} =< X_{ij}^n >_1^\infty, 1 \le i, j \le N$ be *independent* sources.

Given that there exists a system consisting of a partially known medium m and modems h_i at user i, $1 \le i \le N$, interconnected to each other as in Figure 3.2, such that independent sources X_{ij} are communicated from user i to user j and received as Y_{ij} with the help of modems $< h_i^n >_1^\infty$ at user i, $1 \le i \le N$, where the modem at user i is h_i^n when the block-length is n.

Let s and r be two particular users.

Let the source X_{sr} be i.i.d. Note that for now, this i.i.d. assumption is made only for the source X_{sr} and not for the other sources X_{ij}, $(i,j) \ne (s,r)$. It is known that with the above system architecture, the source X_{sr} is communicated to within a probability of excess distortion D_{sr} under an *additive* distortion metric d_{sr} from user s to user r over the partially known medium m . That is, for some $\omega_{sr} =< \omega_{sr,n} >_1^\infty$, $\omega_{sr,n} \to 0$ as $n \to \infty$, end-to-end,

$$p_{X_{sr}^n Y_{sr}^n}\left(\frac{1}{n} d_{sr}^n(X_{sr}^n, Y_{sr}^n) > D_{sr}\right) < \omega_{sr,n} \to 0 \text{ as } n \to \infty \tag{3.6}$$

We will keep referring to Figure 3.7.

Question: assuming that there is common randomness, can source X_{sr} be communicated to within a distortion D from user s to user r by using a separation architecture, that is, by using an architecture which first compresses the source X_{sr} to within distortion level D_{sr} followed by reliable communication of the compressed source over the medium in such a way that

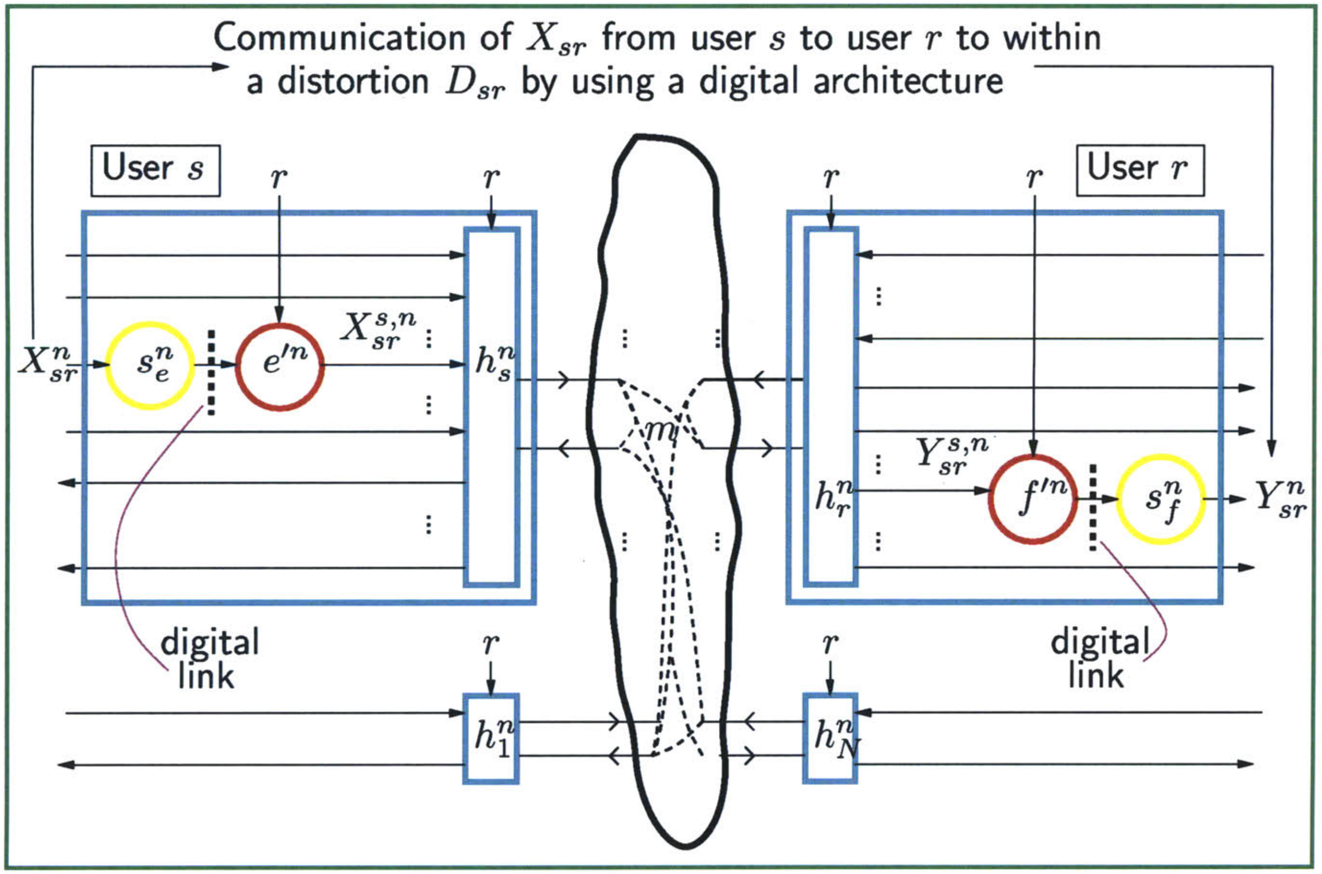


Figure 3.7. Source X_{sr} can be universally communicated from user s to user r to within a distortion D_{sr} over the partially known medium m by using a digital architecture, using the methodology described in Section 3.5

consumption of system resources is unchanged, and such that the communication of the rest of the sources in the system is unchanged *in distribution*?

We will answer this question in the affirmative by using the methodology of Section 3.5 by using the results from the point-to-point setting of Chapter 2.

For now, neglecting the communication of the rest of the sources, and thinking of the communication of source X_{sr} from user s to user r to within a probability of excess distortion D_{sr} as a point-to-point system, by using the results from Chapter 2, it follows that there exist source encoder-decoder $< s_e^n, s_f^n >_1^\infty$ and channel encoder-decoder $< e'^n, f'^n >_1^\infty$ (see Figure 3.7) such that the source X_{sr} is communicated, end-to-end, to within a distortion D_{sr} by use of a digital architecture. However, the question arises whether in this new architecture, the communication of the rest of the sources has been affected. The answer is that it has not been affected and this follows from the construction of e'^n. Recall from Chapter 2 that e'^n generates i.i.d. X_{sr} codes. Thus, the input to h_s^n in the new architecture which we denote by $X_{sr}^{s,n}$ has the same distribution as X_{sr}. From the arguments of Section 3.5, it follows that the communication of the rest of the sources has not been affected in distribution, nor has the consumption of system resources been increased. Since we do not require the exact knowledge of the medium, we are also solving the universal problem.This answers the question raised above in the affirmative.

Next assume that all sources X_{ij} are i.i.d. In general, source X_{ij} needs to be communicated from user i to user j to within a probability of excess distortion D_{ij}, $1 \leq i \leq N$. By comments from Section 3.5, it follows that the same procedure carried out from user s to user r can be carried out for any two pair of users.

The independence assumption on sources X_{ij} is required for reasons of Section 3.5.

Thus, optimality of digital communication for universal multi-user communication with fidelity criteria when the sources are independent of each other and random-coding is permitted, follows.

■ 3.8 Recapitulation

We have proved the optimality of digital communication for communication with fidelity criteria in multi-user settings. There are two crucial assumptions:

1. The sources which the various users want to communicate are independent of each other (the setting is unicast)

2. There is common randomness, that is, random-coding is permitted.

Without these assumptions, the results are false.

The proof is a simple generalization of the point-to-point universal source-channel separation theorem for rate-distortion discussed in Chapter 2: we do an induction over all source pairs in the network. This needs to be done with care, and we do this by maintaining the marginals at the inputs to the medium by using the methodology described in Section 3.5.

Since the proofs just build on the point-to-point proofs, many comments made in Chapter 2 also hold here, and we omit a discussion.

3.9 In the next chapter ...

In the next chapter, we discuss, to what extent, the results of this chapter are applicable to the traditional wireless telephony problem.

Chapter 4

Optimality of digital communication: Partial applicability of result from the previous chapter to the traditional wireless telephony problem

We should make things as simple as possible, but not simpler.

-Albert Einstein

■ 4.1 In this chapter ...

■ 4.1.1 Introduction

In this chapter, we see the application of results from the previous chapter to the problem of traditional wireless telephony. The results will only be partially applicable but lend some insight into the use of separation architectures in traditional wireless telephony. In the rest of this chapter, when we say wireless, we will be referring to the problem of traditional wireless telephony.

■ 4.1.2 Chapter outline

In Section 4.2, we discuss the features of traditional wireless telephony problem. We also prove that under certain assumptions that are not necessarily true, digital communication is optimal for the traditional wireless telephony problem. We do this to understand, to what extent digital communication is optimal for the traditional wireless telephony problem.

One crucial assumption that we make which is not true is that the voice signals of all the users are independent of each other. Why this assumption is not true is discussed in Section 4.2 in brief. In Section 4.3, we discuss a toy problem, an understanding of which can help understand what happens when correlated sources need to communicated over a medium and to what extent if at all, will separation hold in such a scenario.

In Section 4.4, we recapitulate this chapter.

4.2 Partial application to the wireless problem

In this section, we discuss the traditional wireless telephony problem and the partial applicability of the optimality of digital communication to this problem.

4.2.1 The wireless telephony problem

There are $2N$ users, $S_1, S_2, \ldots, S_N$, and $S'_1, S'_2, \ldots, S'_N$. For all i, users S_i and S'_i wish to talk to each other. The voice signal of user S_i is V_i and the voice signal of user S'_i is V'_i.

The question is: how does one design wireless architectures to maximize the number of users communicating at the same time over the wireless medium under certain constraints on resource consumption.

4.2.2 The features of the wireless problem and the assumptions that we make

The wireless problem has the following features:

1. *Pairwise independence of voice signals:* V_i is independent of V_j, V'_j for $j \neq i$. This is because what two users talk to each other is independent of what other users are talking amongst each other. They might be talking about the same subject. For example, if it is close to the elections or valentines day, a lot of conversations will revolve around these particular topic. What we are saying is that the conversation between a pair of users is independent of the conversation between another pair of users.

   However, V_i and V'_i are dependent of each other. This is because the conversation between two users will depend on what they are saying to each other.

   For our result on separation to be applicable, we require that all signals be independent of each other. In particular, we require that V_i be independent of V'_j for all j. As discussed above, this is not true for $j \neq i$. *This is a crucial assumption which is not true which we will make in order to prove the optimality of digital communication for the traditional wireless telephony problem.*

2. *The wireless medium is time varying and only partially known:* Wireless medium, that is, the atmosphere, changes with time and the exact operation of transmission of electromagnetic waves through the atmosphere might not be known. The communication methodology should work irrespective of the state of the atmosphere. Of course, there will be certain "very bad" states of the atmosphere under which the communication

will not be possible at all; however, we would want the communication to happen for some states of the atmosphere.

3. *Voice admits distortion:* Voice admits distortion in the sense that what the listener hears need not be exactly the same as what the speaker speaks in order for the listener to make out, what the speaker spoke.

4. *Other concerns:* There are other important features and concerns, for example, security and delay. Voice communication happens in real-time and only delays of the order of milli-seconds are permitted. Voice communication should happen securely. There are other concerns which we do not talk about here.

We make the following assumptions:

1. All the voice signals V_i, V'_j, $1 \leq i, j \leq N$ are independent of each other, not just pair-wise. As stated above, this is not true. However, we assume that this is the case.

2. Distortion measure on voice is permutation invariant

3. Delays do not matter

4.2.3 Optimality of digital communication for wireless

With the above assumptions, it follows from results from the previous chapter that separation holds: assuming random coding is permitted, it is sufficient to consider separation based architectures where each user first compresses its voice signal and this is followed by the universal reliable communication of the compressed voice signals over the wireless medium. There is sufficiency in the sense that if voice communication of certain number of users can be accomplished over the wireless medium with certain energy, bandwidth and other resource consumption, the same can be accomplished using a separation based architecture, too. In other words, there exists a "best possible" digital architecture.

4.3 The assumption that all the voice signals are independent

As stated above, we assumed that all voice signals V_i, V'_j are independent of each other. However, as we said, this is not true in that V_i and V'_i are not independent of each other.

To understand what happens when signals are dependent, it would be helpful to consider the problem only with two users s and s'. User s wants to communicate a source V to user s' with some distortion D under distortion metric d and user s' wants to communicate a source V' to S with some distortion D' under some distortion metric d'. *The sources V and V' may be dependent on each other.* See Figure 4.3. $\hat{V}$ denotes the reconstruction of V at user s' and $\hat{V}'$ denotes the reconstruction of V' at user s.

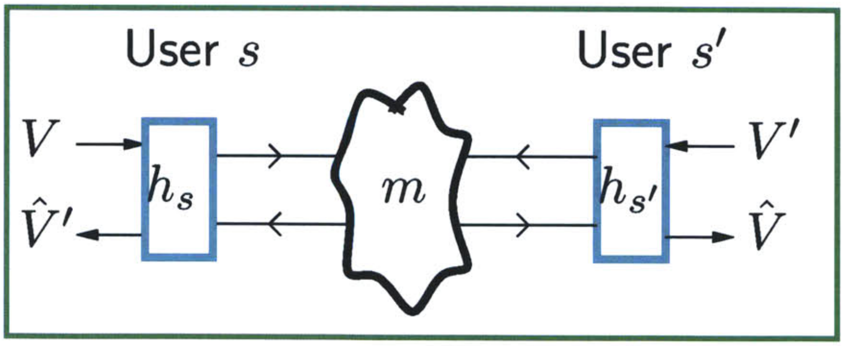


Figure 4.1. How do we build modems h_s and $h_{s'}$ for communication of sources V and V' which may be dependent? To what extent, if at all, and maybe approximately, does separation hold?

If this problem is understood, this would lend further insight into the ramifications of the dependence between V_i and V_i' in the wireless problem. As we discussed in Section 3.2, Gastpar, in [Gas02] discusses two examples where separation does not hold if the sources are correlated. We *do not* expect separation to hold in the example discussed above. However, we have *not* worked this out.

Even if separation does not hold, it would be insightful to understand, to what extent separation holds in the above example, and in general, when sources are correlated and the medium is only partially known. [TCDS] does consider the question of approximate optimality in the case when a user wants to communicate the same source (multi-cast) to various users to within different distortion levels. This is a special

As we stated in Section 3.2, [TCDS] proves, in certain scenarios, approximate optimality results when the sources are correlated with each other. It would be interesting to see if one can do the same for the setting described in this section. The setting in [TCDS] to prove approximate optimality results is very different from the example described in this section: [TCDS] has the the setting where a user wants to communicate the same source to within possibly different distortion levels to other users, and thus, this can be thought of as the case when sources at a particular user are perfectly correlated with each other. In particular, the correlated sources are at the *same* user. The example in this section has correlated sources at *different users*. In spite of these differences, it would be interesting to see if approximate optimality results can be proved for the example in this section, and for the most general scenario of correlated sources at multiple users.

4.4 Recapitulation

In this chapter, we proved the partial applicability of the optimality of digital communication to the traditional wireless telephony problem. This requires certain assumptions which are not true: the main such assumption is that the voice signals which users want to communicate to each other are independent of each other. Some of the other assumptions are that the distortion metric that can be put on voice are permutation invariant and that, delays do not matter, but we do not consider these serious assumptions compared to the assumption of voice signals being independent.

Because of the assumptions that we have to make which are not true, the applicability of our results to the wireless telephony problem is partial and not full.

4.5 In the next chapter ...

In the next chapter, we change gears. We discuss an operational perspective on the optimality of digital communication for universal communication with fidelity criteria. This was discussed partly in Chapter 2 and will be discussed further in the next chapter. I think of this as the second flavor of my thesis.

Chapter 5

Optimality of digital communication: operational view-point

Work on a proof until it is as evident as $2+2=4$.
-Paraphrased from a source which I do not remember

5.1 In this chapter ...

5.1.1 Introduction

In this chapter, we switch gears and discuss the second flavor of this thesis as discussed in Chapter 1: a rigorous operational view of the optimality of digital communication for communication with a fidelity criterion in the point-to-point setting. Since the proof for the multi-user setting was a simple generalization of the proof in the point-to-point setting, this also provides an operational view of the optimality of digital communication for communication with fidelity criteria in the multi-user setting.

In Section 2.14 of Chapter 2, we gave a proof of the universal source-channel separation for communication with a fidelity criterion holds in the point-to-point setting for the uniform X source when the distortion measure is permutation invariant. This proof was based on the assumption that $R_U^P(D) = R_U^P(D, \inf)$. This proof, as we said in Section 2.14 is operational: it uses only the definitions of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a certain distortion level. It does not use the definitions of channel capacity as a maximum mutual information and the rate-distortion function as a minimum mutual information. *Note that when we use the word operational, it does not have to do anything with something being physically operational or practical. What we mean is that we want to try to deal only with mathematical structures which reflect the real meaning of the quantities rather than simplified mathematical structures.*

One of the things that we will prove in this chapter is $R_U^P(D) = R_U^P(D, \inf)$. We do this operationally. This will just provide an operational perspective on universal source-channel separation without any technical assumptions. This proof will require steps which go through the i.i.d. X source. Some readers would question the use of the uniform X source. Traditional information theory literature uses the i.i.d. X source. In my opinion, the uniform X source captures all the ideas: it is simpler than an i.i.d. X source, in that it consists of only one type class and yet, is "close to" an i.i.d. X source because most of the probability of an i.i.d. X source rests on sequences with type "close to" p_X. In my opinion, this is enough.

However, for the reader unsatisfied with the uniform X source, we will generalize the results to the i.i.d. X source. However, for the i.i.d. X source, we will provide an operational proof of *pseudo-universal* source-channel separation for rate-distortion for additive distortion measures instead of universal source-channel separation for rate-distortion. Pseudo-universal differs from universal in that we will not require a uniformity in the rate at which probability of excess distortion $\rightarrow 0$ as block-length $\rightarrow \infty$ when communicating over the channel:

The partially known channel $k \in \mathscr{A}$ is said to be capable of pseudo-universally communicating i.i.d. X source to within a distortion D if there exists an encoder-decoder pair $< e^n, f^n >_1^\infty$ independent of the particular $k \in \mathscr{A}$ such that under the joint distribution $p_{X^nY^n}$ as described above,

$$p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D\right) \rightarrow 0 \text{ as } n \rightarrow \infty \; \forall \; k \in \mathscr{A} \tag{5.1}$$

The reader should compare this definition with the definition of a partially known channel which is capable of universally communicating i.i.d. X source to within a distortion level D in Definition 2.25: there is no $\omega =< \omega^n >_1^\infty$ sequence in the definition anymore which was introduced to enforce the uniformity over the partially known channel $k \in \mathscr{A}$ in the rate at which the probability of excess distortion $\rightarrow 0$ as $n \rightarrow \infty$. Now, we do not ask for this uniformity.

We defined in Chapter 2, a channel which is capable of communicating i.i.d. X source to within a distortion D. Note that we in fact defined a *partially* known channel which is capable of communicating i.i.d. X source to within a distortion D. However, let $k \in \mathscr{A}$ denote a partially known channel which is capable of universally communicating i.i.d. X source to within a distortion D. Thus, there exist encoder-decoder $< e^n, f^n >_1^\infty$ such that for the composition of the encoder, channel and decoder $< e^n \circ k \circ f^n >_1^\infty$, (2.25) holds for some $\omega =< \omega_n >_1^\infty$, $\omega_n \rightarrow 0$ as $n \rightarrow \infty$. We can think of $c \in \mathscr{C}_{\mathscr{A}} =< e^n \circ k \circ f^n >_1^\infty$ as a composite channel with input space $\mathscr{X}$ and output space $\mathscr{Y}$, and we can think of this partially known channel as *directly* communicating i.i.d. X source to within a distortion D.

We define $\mathscr{C}_{X,D}$ to be the set of channels which *pseudo*-directly communicates the i.i.d. X source to within a distortion D. $c \in \mathscr{C}_{X,D}$ can then be thought of as a partially known channel. Note that the ω sequence for different channels in $\mathscr{C}_{X,D}$ might be different, and for this reason we call it pseudo-direct communication and not direct communication.

We will define what we call the *pseudo*-universal capacity of the set of channel $\mathscr{C}_{X,D}$. The pseudo-universal capacity differs from universal capacity in the sense that we do not ask for a uniformity in the rate at which error probability $\to 0$ as block-length $n \to \infty$ over the set of channels; it can be different for different channels; of course, as in the definition of universal capacity, the same encoder-decoder should work for all channels in the set. Similarly, we will define what it means for a partially known channel $k \in \mathscr{A}$ to be capable of pseudo-universally communicating a random source to within a distortion level D: again, the only difference will be that we will not ask for uniformity in the rate at which the probability of excess distortion $\to 0$ as block-length $n \to \infty$ over the particular $k \in \mathscr{A}$; of course, as in the definition of universal capability of a channel to universally communicate a source to within a certain distortion level, the same encoder-decoder should work for all $k \in \mathscr{A}$.

We prove operationally that the pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$ is $\geq R_X^E(D) = R_X^P(D)$. From this, will follow the optimality of digital communication for communication with a fidelity criterion where we use pseudo-universal communication for communication to within a distortion D instead of universal communication to within a distortion D. We will call this the *An operational view-point on optimality of digital communication for pseudo-universal communication with a fidelity criterion* or *An operational view-point on pseudo-universal source-channel separation for rate-distortion.*

On the way, we will need to define the corresponding channel set for the uniform X source which we will denote by $\mathscr{C}_{U,D}$ and we will derive relation between pseudo-universal capacities of various channel sets and the various rate-distortion functions.

We believe that the above steps can also be carried out with universal capacity instead of pseudo-universal capacity with only minor modifications; however, we have not worked them out. For the reader unhappy with this explanation, the reader can think of this as an operational view of the *non-universal* view on the optimality of digital communication, that is the channel is a fully known channel: for a fully known channel, pseudo-universal capacity and the universal capacity are the same because uniformity is trivial for a channel set which consists of just one channel.

Another consideration that we do *not* have in this chapter is that of resource consumption. We prove the optimality of digital communication but do not look at the resource consumption in the digital architecture as opposed to a more general architecture. Again, this is something I believe is something that should be possible quite easily but I have not done it. For the reader unhappy with this explanation, the reader should think of it as the way things are done in the usual information theory literature in the discrete case where resource consumption is not considered at all. I should add that I do not agree with this approach in the literature because resource consumption is a very important issue, and when proving optimality of digital communication, one should prove that it can be done with the same or lesser resource consumption as compared to other architectures; in our case in this chapter, I quite strongly believe that it can be done; just that I have not done it.

We will in fact prove that the pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$ is precisely $R_X^P(D)$, and not just $\geq R_X^P(D)$. We will use this to give the idea for an alternate proof of the rate-distortion theorem for those i.i.d. sources X for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$. I believe that our proof is more insightful than the original proof of Shannon [Sha59]. This also leads to connections between source and channel coding which was talked about in brief Subsection 2.14.10, and is elaborated on a bit more in this chapter.

5.1.2 Chapter outline

In Section 5.3, we discuss the i.i.d. X and the uniform X sources: the two sources which we use in this chapter, and the to sources for which the theorems interplay with each other.

In Section 5.4, we discuss the rate-distortion source-coding problem. We define source-codes and what we call, jump source codes. The definitions of source-codes are made for coding both the i.i.d. X and the uniform X sources. We define various rate-distortion functions for the i.i.d. and the uniform X sources and prove the equality of all the rate-distortion functions. The proof of the equality of the various rate-distortion functions is operational.

In Section 5.5, we discuss the channel-coding problem. We define channels and what we call jump channels. We define, what we means for these channels to pseudo-directly communicate i.i.d. X and uniform X sources to within certain distortion levels, and correspondingly define various sets of channels or jump channels which pseudo-directly communicate the i.i.d. X and the uniform X sources to within certain distortion levels. We define the pseudo-universal capacities of these sets of channels and derive relations between these capacities. The derivation of these relations is operational.

In Section 5.6, we link the source-coding and the channel-coding problem. We prove that the pseudo-universal capacity of the set of channels which pseudo-directly communicate the uniform X source to within a distortion D is equal to the minimum rate required to compress the i.i.d. X source to within a distortion D. A similar statement is proved for the i.i.d. X source. The derivation of this result is operational.

In Section 5.7, we state the pseudo-universal source-channel separation theorem for rate-distortion and use the result of the previous section to finish the final step in proving operationally, the pseudo-universal source-channel separation theorem for communication with a fidelity criterion. Comments are made that we do not take into account resource consumption unlike Chapter 2, and that, we are operationally proving a pseudo-universal, not universal source-channel separation theorem for communication with a fidelity criterion. We also discuss the operational nature of the proof and compare our proof with Shannon's non-operational proof.

Out of our operational proof, also come out connections between source and channel coding. These connections are discussed In Section 5.8. Also, we provide an alternate proof of the rate-distortion theorem for those i.i.d. X sources for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Comments are made on why this proof is in our opinion, more insightful and more operational than Shannon's.

In this chapter, we have proved operationally, a pseudo-universal, not a universal source-channel separation theorem for rate-distortion. In Section 5.9, we comment briefly on what changes might be needed in order to prove a universal source-channel separation theorem for rate-distortion. Further, in this chapter, we have not taken into account resource consumption in the system when constructing the digital architecture. In Section 5.10, we comment very briefly on how we could also take into account, the resource consumption in the system, when proving the universal source-channel separation theorem for rate-distortion.

In Section 5.11, we recapitulate this chapter and make some final remarks on this chapter.

5.2 A note on definitions

We will, en-route, be re-defining many quantities, that have already been defined in Chapter 2. We do this for two reasons:

- To make this chapter complete in itself
- Some definitions like those of random source-codes and random channel-codes will look different from those of Chapter 2 but will be the same in spirit as those of Chapter 2

5.3 Sources

We consider 2 kinds of sources: i.i.d. and uniform. These are defined below.

Let $\mathscr{X}$ be a finite set.

Definition 5.1 (i.i.d. X source). Let X be a random-variable on $\mathscr{X}$. Let X^n denote i.i.d. X sequence of block-length n. X^n is a random-variable on $\mathscr{X}^n$. $< X^n >_1^\infty$ is the i.i.d. X source. By abuse of notation, we denote $X =< X^n >_1^\infty$.

Definition 5.2 (Uniform X source). Let X be a random variable on $\mathscr{X}$. *Let $p_X(x)$ be rational* $\forall x$. Let n_0 be the least positive integer for which $n_0 p_X(x)$ is an integer $\forall x \in \mathscr{X}$. Let $\mathscr{U}^n$ denote the set of sequences with (*exact*) empirical distribution (type) p_X. $\mathscr{U}^n$ is non-empty if and only if n_0 divides n. Let $n' \triangleq n_0 n$. Let $U^{n'}$ denote a random variable which is uniform on $\mathscr{U}^{n'}$ and zero elsewhere. Then, $< U^{n'} >_1^\infty$ is the uniform X source and is denoted by U. *Intuitively, the uniform X source is the source which puts uniform distribution on the set of all sequences whose empirical distribution is p_X.*

Note 5.1. The superscript n' in $\mathscr{U}^{n'}$ denotes that the block-length is n'. It *does not* mean that $\mathscr{U}^{n'}$ is the cartesian product of some set $\mathscr{U}$ with itself n times. Infact, the set $\mathscr{U} = \mathscr{U}^1$ is empty unless $n_0 = 1$. Similarly, the superscript n' in $U^{n'}$ denotes block-length. It *does not* mean that $U^{n'}$ is i.i.d. U source for some random variable U.

Definition 5.3 (n_0). n_0 is the least positive integer for which $n_0 p_X(x)$ is an integer $\forall x \in \mathscr{X}$.

Definition 5.4 (n'). $n' \triangleq n_0 n$.

Note 5.2. *Uniform X source is defined only for those block-lengths which are divisible by n_0.*

Note 5.3. If $p_X(x)$ is irrational for some $x \in \mathscr{X}$, $\mathscr{U}^n$ is empty $\forall n$. Thus, in order to define the uniform X source, the assumption that $p_X(x)$ be rational $\forall x \in \mathscr{X}$ is necessary.

Note 5.4. Let $p_X(x)$ be rational $\forall x \in \mathscr{X}$. The uniform X source and the i.i.d. X source are "close" to each other in the following sense. The uniform X source puts mass only only sequences with empirical distribution *exactly* p_X. For large n, i.i.d. X source puts "most of" its mass on sequences with empirical distribution "close to" p_X. We are interested in i.i.d. X source. Uniform X source is introduced only because some arguments can be made can rigorous for the uniform X source, which, we do not know, how to make rigorous for the i.i.d. X source.

■ 5.4 The rate-distortion problem

In this section, we discuss the rate-distortion problem of source-coding (compressing) the i.i.d. and the uniform X sources. We define source-codes and jump source-codes We define various rate-distortion functions for the i.i.d. X and the uniform X sources with the expected and the probability of excess distortion definitions, taking limits as lim inf and lim sup, and allowing jump source codes when source-coding the i.i.d. X source. We prove the continuity, convexity and the equality of various rate-distortion functions defined.

■ 5.4.1 Source codes

Source-codes are sequences for various block-lengths.

This subsection defines deterministic and random source-codes and jump source-codes for the i.i.d. X source, and deterministic and random source-codes for the uniform X source. The output space of the source-codes will be a finite set $\mathscr{Y}$.

Source codes to encode the i.i.d. X source

When the block-length is n, the input space is $\mathscr{X}^n$, the cartesian product of $\mathscr{X}$ with itself n times, and the output space is $\mathscr{Y}^n$, the cartesian product of $\mathscr{Y}$ with itself, n times.

Let $\mathscr{E}^n_{\mathscr{X}}(R)$ denote the set of all functions with domain X^n and range $\{1,2,\ldots,2^{\lfloor nR \rfloor}\}$. Let $\mathscr{F}^n_{\mathscr{Y}}(R)$ denote the set of all functions with domain $\{1,2,\ldots,2^{\lfloor nR \rfloor}\}$ and range $\mathscr{Y}^n$.

Definition 5.5 (Rate R deterministic source-code). Rate R deterministic source-code is a sequence $s =< s^n >_1^\infty =< e^n, f^n >_1^\infty$. $e^n \in \mathscr{E}^n_{\mathscr{X}}(R)$ and $f^n \in \mathscr{F}^n_{\mathscr{Y}}(R)$. This is interpreted as follows. When the block-length is n, $x^n \in \mathscr{X}^n$ is encoded as $e^n(x^n)$ and $a \in \{1,2,\ldots,2^{\lfloor knR \rfloor}\}$ is decoded as $f^n(a)$. $f^n(e^n(x^n))$ is the actual encoding of x^n by the source-code.

Note 5.5. In the definition of a rate R deterministic source-code, the set $\{1,2,\ldots,2^{\lfloor nR \rfloor}\}$ can be replaced by any set of cardinality $2^{\lfloor nR \rfloor}$.

Random source codes, which consist of a random source encoder and a random source decoder were defined in Section 2.7.2 as transition probabilities. In this chapter, we will view them in another, equivalent way: as a joint distribution on the space of deterministic source encoders and decoders

Definition 5.6 (Rate R random source-code with common randomness). A rate R random source-code with common randomness is a sequence $s =< s^n >_1^\infty$ which should be interpreted as follows. p_{s^n} is a probability distribution on $\mathscr{E}^n_{\mathscr{X}}(R) \times \mathscr{F}^n_{\mathscr{Y}}(R)$. That is, $p_{s^n} \in \mathscr{P}(\mathscr{E}^n_{\mathscr{X}}(R) \times \mathscr{F}^n_{\mathscr{Y}}(R))$. This is interpreted as follows. When block-length is n, the deterministic source-code $(e^n, f^n), e^n \in \mathscr{E}^n_{\mathscr{X}}(R), f^n \in \mathscr{F}^n_{\mathscr{Y}}(R)$ is used as the source-code with probability $p_{s^n}((e^n, f^n))$.

Note 5.6. If a source-code s has rate R, it also has rate $> R$.

Definition 5.7 (Transition probability corresponding to a source-code). Let $\mathscr{A}_{x^n,y^n}$ denote the set of all deterministic encoder-decoder pair (e^n, f^n) which encode x^n as y^n. That is, $\mathscr{A}_{x^n,y^n} \triangleq \{(e^n, f^n) | f^n(e^n(x^n)) = y^n\}$. $p_{s^n}(\mathscr{A}_{x^n,y^n})$ is the probability that x^n is encoded as y^n by the source-code s. This can be used to define a stochastic kernel/transition probability matrix $q_s^n : \mathscr{X}^n \to \mathscr{P}(\mathscr{Y}^n)$ as follows. $q_s^n(y^n|x^n) = p_{s^n}(\mathscr{A}_{x^n,y^n})$. $q_s^n(y^n|x^n)$ is the probability that x^n is encoded as y^n by the source code s. The sequence $q_s =< q_s^n >_1^\infty$ is the transition probability corresponding to a source-code $s =< s^n >_1^\infty$.

Note 5.7. The sequence $q_s =< q_s^n >_1^\infty$ is important because the distortion incurred by source-code s depends only on q_s. This will become clear later.

Definition 5.8 (Rate R deterministic jump source-code). Rate R deterministic source-code is a sequence $s =< s^{kn} >_1^\infty =< e^{kn}, f^{kn} >_1^\infty$ where $k > 0$ is some positive integer, where definitions of e^{kn}, f^{kn} and the interpretation of the source-code is exactly as in Definition 5.5. The only difference is that quantitities in a jump block-code are defined only for block-lengths kn, $1 \le n \le \infty$.

Note 5.8. In the above definition, when we say $< kn >_1^\infty$ refers to $< kn >_{n=1}^\infty$. k remains fixed. This will be the case throughout the rest of the thesis: for $n' = f(n)$, $< a^{n'} >_1^\infty$ refers to $< a^{f(n)} >_{n=1}^\infty$.

Definition 5.9 (Rate R random jump source-code). This is defined exactly as definition 5.6, except that quantities are defined only for block lengths kn, $1 \le n \le \infty$, for some positive integer k.

Note 5.9. Jump source-codes *do not* have physical significance. We define them because proof of some equalities concerning rate-distortion functions require the introduction of jump source-codes. This will be clarified in more detail, later.

Definition 5.10 (Transition probability corresponding to a a jump source-code). This is exactly the same as Definition 5.7, except that quantities are defined only for block-lengths kn, $1 \leq n \leq \infty$, for some positive integer k.

Source-codes to encode the uniform X source

Recall the definitions of n_0 and n'.

When the block-length is n', the input space is $\mathscr{U}^{n'}$. The output space is $\mathscr{Y}^{n'}$, the cartesian product of $\mathscr{Y}$ with itself, n' times.

Let $\mathscr{E}_{\mathscr{U}}^{n'}(R)$ denote the set of all functions with domain $\mathscr{U}^{n'}$ and range $\{1,2,\ldots,2^{\lfloor n'R \rfloor}\}$. Let $\mathscr{F}_{\mathscr{Y}}^{n'}(R)$ denote the set of all functions with domain $\{1,2,\ldots,2^{\lfloor n'R \rfloor}\}$ and range $Y^{n'}$.

Definition 5.11 (Rate R deterministic source-code). Rate R deterministic source-code is a sequence $s =< s^{n'} >_1^{\infty} =< e^{n'}, f^{n'} >_1^{\infty}$. $e^{n'} \in \mathscr{E}_{\mathscr{U}}^{n'}(R)$ and $f^{n'} \in \mathscr{F}_{\mathscr{Y}}^{n'}(R)$. This is interpreted as follows. When the block-length is n', $u^{n'} \in \mathscr{U}^{n'}$ is encoded as $e^{n'}(u^{n'})$ and $a \in \{1,2,\ldots,2^{\lfloor n'R \rfloor}\}$ is decoded as $f^{n'}(a)$. $f^{n'}(e^{n'}(x^{n'}))$ is the actual encoding of $x^{n'}$ by the source-code.

Note 5.10. In the definition of a rate R deterministic source-code, the set $\{1,2,\ldots,2^{\lfloor n'R \rfloor}\}$ can be replaced by any set of cardinality $2^{\lfloor n'R \rfloor}$.

Definition 5.12 (Rate R random source-code with common randomness). A rate R random source-code with common randomness is a sequence $s =< s^{n'} >_1^{\infty}$ which should be interpreted as follows. $p_{s^{n'}}$ is a probability distribution on $\mathscr{E}_{\mathscr{U}}^{n'}(R) \times \mathscr{F}_{\mathscr{Y}}^{n'}(R)$. That is, $p_{s^{n'}} \in \mathscr{P}(\mathscr{E}_{\mathscr{U}}^{n'}(R) \times \mathscr{F}_{\mathscr{Y}}^{n'}(R))$. This is interpreted as follows. When block-length is n', the deterministic source-code $(e^{n'}, f^{n'}), e^{n'} \in \mathscr{E}_{\mathscr{U}}^{n'}(R), f^{n'} \in \mathscr{F}_{\mathscr{Y}}^{n'}(R)$ is used as the source-code with probability $p_{s^{n'}}((e^{n'}, f^{n'}))$.

Note 5.11. Since the uniform source is defined only for block-lengths $n' = n_0 n$, source-codes to encode the uniform X source are also defined only for block-lengths $n' = n_0 n$. This *does not* mean that these source-codes are jump source-codes.

Note 5.12. If a source-code s has rate R, it also has rate $> R$.

Definition 5.13 (Transition probability corresponding to a source-code). Let $\mathscr{A}_{u^{n'}, y^{n'}}$ denote the set of all deterministic encoder-decoder pair $(e^{n'}, f^{n'})$ which encode $u^{n'}$ as $y^{n'}$. That is, $\mathscr{A}_{u^{n'}, y^{n'}} \triangleq \{(e^{n'}, f^{n'}) | f^{n'}(e^{n'}(u^n)) = y^{n'}\}$. $p_{s^{n'}}(\mathscr{A}_{u^{n'}, y^{n'}})$ is the probability that $u^{n'}$ is encoded as $y^{n'}$ by the source-code s. This can be used to define a stochastic kernel/transition probability matrix $q_s^{n'} : \mathscr{U}^n \to \mathscr{P}(\mathscr{Y}^{n'})$ as follows. $q_s^{n'}(y^{n'} | u^{n'}) = p_{s^{n'}}(\mathscr{A}_{u^{n'}, y^{n'}})$. $q_s^{n'}(y^{n'} | u^{n'})$ is the probability that $u^{n'}$ is encoded as $y^{n'}$ by the source code s. The sequence $q_s =< q_s^{n'} >_1^{\infty}$ is the transition probability corresponding to a source-code $s =< s^{n'} >_1^{\infty}$.

Note 5.13. The sequence $q_s =< q_s^n >_1^\infty$ is important because the distortion incurred by source-code s depends only on q_s. This will become clear later.

Note 5.14. We do not need to define jump source-codes for encoding the uniform X source.

■ 5.4.2 Distortion produced by a source-code and a jump source-code

In this subsection, we define the distortion produced by a source-code. We consider two definitions of distortion: the expected distortion and the probability of excess distortion. Both the expected distortion and the probability of excess distortion can be defined by taking limits in two ways: lim inf and lim sup. When the source is i.i.d. X, distortion is defined for both source-codes and jump source-codes. For the uniform X source, distortion is defined only for source-codes: we would not need the definition with jump source-codes.

Assume that $\mathcal{X} = \mathcal{Y}$. *This assumption will be used throughout the rest of this chapter.*

$d : \mathcal{X} \times \mathcal{Y} \to [0,\infty)$ is the distortion function. Let $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. $d(x,y)$ should be thought of as the distortion between x and y, or as the distortion incurred if x is received as y.

Let d be such that $d(x,x) = 0 \forall x \in \mathcal{X}$. This assumption will be used throughout the rest of this chapter.

For $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, the distortion between x^n and y^n is defined additively:

$$d^n(x^n, y^n) \triangleq \sum_{i=1}^{n} d(x^n(i), y^n(i)) \tag{5.2}$$

Denote $D_{\max} \triangleq \max_{x\in\mathcal{X}, y\in\mathcal{Y}} d(x,y)$.

Distortion produced by a source code which encodes i.i.d. X source

Consider a source-code $s =< s^n >_1^\infty$.

The action of s^n on source X^n results in a joint probability distribution on the input-output $\mathcal{X}^n \times \mathcal{Y}^n$ space. This action is that of the kernel q_s^n on the source X^n. The output random-variable is Y^n. The joint random-variable on the $\mathcal{X}^n \times Y^n$ space is $X^n Y^n$.

Definition 5.14. [Achievability of expected distortion D by source-code s when encoding i.i.d. X Source] Distortion D is achievable in the expected sense (or that, distortion D is E-achievable, or that expected distortion D is achievable) by the source-code s for the i.i.d. X source if there exists

$$\limsup_{n\to\infty} \mathrm{E}_{X^n Y^n}\left[\frac{1}{n} d^n(X^n, Y^n)\right] \leq D \tag{5.3}$$

Note 5.15. By definition, if distortion D is achievable in the expected sense (or that, distortion D is P-achievable) by the source-code s for the i.i.d. X source, then distortion $D' > D$ is also achievable in the expected sense by the source-code s for the i.i.d. X source.

Definition 5.15. [Achievability of probability of excess distortion D by source-code s when encoding i.i.d. X Source] Distortion D is achievable in the probability of excess distortion sense (or that, distortion D is P-achievable) by the source-code s for the i.i.d. X source if

$$\lim_{n\to\infty} p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D\right) = 0 \tag{5.4}$$

Note 5.16. In the above definition, we use $\lim\limits_{n\to\infty}$ and not $\limsup\limits_{n\to\infty}$ because both definitions are the same. This is because, if $a_n \geq 0, 1 \leq n < \infty$, then, $\limsup\limits_{n\to\infty} a_n = 0$ if and only if $\lim\limits_{n\to\infty} a_n = 0$.

Note 5.17. By definition, if distortion D is achievable in the probability of excess distortion sense by the source-code s for the i.i.d. X source, then distortion $D' > D$ is also achievable in the probability of excess distortion sense by the source-code s for the i.i.d. X source.

Definition 5.16. [Inf-achievability of expected distortion D by source-code s when encoding i.i.d. X Source] Expected distortion D is *inf*-achievable by the source-code s for i.i.d. X source if

$$\liminf_{n\to\infty} \mathrm{E}_{X^nY^n}\left[\frac{1}{n}d^n(X^n,Y^n)\right] \leq D \tag{5.5}$$

Note 5.18. By definition, if expected distortion D is inf-achievable by the source-code s for the i.i.d. X source, then expected distortion $D' > D$ is also inf-achievable by the source-code s for the i.i.d. X source.

Note 5.19. When we say that expected distortion D is achievable, we would mean that expected distortion D is achievable with Definition 5.14. When we want to talk about achievability of probability of excess distortion D in the sense of Definition 5.16, *we would explicitely refer to it as inf-achievability.*

Definition 5.17. [inf-achievability of probability of excess distortion D by source-code s when encoding i.i.d. X Source] Probability of excess distortion D is *inf*-achievable by the source-code s for i.i.d. X source if

$$\liminf_{n\to\infty} p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D\right) = 0 \tag{5.6}$$

Note 5.20. By definition, if probability of excess distortion D is inf-achievable by the source-code s for the i.i.d. X source, then probability of excess distortion $D' > D$ is also inf-achievable by the source-code s for the i.i.d. X source.

Note 5.21. When we say that expected distortion D is achievable, we would mean that probability of excess distortion D is achievable with Definition 5.15. When we want to talk about achievability of probability of excess distortion D in the sense of Definition 5.17, *we would explicitely refer to it as inf-achievability.*

Distortion produced by a jump-source code which encodes i.i.d. X source

Consider a jump source-code $s =< s^{kn} >_1^\infty$. *Denote* $n'' = kn$

The action of $s^{n''}$ on source $X^{n''}$ results in a joint probability distribution on the input-output $\mathcal{X}^{n''} \times \mathcal{Y}^{n''}$ space. This action is that of the kernel $q_s^{n''}$ on the source $X^{n''}$. The output random-variable is $Y^{n''}$. The joint random-variable on the $\mathcal{X}^{n''} \times Y^{n''}$ space is $X^{n''}Y^{n''}$.

Definition 5.18. [Achievability of expected distortion D by the jump source-code s when encoding i.i.d. X Source] This definition is the same as Definition 5.14 except that limits are taken along n''instead of n.

Note 5.22. Note 5.15 holds for jump source-codes.

Definition 5.19. [Achievability of probability of excess distortion D by jump source-code s when encoding i.i.d. X Source] This definition is the same as definition 5.15 except that limits are taken along n'' instead of n.

Note 5.23. Note 5.16 holds for jump block-codes.

Note 5.24. Note 5.17 holds for jump source-codes.

Note 5.25. We do not talk about inf-achievability with jump block codes. This is because of the following reason. Suppose we are given a jump source-code $s =< s^{kn} >_1^\infty$ for which we want to define inf-achievability. Thus, there would be some subsequence kn_i along which limits are taken. Now, consider a source-code $t =< t^n >_1^\infty$ which is the same as source-code s for block-lengths kn and arbitrarily defined, for other block-lengths. Limits can then be taken along the same block-lengths kn_i for this code without jump. Thus, for the definition of inf-achievability, a jump code can first be converted into a code without jump by the above procedure. Thus, one does not gain anything by allowing jumps when considering inf-achievabilty.

Distortion produced by a source-code which encodes uniform X source

Let X be such that $p_X(x)$ is rational $\forall x \in \mathcal{X}$.

Recall that the uniform X source is defined only for block lengths which are multiples of n_0. Recall that $n' \triangleq n_0 n$. Also, source-codes for coding the uniform X source are defined only for block-lengths n'.

Consider a source-code $s =< s^{n'} >_1^\infty$.

The action of $s^{n'}$ on source $U^{n'}$ results in a joint probability distribution on the input-output $\mathcal{U}^{n'} \times \mathcal{Y}^{n'}$ space. This action is that of the kernel $q_s^{n'}$ on the source $U^{n'}$. The output random-variable is $Y^{n'}$. The joint random-variable on the $\mathcal{U}^{n'} \times \mathcal{Y}^{n'}$ space is $U^{n'}Y^{n'}$.

Definition 5.20. [Achievability of expected distortion D by source-code s when encoding uniform X Source] This definition is the same as Definition 5.14 except that limits are taken along n', and that, expectation is taken with respect to the joint random variable $U^{n'}Y^{n'}$.

Note 5.26. Note 5.26 holds for the uniform X source.

Definition 5.21. [Achievability of probability of excess distortion D by source-code s when encoding uniform X Source] This definition is the same as Definition 5.15 except that limits are taken along n', and that, probability is taken with respect to the joint random variable $U^{n'}Y^{n'}$.

Note 5.27. Note 5.16 holds.

Note 5.28. Note 5.17 holds for the uniform X source.

Definition 5.22. [Inf-achievability of expected distortion D by source-code s when encoding uniform X Source] This definition is the same as Definition 5.16 except that limits are taken along n', and that, expectation is taken with respect to the joint random variable $U^{n'}Y^{n'}$.

Note 5.29. Note 5.18 holds for the uniform X source.

Note 5.30. When we say that expected distortion D is achievable, we would mean that expected distortion D is achievable with Definition 5.20. When we want to talk about achievability of probability of excess distortion D in the sense of Definition 5.22, *we would explicitely refer to it as inf-achievability.*

Definition 5.23. [inf-achievability of probability of excess distortion D by source-code s when encoding uniform X Source] This definition is the same as Definition 5.17 except that limits are taken along n', and that, probability is taken with respect to the joint random variable $U^{n'}Y^{n'}$.

Note 5.31. Note 5.31 holds for the uniform X source.

Note 5.32. When we say that expected distortion D is achievable, we would mean that probability of excess distortion D is achievable with Definition 5.21. When we want to talk about achievability of probability of excess distortion D in the sense of Definition 5.23, *we would explicitely refer to it as inf-achievability.*

Note 5.33. As stated before, jump source-codes are not considered, when encoding the uniform X source.

Note 5.34. Recall the assumptions that $\mathcal{X} = \mathcal{Y}$ and $d(x,x) = 0 \forall x \in \mathcal{X}$. By this assumption, distortion $D = 0$ is achievable for all the above definitions, and $R_X^E(0)$, $R_X^P(0)$, $R_X^E(0,\text{inf})$, $R_X^P(0,\text{inf})$, $R_X^E(0,j)$, $R_X^P(0,j)$, $R_U^E(0)$, $R_U^P(0)$, $R_U^E(0,\text{inf})$, and $R_U^P(0,\text{inf})$, are all $\leq \log|\mathcal{X}|$. In particular, $R_X^E(0)$, $R_X^P(0)$, $R_X^E(0,\text{inf})$, $R_X^P(0,\text{inf})$, $R_X^E(0,j)$, $R_X^P(0,j)$, $R_U^E(0)$, $R_U^P(0)$, $R_U^E(0,\text{inf})$, and $R_U^P(0,\text{inf})$ are all defined for $D \in [0,\infty)$.

5.4.3 The rate-distortion function

In this subsection, we define the rate-distortion function when encoding the i.i.d. X and the uniform X sources. The rate-distortion function can be defined for the expected distortion definition and the probability of excess distortion definition, under both the lim inf and the lim sup definitions.

For i.i.d. X sources, the rate-distortion function can be defined when jump source-codes are allowed or when only source-codes are allowed. This leads to six different definitions of the rate-distortion function when encoding the i.i.d. X source: $R^E_X(D)$, $R^E_X(D, \inf)$, $R^E_X(D, j)$, $R^P_X(D)$, $R^P_X(D, \inf)$, and$R^P_X(D, j)$.

For the uniform X source, rate-distortion function is defined when source-codes are allowed. This leads to four different definitions of the rate-distortion function when encoding the uniform X source: $R^E_U(D)$, $R^E_U(D, \inf)$, $R^P_U(D)$, and $R^P_U(D, \inf)$

Rate-distortion function corresponding to source-codes which encode the i.i.d. X source

Definition 5.24 (Rate-distortion function $R^E_X(D)$). Rate R is E-achievable corresponding to distortion level D for the i.i.d. X source if there exists a rate R source code s which achieves expected distortion D when encoding the i.i.d. X source. The infimum of all E-achievable rates for distortion level D is the rate-distortion function $R^E_X(D)$.

Definition 5.25 (Rate-distortion function $R^P_X(D)$). Rate R is P-achievable corresponding to distortion level D for the i.i.d. X source if there exists a rate R source code s which achieves probability of excess distortion D for the i.i.d. X source. The infimum of all P-achievable rates for distortion level D is the rate-distortion function $R^P_X(D)$.

Definition 5.26 (Rate-distortion function $R^E_X(D, \inf)$). Rate R is inf-E-achievable for the i.i.d. X source if there exists a rate R source code s which inf-achieves expected distortion D for the i.i.d. X source. The infimum of all E-achievable rates for distortion level D is the rate-distortion function $R^E_X(D, \inf)$.

Definition 5.27 (Rate-distortion function $R^P_X(D, \inf)$). Rate R is inf-P-achievable for the i.i.d. X source if there exists a rate R source code s which inf-achieves probability of excess distortion D for the i.i.d. X source. The infimum of all E-achievable rates for distortion level D is the rate-distortion function $R^P_X(D, \inf)$.

Rate-distortion function corresponding to jump source-codes which encode the i.i.d. X source

Definition 5.28 (Rate-distortion function $R^E_X(D, j)$). Rate R is E-achievable corresponding to distortion level D for the i.i.d. X source if there exists a rate R jump source code s which achieves expected distortion D when encoding the i.i.d. X source. Jump source-codes with all possible jump sizes k are under consideration. The infimum of all E-achievable rates for distortion level D is the rate-distortion function $R_X^E(D,j)$.

Definition 5.29 (Rate-distortion function $R_X^P(D,j)$). Rate R is P-achievable corresponding to distortion level D for the i.i.d. X source if there exists a rate R jump source code s which achieves probability of excess distortion D for the i.i.d. X source. Jump source-codes with all possible jump sizes k are under consideration. The infimum of all P-achievable rates for distortion level D is the rate-distortion function $R_X^P(D,j)$.

Rate-distortion function corresponding to source-codes which encode the uniform X source

Definition 5.30 (Rate-distortion function $R_U^E(D)$). $R_U^E(D)$ is defined analogously to $R_X^E(D)$.

Definition 5.31 (Rate-distortion function $R_U^P(D)$). $R_U^P(D)$ is defined analogously to $R_X^P(D)$.

Definition 5.32 (Rate-distortion function $R_U^E(D,\text{inf})$). $R_U^E(D,\text{inf})$ is defined analogously to $R_X^E(D,\text{inf})$.

Definition 5.33 (Rate-distortion function $R_U^P(D,\text{inf})$). $R_U^P(D,\text{inf})$ is defined analogously to $R_X^P(D,\text{inf})$.

5.4.4 Properties and equalities of the various rate-distortion functions for i.i.d. and uniform sources

In this subsection, we prove various the convexity, continuity, and equality of various rate-distortion functions. Our goal is to prove 2 results

1. For X such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$, for $D \in (0,\infty)$, $R_X^E(D)$, $R_X^P(D)$, $R_X^E(D,j)$, $R_X^P(D,j)$, $R_X^E(D,\text{inf})$, $R_X^P(D,\text{inf})$, $R_U^E(D)$, $R_U^P(D)$, $R_U^E(D,\text{inf})$, and $R_U^P(D,\text{inf})$ are convex, continuous functions of D and are all equal
2. For arbitrary X, for $D \in (0,\infty)$, $R_X^E(D)$, $R_X^E(D,j)$, $R_X^E(D,\text{inf})$, $R_X^P(D)$, $R_X^P(D,j)$, and $R_X^P(D,\text{inf})$ are all convex continuous functions of D and are all equal

Proving the above will be carried out in various steps which the reader might want to refer to now, or in the future, as the proofs progress. The statements below are *not* precise, in particular, they do not mention the range of D for which the results hold.

1. Results concerning achievability with deterministic source-codes
2. Construction of time-sharing code. As a consequence, the proof of convexity and continuity of $R_X^E(D,j)$ and $R_X^P(D,j)$

3. Construction of interpolation code. As a consequence, the proof of $R_X^E(D) = R_X^E(D,j)$ and $R_X^P(D) = R_X^P(D,j)$

4. Construction of jump repetition code. As a consequence, the proof of $R_X^E(D,\inf) = R_X^E(D,j)$ and $R_X^P(D,\inf) = R_X^P(D,j)$

5. Thus, $R_X^E(D) = R_X^E(D,j) = R_X^E(D,\inf)$ and $R_X^P(D) = R_X^P(D,j) = R_X^P(D,\inf)$

6. Description of two constructions describing how to construct a jump source-code for i.i.d. X' source given a source-code for the uniform X source, and how to construct a source-code for the uniform X source given a source-code for the i.i.d. X' source. As a consequence, the proofs of $R_X^E(D) = R_U^E(D)$, $R_X^P(D) = R_U^P(D)$, $R_X^E(D,\inf) = R_U^E(D,\inf)$ and $R_X^P(D,\inf) = R_U^P(D,\inf)$.

7. Proof of $R_U^E(D) = R_U^P(D)$

8. Thus, the proof of equality of $R_X^E(D)$, $R_X^P(D)$, $R_X^E(D,j)$, $R_X^P(D,j)$, $R_X^E(D,\inf)$, $R_X^P(D,\inf)$, $R_U^E(D)$, $R_U^P(D)$, $R_U^E(D,\inf)$, and $R_U^P(D,\inf)$

9. The proof of equality of $R_X^E(D)$, $R_X^E(D,j)$, $R_X^E(D,\inf)$, $R_X^P(D)$, $R_X^P(D,j)$, and $R_X^P(D,\inf)$ by use of limiting arguments

First we prove that there is no loss of generality if one restricts attention to deterministic source-codes and jump source-codes.

Achievability with deterministic codes

We prove that if rate R is E- achievable with expected distortion D for a source, rate R is also E achievable with distortion D with a deterministic source-code for the source. Similarly, if rate R is P-achievable with distortion D for a source, rate R is also P-achievable with probability of excess distortion D with a deterministic source-code.

Lemma 5.1. *1. Let expected distortion D be achievable for the i.i.d. X source with some rate R source code. Then, expected distortion D is also achievable for the i.i.d. X source with a rate R deterministic source-code*

2. Let expected distortion D be inf-achievable for the i.i.d. X source with some rate R source code. Then, expected distortion D is also inf-achievable for the i.i.d. X source with a rate R deterministic source-code

3. Let expected distortion D be achievable for the i.i.d. X source with some rate R jump source code. Then, expected distortion D is also achievable for the i.i.d. X source with a rate R deterministic jump source-code

4. *Let expected distortion D be achievable for the uniform X source with some rate R source code. Then, expected distortion D is also achievable for the uniform X source with a rate R deterministic source-code*

5. *Let expected distortion D be inf-achievable for the uniform X source with some rate R source code. Then, expected distortion D is also inf-achievable for the uniform X source with a rate R deterministic source-code*

Proof. We prove the first statement above. The proofs of the rest of the statements are similar; the only difference is that limits might be taken along particular block-lengths or that, the source might be different.

Let expected distortion D be achievable for the i.i.d. X source with rate R source-code $s =< s^n >_1^\infty$. That is,

$$\mathrm{E}_{X^nY^n}\left[\frac{1}{n}d^n(X^n,Y^n)\right] = D_n \text{ where } \limsup_{n\to\infty} D_n \leq D \tag{5.7}$$

$$\begin{aligned}
&\mathrm{E}_{X^nY^n}\left[\frac{1}{n}d^n(X^n,Y^n)\right] \\
&= \sum_{(e^n,f^n)\in\mathcal{E}^n_{\mathcal{X}}(R)\times\mathcal{F}^n_{\mathcal{X}}(R)} p_{s^n}((e^n,f^n))\mathrm{E}_{X^nY^n}\left[\frac{1}{n}d^n(X^n,Y^n)|s^n=(e^n,f^n)\right] \\
&= \sum_{(e^n,f^n)\in\mathcal{E}^n_{\mathcal{X}}(R)\times\mathcal{F}^n_{\mathcal{X}}(R)} p_{s^n}((e^n,f^n))\mathrm{E}_{X^n}\left[\frac{1}{n}d^n(X^n,f^n(e^n(X^n)))\right] \leq D_n
\end{aligned} \tag{5.8}$$

Thus, there exists an $(e^n_*,f^n_*)\in\mathcal{E}^n_{\mathcal{X}}(R)\times\mathcal{F}^n_{\mathcal{X}}(R)$ such that

$$\mathrm{E}_{X^n}\left[\frac{1}{n}d^n(X^n,f^n_*(e^n_*(X^n)))\right] \leq D_n \tag{5.9}$$

For the deterministic source-code $< e^n_*,f^n_* >_1^\infty$,

$$\lim_{n\to\infty}\mathrm{E}_{X^n}\left[\frac{1}{n}d^n(X^n,f^n_*(e^n_*(X^n)))\right] \leq \lim_{n\to\infty} D_n \leq D \tag{5.10}$$

Since s has rate R, $< e^n_*,f^n_* >_1^\infty$ also has rate R. $< e^n_*,f^n_* >_1^\infty$ is thus a deterministic rate R source-code with expected distortion D for the i.i.d. X source. □

Note 5.35. It follows from the above lemma that the rate-distortion function $R^E_X(D)$, $R^E_X(D,\text{inf})$, $R^E_X(D,j)$, $R^E_U(D)$, and $R^E_U(D,\text{inf})$ are unchanged by restriction to deterministic source-codes.

Note 5.36. With appropriate definitions, the above lemma infact holds for arbitrary sources, not just i.i.d. and uniform.

Lemma 5.2. *1. Let probability of excess distortion D be achievable for the i.i.d. X source with some rate R source code. Then, probability of excess distortion D is also achievable for the i.i.d. X source with a rate R deterministic source-code*

2. *Let probability of excess distortion D be inf-achievable for the i.i.d. X source with some rate R source code. Then, probability of excess distortion D is also inf-achievable for the i.i.d. X source with a rate R deterministic source-code*

3. *Let probability of excess distortion D be achievable for the i.i.d. X source with some rate R jump source code. Then, probability of excess distortion D is also achievable for the i.i.d. X source with a rate R deterministic jump source-code*

4. *Let probability of excess distortion D be achievable for the uniform X source with some rate R source code. Then, probability of excess distortion D is also achievable for the uniform X source with a rate R deterministic source-code*

5. *Let probability of excess distortion D be inf-achievable for the uniform X source with some rate R source code. Then, probability of excess distortion D is also inf-achievable for the uniform X source with a rate R deterministic source-code*

Proof. We prove the first statement above. The proofs of the rest of the statements are similar; the only difference is that limits might be taken along particular block-lengths or that, the source might be different (or jump source-codes, as the case may be).

Let probability of excess distortion D be achievable for the i.i.d. X source with rate R source-code $s =< s^n >_1^\infty$. That is,

$$p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D\right) = \epsilon_n \text{ where } \lim_{n\to\infty}\epsilon_n = 0 \tag{5.11}$$

$$\begin{aligned}
&p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D\right)\\
=&\sum_{(e^n,f^n)\in\mathscr{E}^n_{\mathscr{X}}(R)\times\mathscr{F}^n_{\mathscr{X}}(R)} p_{s^n}((e^n,f^n))p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D|s^n=(e^n,f^n)\right)\\
=&\sum_{(e^n,f^n)\in\mathscr{E}^n_{\mathscr{X}}(R)\times\mathscr{F}^n_{\mathscr{X}}(R)} p_{s^n}((e^n,f^n))p_{X^n}\left(\frac{1}{n}d^n(X^n,f^n(e^n(X^n))) > D\right)\\
=&\epsilon_n
\end{aligned} \tag{5.12}$$

Thus, there exists an $(e^n_*,f^n_*)\in\mathscr{E}^n_{\mathscr{X}}(R)\times\mathscr{F}^n_{\mathscr{X}}(R)$ such that

$$p_{X^n}\left(\frac{1}{n}d^n(X^n,f^n_*(e^n_*(X^n))) > D\right) \le \epsilon_n \tag{5.13}$$

For the deterministic source-code $< e_*^n, f_*^n >_1^\infty$,

$$\lim_{n\to\infty} p_{X^n}\left(\frac{1}{n}d^n(X^n, f_*^n(e_*^n(X^n))) > D\right) \leq \lim_{n\to\infty} \epsilon_n = 0 \tag{5.14}$$

Since s has rate R, $< e_*^n, f_*^n >_1^\infty$ also has rate R. $< e_*^n, f_*^n >_1^\infty$ is thus a deterministic rate R source-code with probability of excess distortion D for the i.i.d. X source. □

Note 5.37. It follows from the above lemma that the rate-distortion function $R_X^P(D)$, $R_X^P(D, \inf)$, $R_X^P(D, j)$, $R_U^P(D)$, and $R_U^P(D, \inf)$ are unchanged by restriction to deterministic source-codes(or jump source-codes, as the case may be).

Note 5.38. With appropriate definitions, the above lemma infact holds for arbitrary sources, not just i.i.d. and uniform.

Next we prove theconvexity and continuity of various rate-distortion functions for $D \in (0,\infty)$.

Convexity and continuity of rate-distortion functions

We prove the convexity of $R_X^E(D, j)$ and $R_X^P(D, j)$ for $D \in [0,\infty)$. As a consequence, will follow, the continuity of $R_X^E(D, j)$ and $R_X^P(D, j)$ for $D \in (0,\infty)$. For this, we first define equal time sharing between jump source-codes.

Definition 5.34 (Equal time sharing between deterministic jump source-codes which code the i.i.d. X source). Let $s =< s^{kn} >_1^\infty =< e^{kn}, f^{kn} >_1^\infty$ be a rate R deterministic jump source-code to code the i.i.d. X source. Let $s' =< s'^{k'n} >_1^\infty$ be a rate R' deterministic source-code to code the i.i.d. X source . The jump source-code code $t =< t^{2kk'n} >_1^\infty =< g^{2kk'n}, h^{2kk'n} >_1^\infty$ which time-shares *equally* between s and s' is defined as follows

Denote $n'' = kk'n$.

Let

$$\begin{aligned} g^{2n''}(x^{2n''}) \triangleq (&e^{kn}(x(1..kn)), e^{kn}(x(kn+1..2kn)), \dots, e^{kn}(x(k(k'-1)n+1..n'')),\\ &e'^{k'n}(x(n''..n''+k'n)), e'^{k'n}(x(n''+k'n+1..n''+2k'n)), \dots,\\ &e'^{k'n}(x(n''+k'(k-1)n..2n''))) \end{aligned} \tag{5.15}$$

The range of $g^{2n''}$ is $\{1,2,\dots 2^{\lfloor knR \rfloor}\}^{k'} \times \{1,2,\dots 2^{\lfloor k'nR' \rfloor}\}^k$. For $a^{k'+k}$, a vector of length $k'+k$, $\in \{1,2,\dots 2^{\lfloor nR \rfloor}\}^{k'} \times \{1,2,\dots 2^{\lfloor nR' \rfloor}\}^k$, define

$$\begin{aligned} h^{2n''}(a^{k'+k}) \triangleq (&f^{kn}(a^{k+k'}(1)), f^{kn}(a^{k+k'}(2)), \dots, f^{kn}(a^{k+k'}(k')),\\ &f'^{k'n}(a^{k+k'}(k'+1)), f'^{k'n}(a^{k+k'}(k'+2)), \dots, f'^{k'n}(a^{k+k'}(k'+k))) \end{aligned} \tag{5.16}$$

Lemma 5.3. *The jump source-code t has rate* $\dfrac{R+R'}{2}$.

Proof. The range of $g^{2n''}$ has cardinality $2^{k'\lfloor knR\rfloor + k'\lfloor k'nR'\rfloor}$. $k'\lfloor knR\rfloor + k'\lfloor k'nR'\rfloor \leq \lfloor k'knR + kk'nR'\rfloor = n''(R+R')$. It follows that the cardinality of the image of $g^{2n''} \leq 2^{\lfloor n''(R+R')\rfloor}$ $= 2^{\lfloor 2n''\frac{R+R'}{2}\rfloor}$. Thus, the jump source-code t has rate $\frac{R+R'}{2}$. □

Note 5.39. By an extension of the above definition, equal time sharing can be defined for random jump source codes. We would have no need for it, and the notation becomes complicated; hence, we do not define it.

Lemma 5.4. *$\forall X$, $R^E_X(\cdot, j)$ is a convex function of D for $D \in (0,\infty)$. As a corollary, $R^E_X(D,j)$ is continuous function of D for $D \in (0,\infty)$.*

Proof. Let $s =< s^{kn} >_1^\infty$ be a rate R jump source-code such that expected distortion D is achievable when encoding i.i.d. X source. By Lemma 5.1, there exists a rate R deterministic jump source-code with expected distortion D when encoding i.i.d. X source. Thus, without loss of generality, assume that s is deterministic. Let $s' =< s'^{k'n} >_1^\infty$ be a rate R' jump source-code such that expected distortion D' is achievable when encoding the i.i.d. X source. By Lemma 5.1, there exists a rate R deterministic jump source-code with expected distortion D' when encoding i.i.d. X' source. Thus, without loss of generality, assume that s' is deterministic. Denote $n'' = kk'n$. Define the random variables

$$D_{1,kn} \triangleq \frac{1}{kn} d^{kn}\left((X^{2n''}(1..kn)), Y^{2n''}(1..kn)\right) \tag{5.17}$$

$$D_{2,kn} \triangleq \frac{1}{kn} d^{kn}\left((X^{2n''}(kn+1..2kn)), Y^{2n''}(kn+1..2kn)\right) \tag{5.18}$$

$$\vdots$$

$$D_{k',kn} \triangleq \frac{1}{kn} d^{kn}\left((X^{2n''}(k(k'-1)n+1..n'')), Y^{2n''}(k(k'-1)n+1..n'')\right) \tag{5.19}$$

and

$$D'_{1,k'n} \triangleq \frac{1}{k'n} d^{k'n}\left(X^{2n''}(n''+1..n''+k'n), Y^{2kk'n}(n''+1..n''+k'n)\right) \tag{5.20}$$

$$D'_{2,k'n} \triangleq \frac{1}{k'n} d^{k'n}\left(X^{2n''}(n''+k'n+1..n''+2k'n), Y^{2kk'n}(n''+k'n+1..n''+2k'n)\right) \tag{5.21}$$

$$\vdots$$

$$D'_{k,k'n} \triangleq \frac{1}{k'n} d^{k'n}\left(X^{2n''}(n''+k'(k-1)n..2n''), Y^{2kk'n}(n''+k'(k-1)n..2n'')\right) \tag{5.22}$$

The expected distortion produced by the jump source-code t which equally time shares between s and s' satisfies, by construction,

$$\limsup_{n\to\infty} E[D_{i,kn}] \leq D, 1 \leq i \leq k' \text{ and } \limsup_{n\to\infty} E[D'_{j,k'n}] \leq D', 1 \leq j \leq k \tag{5.23}$$

The expected distortion produced by the jump source-code t is

$$\begin{aligned}
&\limsup_{n\to\infty} E\left[\frac{1}{2n''} d^{2n''}\left(X^{2n''}, Y^{2n''}\right)\right] \\
=&\limsup_{n\to\infty} E\left[\sum_{i=1}^{k'} \frac{D_{i,kn}}{2k'} + \sum_{j=1}^{k} \frac{D_{j,k'n}}{2k}\right] \\
=&\limsup_{n\to\infty} \left(\sum_{i=1}^{k'} \frac{1}{2k'} E[D_{i,kn}] + \sum_{j=1}^{k} \frac{1}{2k} E[D_{j,k'n}]\right) \\
\leq&\sum_{i=1}^{k'} \frac{1}{2k'} \limsup_{n\to\infty} E[D_{i,kn}] + \sum_{j=1}^{k} \frac{1}{2k} \limsup_{n\to\infty} E[D_{i,k'n}] \\
\leq&\sum_{i=1}^{k'} \frac{1}{2k'} D + \sum_{j=1}^{k} \frac{1}{2k} D' \\
=&\frac{D+D'}{2}
\end{aligned} \tag{5.24}$$

Thus, expected distortion $\frac{D+D'}{2}$ is achievable with the rate $\frac{R+R'}{2}$ jump source-code t when encoding the i.i.d. X source. It follows that $R_X^E(D,j)$ is mid-point convex for $D \in [0,\infty)$. $R_X^E(D,j)$ is a decreasing function of D, and hence, the inverse image of an interval is an interval. Thus, $R_X^E(D,j)$ is a measurable function of D. Thus, $R_X^E(D)$ is mid-point convex for $D \in [0,\infty)$ and Lebesgue measurable for $D \in [0,\infty)$. By Sierpinski's theorem, $R_X^E(D,j)$ is a convex function of D for $D \in (0,\infty)$. As a corollary, $R_X^E(D,j)$ is a continuous function of D for $D \in (0,\infty)$. □

Note 5.40. The above proof does not hold for the uniform X source. The special structure of the i.i.d. X source is used.

Note 5.41. Sierpenski's theorem says that a Lebesgue measurable, mid-point convex function on an open interval is convex. Sierpinski's theorem was invoked above to prove that $R_X^E(D,j)$ is convex. In fact, we do not need to use Sierpinski's theorem. There's another theorem which says that a mid-point convex function on an open interval which is not convex is everywhere discontinuous. Now, $R_X^E(D,j)$ is a decreasing function of D and thus, has at most countably many discontinuities. Also, as proved above, $R_X^E(D,j)$ is a mid-point convex function of D. Thus, $R_X^E(D,j)$ is convex for $D \in (0,\infty)$.

Lemma 5.5. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_r$ be measurable sets such that* $\Pr(\mathcal{A}_i) \geq 1 - \delta_i$. *Then* $\Pr(\cap_{i=1}^r \mathcal{A}_i) \geq 1 - \sum_{i=1}^r \delta_i$.

Proof.

$$\Pr(\mathcal{A}_1 \cap \mathcal{A}_2) = \Pr(\mathcal{A}_1) + \Pr(\mathcal{A}_2) - \Pr(\mathcal{A}_1 \cup \mathcal{A}_2) \geq (1-\delta_1) + (1-\delta_2) - 1 = 1 - (\delta_1 + \delta_2) \quad (5.25)$$

By induction,

$$\Pr(\cap_{i=1}^r \mathcal{A}_i) \geq 1 - \sum_{i=1}^r \delta_i \quad (5.26)$$

□

Lemma 5.6. $\forall X$, $R_X^P(D,j)$ *is convex on* $D \in (0,\infty)$. *As a corollary,* $R_X^P(D,j)$ *is continuous for* $D \in (0,\infty)$.

Proof. Let $s = < s^{kn} >_1^\infty$ be a rate R jump source-code such that probability of excess distortion D is achievable when encoding i.i.d. X source. By Lemma 5.1, there exists a rate R deterministic jump source-code with probability of excess distortion D when encoding i.i.d. X source. Thus, without loss of generality, assume that s is deterministic. Let $s' = < s'^{k'n} >_1^\infty$ be a rate R' jump source-code such that probability of excess distortion D' is achievable when encoding the i.i.d. X source. By Lemma 5.1, there exists a rate R deterministic jump source-code with probability of excess distortion D' when encoding the i.i.d. X' source. Thus, without loss of generality, assume that s' is deterministic. Denote $n'' = kk'n$. Define the random variables $D_{i,kn}, 1 \leq i \leq k'$, $D_{j,k'n}, 1 \leq j \leq k$ as in Lemma 5.4. Then probability of excess distortion produced by the jump source-code t which equally time shares between s and s' satisfies, by construction,

$$\Pr\left(D_{i,kn} > D\right) = \epsilon_{kn} \to 0 \text{ as } n \to \infty \;\forall 1 \leq i \leq k' \quad (5.27)$$

$$\Pr\left(D_{j,k'n} > D'\right) = \epsilon_{k'n} \to 0 \text{ as } n \to \infty \;\forall 1 \leq j \leq k \quad (5.28)$$

Thus,

$$\Pr\left(D_{i,kn} \leq D\right) = 1 - \epsilon_{kn} \quad (5.29)$$

$$\Pr\left(D_{j,k'n} \leq D'\right) = 1 - \epsilon_{k'n} \quad (5.30)$$

For the jump source-code t, when the block-length n,

$$\Pr\left(\frac{1}{n} d^{2n''}(X^{2n''}, Y^{2n''}) \leq \frac{D+D'}{2}\right)$$

$$
\begin{aligned}
&= \Pr\left(\sum_{i=1}^{k'} \frac{D_{i,kn}}{2k'} + \sum_{j=1}^{k} \frac{D_{j,k'n}}{2k} \le \frac{D+D'}{2}\right) \\
&\le \Pr\left(\frac{D_{i,kn}}{2k'} \le \frac{D}{2k'}, 1 \le i \le k' \text{ and } \frac{D_{j,kn}}{2k} \le \frac{D'}{2k}, 1 \le j \le k\right) \\
&= \Pr\left(D_{i,kn} \le D, 1 \le i \le k' \text{ and } D_{j,kn} \le D', 1 \le j \le k\right) \\
&= \Pr\left(\cap_{i=1}^{k'} \{D_{i,kn} \le D\} \cap_{j=1}^{k} \{D_{j,k'n} \le D'\}\right) \\
&\ge 1 - \left(\sum_{i=1}^{k'} \epsilon_{kn} + \sum_{j=1}^{k} \epsilon_{k'n}\right) \text{ by Lemma 5.5} \\
&= 1 - (k\epsilon_{k'n} + k'\epsilon_{kn}) \to 1 \text{ as } n \to \infty
\end{aligned}
\tag{5.31}
$$

Thus,

$$
\lim_{n\to\infty} \Pr\left(\frac{1}{n} d^{2n''}(X^{2n''}, Y^{2n''}) > \frac{D+D'}{2}\right) = 1 \tag{5.32}
$$

Thus, probability of excess distortion $\frac{D+D'}{2}$ is achievable with the rate $\frac{R+R'}{2}$ jump source-code t when encoding the i.i.d. X source. It follows that $R_X^P(D,j)$ is mid-point convex for $D \in [0,\infty)$. $R_X^P(D,j)$ is a decreasing function of D. By an argument similar to Lemma 5.4, it follows that $R_X^P(D,j)$ is a convex function of D for $D \in (0,\infty)$ and hence, also, a continuous function of D for $D \in (0,\infty)$. □

Note 5.42. The above proof does not hold for the uniform X source. The special structure of the i.i.d. X source is used.

Note 5.43. *The proofs of Lemmas 5.4 and 5.6 provided above do not work for the functions $R_X^E(D)$ and $R_X^P(D)$. This is one of the reasons for defining the jump source-codes and the jump rate-distortion functions $R_X^E(D,j)$ and $R_X^P(D,j)$. There are other reasons which will be provided at the right place.*

Next we prove the equality of the rate-distortion function with jump source-codes and when jumps are not allowed.

Equality of rate-distortion function with jump and usual source-codes: $R_X^E(D,j) = R_X^E(D)$ and $R_X^P(D,j) = R_X^P(D)$

We prove results concerning relations between achievable distortion levels and rates for source-codes and jump source-codes for the i.i.d. X source, and as a consequence, prove that $R_X^E(D,j) = R_X^E(D)$ and $R_X^P(D,j) = R_X^P(D)$.

For this, we first need to define an interpolation code.

Definition 5.35 (Interpolation code). Let $s =< s^{kn} >_1^{\infty}$ be a jump source-code. We construct the interpolation code $t =< t^n >_1^{\infty}$ as follows:

Let $r \in \{0, 1, 2, \dots, k-1\}$.

Let $(e^{kn}, f^{kn}) \in \mathcal{E}_{\mathcal{X}}^{kn}(R) \times \mathcal{F}_{\mathcal{X}}^{kn}(R)$. Fix a $y \in \mathcal{Y}$ arbitrarily. Define $(e^{kn+r}, f^{kn+r}) \in \mathcal{E}_{\mathcal{X}}^{kn+r}(R) \times \mathcal{F}_{\mathcal{X}}^{kn+r}(R)$ as follows.

$$e^{kn+r}(x^{kn+r}) = e^{kn}(x^{kn}) \tag{5.33}$$

$$\tag{5.34}$$

Note that the range of e^{kn+r} is $\{1, 2, \dots, 2^{\lfloor knR \rfloor}\} \subset \{1, 2, \dots, 2^{\lfloor (kn+r)R \rfloor}\}$. If $a \in \{1, 2, \dots, 2^{\lfloor knR \rfloor}\}$,

$$f^{kn+r}(a) = (f^{kn}(a), y, y, \dots, y) \tag{5.35}$$

where, in the above expression, the number of y is r. If $a \in \{2^{\lfloor knR \rfloor}+1, \dots, 2^{\lfloor (kn+r)R \rfloor}\}$, define

$$f^{kn+r}(a) = (y, y, \dots, y) \tag{5.36}$$

where the number of y in the above expression is $kn + r$.

Define $p_{t^{kn+r}}(e^{kn+r}, f^{kn+r}) = p_{s^{kn}}(e^{kn}, f^{kn})$ where $(e^{kn}, f^{kn}) \in \mathcal{E}_{\mathcal{X}}^{kn}(R) \times \mathcal{F}_{\mathcal{X}}^{kn}(R)$ and (e^{kn+r}, f^{kn+r}) is as above.

This defines the source-code $t =< t^n >_1^{\infty}$.

Note that $t^{kn} = s^{kn}$.

Note 5.44. In the interpolation code, one can think of interpolation as being done in the trivial way. Also, for $r \in \{0, 2, \dots, k-1\}$, t^{kn+r} acts as s^{kn} on the initial block of length kn.

Lemma 5.7. *If $s =< s^{kn} >_1^{\infty}$ has rate R, the interpolation code $t =< t^n >_1^{\infty}$ also has rate R.*

Proof. This is clear from the definition of the interpolation code. □

Lemma 5.8. *If expected distortion D is achievable with a jump source-code $s =< s^{kn} >_1^{\infty}$ for the i.i.d. X source, expected distortion D is also achievable with the interpolation source-code $t =< t^n >_1^{\infty}$ for the i.i.d. X source.*

Proof. Expected distortion D is achievable with jump source-code $s =< s^{kn} >_1^{\infty}$ for the i.i.d. X source. That is, for the source-code s,

$$\mathrm{E}_{X^{kn}Y^{kn}}\left[\frac{1}{kn} d^{kn}(X^{kn}, Y^{kn})\right] = D_{kn} \text{ where } \limsup_{n\to\infty} D_{kn} \leq D \tag{5.37}$$

For the interpolation code t, for $r \in \{0,1,2,\ldots,k-1\}$, for block length $kn+r$,

$$
\begin{aligned}
&\mathrm{E}_{X^{kn+r}Y^{kn+r}}\left[\frac{1}{kn+r}d^{kn+r}(X^{kn+r},Y^{kn+r})\right] \\
=&\mathrm{E}_{X^{kn+r}Y^{kn+r}}\left[\frac{1}{kn+r}\left(d^{kn}(X^{kn+r}(1..kn),Y^{kn+r}(1..kn))+\right.\right. \\
&\qquad\left.\left. d^{r}(X^{kn+r}(kn+1..kn+r),Y^{r}(kn+1..kn+r))\right)\right] \\
=&\frac{kn}{kn+r}\mathrm{E}_{X^{kn+r}Y^{kn+r}}\left[\frac{1}{kn}d^{kn}(X^{kn+r}(1..kn),Y^{kn+r}(1..kn))\right] \\
&+\frac{1}{kn+r}\mathrm{E}_{X^{kn+r}Y^{kn+r}}\left[d^{r}(X^{kn+r}(kn+1..kn+r),Y^{r}(kn+1..kn+r))\right] \\
\leq&\mathrm{E}_{X^{kn+r}Y^{kn+r}}\left[\frac{1}{kn}d^{kn}(X^{kn+r}(1..kn),Y^{kn+r}(1..kn))\right]+\frac{rD_{\max}}{kn+r} \qquad (5.38)
\end{aligned}
$$

By construction of the interpolation-code, t^{kn+r} acts as s^{kn} on the initial block of length kn. It follows that

$$\mathrm{E}_{X^{kn+r}Y^{kn+r}}\left[\frac{1}{kn+r}d^{kn+r}(X^{kn+r},Y^{kn+r})\right]\leq D_{kn}+\frac{rD_{\max}}{kn+r} \qquad (5.39)$$

It follows that

$$\limsup_{n\to\infty}\mathrm{E}_{X^{kn+r}Y^{kn+r}}\left[\frac{1}{kn+r}d^{kn+r}(X^{kn+r},Y^{kn+r})\right]\leq D \qquad (5.40)$$

Thus, expected distortion D is achievable for the interpolation source-code t for the i.i.d. X source. ☐

Lemma 5.9. *If probability of excess distortion D is achievable with a jump source-code $s =< s^{kn} >_1^\infty$ for the i.i.d. X source, probability of excess distortion $D+\epsilon$ is achievable with the interpolation source-code $t =< t^n >_1^\infty$ for the i.i.d. X source, $\forall\epsilon > 0$.*

Proof. Probability of excess distortion D is achievable with jump source-code $s =< s^{kn} >_1^\infty$ for the i.i.d. X source. That is, for the source-code s,

$$p_{X^{kn}Y^{kn}}\left(\frac{1}{kn}d^{kn}(X^{kn},Y^{kn})>D\right)=\epsilon_{kn}\text{ where }\lim_{n\to\infty}\epsilon_{kn}=0 \qquad (5.41)$$

For the interpolation code t, for $r \in \{0,1,2,\ldots,k-1\}$, for block length $kn+r$,

$$
\begin{aligned}
&p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn+r}d^{kn+r}(X^{kn+r},Y^{kn+r})>D+\epsilon\right) \\
=&p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn+r}\left(d^{kn}(X^{kn+r}(1..kn),Y^{kn+r}(1..kn))\right.\right.
\end{aligned}
$$

$$+ \, d^r(X^{kn+r}(kn+1..kn+r), Y^r(kn+1..kn+r))\Big) > D+\epsilon\Big)$$

$$= p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn+r} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) \right.$$
$$\left. + \frac{1}{kn+r} d^r(X^{kn+r}(kn+1..kn+r), Y^r(kn+1..kn+r)) > D+\epsilon\right)$$
$$\le p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) \right.$$
$$\left. + \frac{1}{kn+r} d^r(X^{kn+r}(kn+1..kn+r), Y^r(kn+1..kn+r)) > D+\epsilon\right) \quad (5.42)$$

$\frac{1}{kn+r} d^r(x^r, y^r) \le \frac{rD_{\max}}{kn+r} \le \epsilon$ for n sufficiently large, $\forall x^r \in \mathcal{X}^r, y^r \in \mathcal{Y}^r$. It follows that for n sufficiently large,

$$p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) \right.$$
$$\left. + \frac{1}{kn+r} d^r(X^{kn+r}(kn+1..kn+r), Y^r(kn+1..kn+r)) > D+\epsilon\right)$$
$$\le p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) + \epsilon > D+\epsilon\right)$$
$$= p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) > D\right) \quad (5.43)$$

By construction of the interpolation code, t, for the first kn block, t^{kn+r} behaves as s^{kn}. It follows that for n sufficiently large,

$$p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) > D\right) = \epsilon_{kn} \to 0 \text{ as } n \to \infty \quad (5.44)$$

Thus, for n sufficiently large,

$$p_{X^{kn+r}Y^{kn+r}}\left(\frac{1}{kn+r} d^{kn+r}(X^{kn+r}, Y^{kn+r}) > D+\epsilon\right) \le \epsilon_{kn} \to 0 \text{ as } n \to \infty \quad (5.45)$$

$$(5.46)$$

Thus, $\forall \epsilon > 0$ probability of excess distortion D is achievable for the interpolation source-code t for the i.i.d. X source. □

Lemma 5.10. *For $D \in [0,\infty)$, $R_X^E(D) = R_X^E(D,j)$ and for $D \in (0,\infty)$, $R^P(D) = R_X^P(D,j)$. $R_X^P(D)$ and $R_X^E(D)$ are convex and continuous functions of D for $D \in (0,\infty)$.*

Proof. For every rate R, the set of rate R source codes which code the i.i.d. X source is a subset of the set of rate R jump source-codes which code the i.i.d. X source. It follows that

$R_X^E(D) \geq R_X^E(D,j)$. From Lemmas 5.7 and 5.8, it follows that $R_X^E(D) \leq R_X^E(D,j)$. Thus, $R_X^E(D) = R_X^E(D,j)$. From Lemma 5.4, it follows that $R_X^E(D)$ is a convex and continuous function of D for $D \in (0,\infty)$.

For every rate R, the set of rate R source codes which code the i.i.d. X source is a subset of the set of rate R jump source-codes which code the i.i.d. X source. It follows that $R_X^P(D) \geq R_X^P(D,j)$. From Lemmas 5.7 and 5.9, it follows that$\forall \epsilon > 0$, $R_X^P(D+\epsilon) \leq R_X^P(D,j)$. In other words, $R_X^P(D) \leq R_X^P(D-\epsilon,j)$. Taking limit as $\epsilon \to 0$, and by the continuity of $R_X^P(D)$ for $D \in (0,\infty)$ (Lemmmma 5.6), it follows that for $D \in (0,\infty)$, $R_X^P(D) = R_X^P(D,j)$. Since $R_X^P(0) \geq R_X^P(D,j)$ and since $R_X^P(D)$ is convex for $D \in (0,\infty)$ (Lemma 5.6), it follows that $R_X^P(D)$ is convex for $D \in (0,\infty)$. By the continuity of $R_X^P(D,j)$ for $D \in (0,\infty)$ (Lemma 5.6), it follows that $R_X^P(D)$ is a continuous function of D for $D \in (0,\infty)$. □

Next we prove the equality of the rate-distortion function when jump source-codes are allowed and with the inf definition

Proofs that $R_X^E(D,\text{inf}) = R_X^E(D,j)$ and $R_X^P(D,\text{inf}) = R_X^P(D,j)$

Next we compare achievable distortions levels and rates for jump source-codes with inf-achievable distortions and inf-achievable rates for source-codes, and as a consequence, will follow that $R_X^E(D,\text{inf}) = R_X^E(D,j)$ and $R_X^P(D,\text{inf}) = R_X^P(D,j)$. For this, we need to define jump repetition code.

Definition 5.36 (Jump repetition code). Let $s^k = (e^k, f^k) \in \mathscr{E}_{\mathscr{X}}^k(R) \times \mathscr{F}_{\mathscr{X}}^k(R)$. We construct a jump repetition source-code $s = < s^{kn} >_1^\infty = < e^{kn}, d^{kn} >_1^\infty$ as follows:

Let $x^{kn} \in \mathscr{X}^{kn}$. Define

$$e^{kn}(x^{kn}) = (e^k(x^{kn}(1..k)), e^k(x^{kn}(k+1..2k)), \ldots, e^k(x^{kn}((n-1)k+1..nk))) \quad (5.47)$$

For $a^n \in \{1,2,\ldots,2^{\lfloor k \rfloor}\}^n$, define

$$f^{kn}(a^n) = (f^k(a^n(1)), f^k(a^n(2)), \ldots, f^k(a^n(n))) \quad (5.48)$$

This defines the jump repetition code s.

Note 5.45. The jump repetition s code can be defined, analogously if $s^k \in \mathscr{P}(\mathscr{E}_{\mathscr{X}}^k(R) \times \mathscr{D}_{\mathscr{X}}^k(R)$. In this case, the jump repetition code will be random. We will not have need for this, and hence, we omit a precise definition.

Lemma 5.11. *If $s^k = (e^k, f^k) \in \mathscr{E}_{\mathscr{X}}^k(R) \times \mathscr{F}_{\mathscr{X}}^k(R)$, the jump repetition code $s = < s^{kn} >_1^\infty = < e^{kn}, f^{kn} >_1^\infty$ corresponding to s^k has rate R*

Proof. The image space of e^{kn} is the set $\{1,2,\ldots,2^{\lfloor k \rfloor}\}^n$. The cardinality of this set is $2^{n\lfloor k \rfloor} \leq 2^{\lfloor kn \rfloor}$. Thus, the jump repetition code s has rate R. □

Lemma 5.12. *Let s be a rate R source-code for which expected distortion D is inf-achievable when encoding the i.i.d. X source. Then, $\forall \epsilon > 0$, there exists a rate R jump source-code for which expected distortion $D+\epsilon$ is achievable when encoding the i.i.d. X source.*

Proof. Let $s =< s^n >_1^\infty$ be a rate R source-code such that expected distortion D is inf-achievable when coding the i.i.d. X source. By Lemma 5.1, there exists rate R deterministic source-code such that expected distortion D is inf-achievable when coding the i.i.d. X source. Without loss of generality, assume that s is deterministic. There exists a sequence $n_i \nearrow \infty$ such that for the source-code s,

$$\mathrm{E}_{X^{n_i}Y^{n_i}}\left[\frac{1}{n_i}d^{n_i}(X^{n_i},Y^{n_i})\right] \leq D+\epsilon_i, \text{ where } \epsilon_i \to 0 \text{ as } n_i \nearrow \infty \tag{5.49}$$

Denote

$$D_{n_i} \triangleq \frac{1}{n_i}d^{n_i}(X^{n_i},Y^{n_i}) \tag{5.50}$$

Let $t_{n_i} =< t_{n_i}^{n_i n} >_1^\infty$ denote the jump repetition code corresponding to s^{n_i}. Fix n_i and denote $n_i n \triangleq n''$. For the source-code t_{n_i}, denote,

$$D_{n,1} \triangleq \frac{1}{n_i}d^{n_i}(X^{n_i n}(1..n_i), Y^{n_i n}(1..n_i)) \tag{5.51}$$

$$D_{n,2} \triangleq \frac{1}{n_i}d^{n_i}(X^{n_i n}(n_i+1..2n_i), Y^{n_i n}(n_i+1..2n_i)) \tag{5.52}$$

$$\vdots$$

$$D_{n,n_i} \triangleq \frac{1}{n_i}d^{n_i}(X^{n_i n}((n-1)n_i+1..nn_i), Y^{n_i n}((n-1)n_i+1..nn_i)) \tag{5.53}$$

By the definition of the jump repetition code, $D_{n,k}, 1 \leq k \leq n_i$ are independent and identically distributed random variables, each having distribution D_{n_i}. For the source-code t_{n_i},

$$\begin{aligned}
&\mathrm{E}_{X^{n_i n}Y^{n_i n}}\left[\frac{1}{n_i n}d^{n_i n}(X^{n_i n},Y^{n_i n})\right] \\
=&\mathrm{E}\left[\sum_{j=1}^{n} D_{j,n_i}\right] \\
=&\frac{1}{n}\sum_{j=1}^{n} ED_{j,n_i}
\end{aligned}$$

$$=\frac{1}{n}\sum_{j=1}^{n} ED_{n_i}$$
$$=ED_{n_i}$$
$$\leq D+\epsilon_i \tag{5.54}$$

Thus, expected distortion $D+\epsilon_i$ is achievable by use of the rate R jump source-code t_{n_i} when source-coding the i.i.d. X source. Note that $\epsilon_i \to 0$ as $n_i \nearrow \infty$. It follows that $\forall \epsilon > 0$, there exists a rate R jump source-code such that expected distortion $D+\epsilon$ is achievable when source-coding the i.i.d. X source. □

Lemma 5.13. *Let s be a rate R source-code for which probability of excess distortion D is inf-achievable when encoding the i.i.d. X source. Then, $\forall \epsilon > 0$, there exists a rate R jump source-code for which probability of excess distortion $D+\epsilon$ is achievable when encoding the i.i.d. X source.*

Proof. Let $s =< s^n >_1^\infty$ be a rate R source-code such that probability of excess distortion distortion D is inf-achievable when coding the i.i.d. X source. By Lemma 5.1, there exists rate R deterministic source-code such that probability of excess distortion distortion D is inf-achievable when coding the i.i.d. X source. Without loss of generality, assume that s is deterministic. There exists a sequence $n_i \nearrow \infty$ such that for the source-code s,

$$p_{X^{n_i}Y^{n_i}}\left(\frac{1}{n_i}d^{n_i}(X^{n_i},Y^{n_i}) > D\right) = \epsilon_i, \text{ where } \epsilon_i \to 0 \text{ as } n_i \nearrow \infty \tag{5.55}$$

Denote D_{n_i} as in the previous lemma.

$ED_{n_i} \leq (1-\epsilon_i)D + \epsilon_i D_{\max}$ and D_{n_i} has finite variance.

Let t_{n_i} be as in the previous lemma. Denote $D_{n,n_1}, D_{n,n_2}, \dots, D_{n,n_i}$, as in the previous lemma.

By the definition of the jump repetition code, $D_{n,k}, 1 \leq k \leq n_i$ are independent and identically distributed random variables, each having distribution D_{n_i}.

Let $\delta > 0$.

For the source-code t_{n_i},

$$\Pr\left(\frac{1}{n_i n}d^{n_i n}(X^{n_i n},Y^{n_i n}) > (1-\epsilon_i)D + \epsilon_i D_{\max} + \delta\right)$$
$$=\Pr\left(\sum_{j=1}^{n} D_{j,n_i} > (1-\epsilon_i)D + \epsilon_i D_{\max} + \delta\right)$$
$$\to 0 \text{ as } n \to \infty \text{ by the weak law of large numbers} \tag{5.56}$$

Thus, probability of excess distortion $(1-\epsilon_i)D + \epsilon_i D_{\max} + \delta$ is achievable by use of the rate R jump source-code t_{n_i} when source-coding the i.i.d. X source. Note that $\epsilon_i \to 0$ as

$n_i \nearrow \infty$ and $\delta > 0$ is arbitrary. It follows that $\forall \epsilon > 0$, there exists a rate R jump source-code such that probability of excess distortion $D + \epsilon$ is achievable when source-coding the i.i.d. X source. □

Lemma 5.14. *For $D \in (0,\infty)$, $R_X^E(D,\inf) = R_X^E(D,j)$ and $R_X^P(D,\inf) = R_X^P(D,j)$. $R_X^E(D,\inf)$ and $R_X^P(D,\inf)$ are convex and continuous functions of D for $D \in (0,\infty)$.*

Proof. First, we prove that $R_X^E(D,\inf) = R_X^E(D,j)$. By definition of $R_X^E(D,\inf)$ and $R_X^E(D,j)$, it follows that $\forall D \in [0,\infty), R_X^E(D,\inf) \leq R_X^E(D,j)$. By Lemma 5.12, it follows that $\forall \epsilon > 0$, $R_X^E(D+\epsilon,j) \leq R_X^E(D,\inf)$. By Lemma 5.4, $R_X^E(D,j)$ is a continuous function of D for $D \in (0,\infty)$. It follows that by taking $\epsilon \to 0$, that for $D \in (0,\infty)$, $R_X^E(D,\inf) = R_X^E(D,j)$. By Lemma 5.4, $R_X^E(D,\inf)$ is a convex and continuous function of D for $D \in (0,\infty)$.

Next, we prove that $R_X^P(D,\inf) = R_X^P(D,j)$. By definition of $R_X^P(D,\inf)$ and $R_X^P(D,j)$, it follows that $\forall D \in [0,\infty), R_X^P(D,\inf) \leq R_X^P(D,j)$. By Lemma 5.13, it follows that $\forall \epsilon > 0$, $R_X^P(D+\epsilon,j) \leq R_X^P(D,\inf)$. By Lemma 5.6, $R_X^P(D,j)$ is a continuous function of D for $D \in (0,\infty)$. It follows that by taking $\epsilon \to 0$, that for $D \in (0,\infty)$, $R_X^P(D,\inf) = R_X^P(D,j)$. By Lemma 5.6, $R_X^P(D,\inf)$ is a convex and continuous function of D for $D \in (0,\infty)$. □

Note 5.46. *Another reason for the introduction of jump source-codes is that in order to prove that $R_X^E(D) = R_X^E(D,\inf)$ and $R_X^P(D) = R_X^P(D,\inf)$, in our proof technique, we need to pass through jump-source-codes: we prove that $R_X^E(D) = R_X^E(D,j)$ and $R_X^E(D,\inf) = R_X^E(D,j)$ from which it follows that $R_X^E(D) = R_X^E(D,\inf)$, and similarly, we prove that $R_X^P(D) = R_X^P(D,j)$ and $R_X^P(D,\inf) = R_X^P(D,j)$ from which it follows that $R_X^P(D) = R_X^P(D,\inf)$. This is because, in order to prove that $R_X^E(D) = R_X^E(D,j)$, we convert a jump source-code into an interpolation-code, and in order to prove that $R_X^E(D,\inf) = R_X^E(D,j)$, we convert a code for the* inf *problem into a repetition code, which is a sequence of jump source codes. Exactly the same happens with the probability of excess distortion criterion. Passing through jump source-codes cannot be avoided in our proof technique. There is one more reason for the introduction of jump source-codes which will be provided at the right time.*

Next, we prove relations between the rate-distortion functions for i.i.d. X and uniform X sources.

Relation between the rate-distortion functions of the i.i.d. X and uniform X sources, and proofs of $R_X^E(D) = R_U^E(D)$ and $R_X^P(D) = R_U^P(D)$, $R_X^E(D,\inf) = R_U^E(D,\inf)$ and $R_X^P(D,\inf) = R_U^P(D,\inf)$

Let X be such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. We prove relations concerning rates and achievable distortions for i.i.d. X and uniform X sources. As a consequence, we prove that $R_X^E(D) = R_U^E(D)$ and $R_X^P(D) = R_U^P(D)$. Similarly, we prove relations concerning rates and inf-achievable distortions for i.i.d. X and uniform X sources, and as a consequence, prove that $R_X^E(D,\inf) = R_U^E(D,\inf)$ and $R_X^P(D,\inf) = R_U^P(D,\inf)$.

Definition 5.37 (Hamming distance between codewords). For $x, x' \in \mathscr{X}$, $d_H(x,x') = 0$ if $x' = x$ and $d_H(x,x') = 1$ if $x' \neq x$. For $x^n, x'^n \in \mathscr{X}^n$, $d_H^n(x^n, x'^n) \triangleq \sum_{i=1}^{n} d_H(x^n(i), y^n(i))$.

Lemma 5.15. *For* $x^n, x'^n \in \mathscr{X}^n$, $y^n \in \mathscr{Y}^n$, $d^n(x^n,y^n) \leq D_{\max} d_H^n(x^n,x'^n) + d^n(x'^n,y^n)$

Proof.

$$\begin{aligned}
d^n(x^n,y^n) = & \\
= & \sum_{\{i|x'^n(i) \neq x^n(i)\}} d(x^n(i),y^n(i)) + \sum_{\{i|x'^n(i)=x^n(i)\}} d(x^n(i),y^n(i)) \\
= & \sum_{\{i|x'^n(i) \neq x^n(i)\}} d(x^n(i),y^n(i)) + \sum_{\{i|x'^n(i)=x^n(i)\}} d(x'^n(i),y^n(i)) \\
\leq & \; d_H^n(x^n,x'^n) D_{\max} + d^n(x'^n,y^n)
\end{aligned} \tag{5.57}$$

□

Definition 5.38 (l^1 distance between probability distributions). Let p and q be probability distributions on $\mathscr{X}$. The l^1 distance between p and q is $d_1(p,q) \triangleq \sum_{x \in \mathscr{X}} |p(x) - q(x)|$

Let X' be an arbitrary random variable on $\mathscr{X}$. $X' =< X'^n >_1^\infty$ is the i.i.d. X' source.

The following two constructions will be needed. Let X be such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$.

1. Given a source-code s for uniform X source, construct a jump source-code s' for i.i.d. X' source. This construction is such that if X' has the same distribution as X, the source-code s' is "close to" s
2. Given a source-code t for i.i.d. X' source, construct a source-code t' for uniform X source. This construction is such that if X' has the same distribution as X, the source-code t' is "close to" t

The two constructions are described in detail below.

Construction 5.1. Let X' is arbitrary. Let X satisfy that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Recall that n_0 is the least positive integer for which $n_0 p_X(x)$ is rational $\forall x \in \mathscr{X}$.Recall that $n' = n_0 n$.

Let $s =< s^{n'} >_1^\infty$ be a (possibly random) rate R source-code used for source-coding uniform X.

Definition 5.39. Let $\epsilon > 0$. For $x^{n'} \in \mathcal{X}^{n'}$, define

$$\xi_\epsilon^{n'}(x^{n'}) = \left\{ u^{n'} \in \mathcal{U}^{n'} : \frac{1}{n'} d_H^{n'}(x^{n'}, u^{n'}) \leq d_1(p_X, p_{X'}) + \epsilon \right\} \tag{5.58}$$

Definition 5.40. The transition probability $k'^{n'} : \mathcal{X}^{n'} \to \mathcal{P}(\mathcal{U}^{n'})$ defined as

$$k'^{n'}(u^{n'}|x^{n'}) = \begin{cases} \frac{1}{|\xi_\epsilon^{n'}(x^{n'})|} & \text{if } u^{n'} \in \xi_\epsilon^{n'}(x^{n'}) \text{ and } \xi_\epsilon^{n'}(x^{n'}) \neq \phi \\ 0 & \text{if } u^{n'} \notin \xi_\epsilon^{n'}(x^{n'}) \text{ and } \xi_\epsilon^{n'}(x^{n'}) \neq \phi \\ \text{uniform on } \mathcal{U}^{n'} & \text{if } \xi_\epsilon^{n'}(x^{n'}) = \phi \end{cases} \tag{5.59}$$

Definition 5.41 (Construction of jump source-code s' from source-code s). Consider the transition probability $q'^{n'} : \mathcal{X}^{n'} \to \mathcal{P}(\mathcal{Y}^{n'})$ defined as $q'^{n'} = k'^{n'} \circ q_s^{n'}$. Recall Definition 5.7 for the definition of transition probability q_s corresponding to a source-code s.

There exist rate R jump source-codes $s' =< s'^{n'} >_1^\infty$ with input space $< \mathcal{X}^n >_1^\infty$ and output space $< \mathcal{Y}^n >_1^\infty$, such that $q_{s'}^{n'} = q'^{n'}$. One such jump source-code is the following. Let $< e^{n'}, f^{n'} >_1^\infty$ be a deterministic jump source-code with domain $< \mathcal{U}^{n'} >_1^\infty$ and range $< \mathcal{Y}^n >_1^\infty$. The source-code s gives a probability to this deterministic source-code given by $p_{s^{n'}}(e^{n'}, f^{n'})$. Let $< g^{n'} >_1^\infty$ be a sequence of deterministic functions where $g^{n'}$ has domain $\mathcal{X}^{n'}$ and range $\mathcal{U}^{n'}$. Then, $< g^{n'} \circ e^{n'}, f^{n'} >_1^\infty$ is a deterministic jump source-code with input space $< \mathcal{X}^n >_1^\infty$ and output space $< \mathcal{Y}^n >_1^\infty$. Define the jump source-code s' as

$$p_{s'^{n'}}(g^{n'} \circ e^{n'}, f^{n'}) \triangleq \left[\prod_{x^{n'} \in \mathcal{X}^{n'}} k'^{n'}(g^{n'}(x^{n'})|x^{n'}) \right] p_{s^{n'}}(e^{n'}, f^{n'}) \tag{5.60}$$

Note that

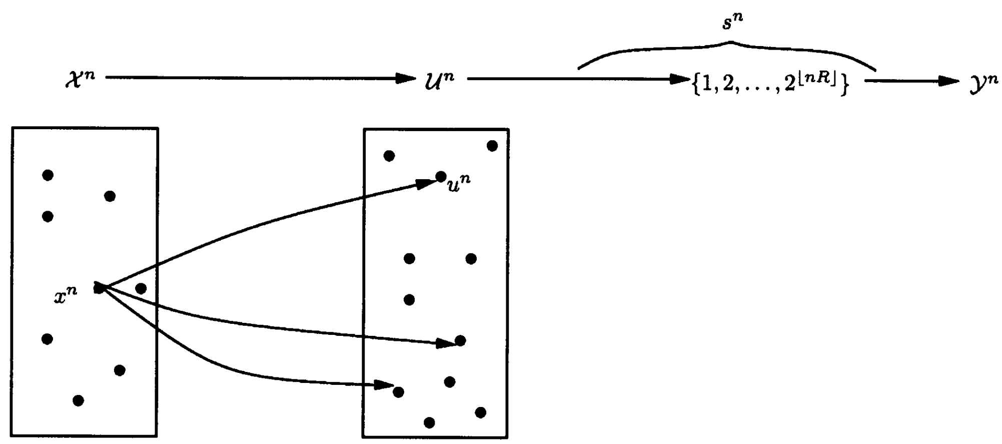
$$\sum_{x^{n'}, e^{n'}, f^{n'}, g^{n'}} p_{s'^{n'}}(g^{n'} \circ e^{n'}, f^{n'}) \triangleq \left[\prod_{x^{n'} \in \mathcal{X}^{n'}} k'^{n'}(g^{n'}(x^{n'})|x^{n'}) \right] p_{s^{n'}}(e^{n'}, f^{n'}) = 1 \tag{5.61}$$

(5.60), thus, defines a jump source-code with input space $< \mathcal{X}^{n'} >_1^\infty$ and output space $< \mathcal{Y}^{n'} >_1^\infty$. It can be checked that $q_{s'^{n'}} = q'^{n'}$.

See Figure 5.1.

Lemma 5.16. *Jump source-code s' has rate R*

Proof. The only deterministic jump source-codes on which $s'^{n'}$ possibly puts positive mass are those of the form $(g^{n'} \circ e^{n'}, f^{n'})$. $g^{n'} \circ e^{n'}$ has domain $\mathcal{X}^{n'}$ and range $\{1, 2, \ldots, 2^{\lfloor n'R \rfloor}\}$ and $f^{n'}$ has domain $\{1, 2, \ldots, 2^{\lfloor n'R \rfloor}\}$ and range $\mathcal{Y}^{n'}$. Thus, the only deterministic functions that $s'^{n'}$ possibly puts positive mass on $\in \mathcal{E}^{n'}(R) \times \mathcal{F}^{n'}(R)$, where $\mathcal{E}^{n'}(R)$ and $\mathcal{F}^{n'}(R)$ are defined in Subsection 5.4.1 It follows that the jump source-code s' has rate R. □



x^n is mapped to u^n with probability $\frac{1}{\xi_\epsilon^n(x^n)}$, if $u^n \in \xi_\epsilon^n(x^n)$

Figure 5.1. Construction of jump source-code s' from source-code s

Lemma 5.17. $\Pr(\xi_\epsilon^{n'}(X'^{n'}) = \phi) \to 0$ *as* $n \to \infty$.

Proof. The idea of the proof is the following. If $x^{n'}$ is generated i.i.d. X', with high probability, we will show the existence of a $u^{n'}$ such that

$$\frac{1}{n'} d_H^{n'}(x^{n'}, u^{n'}) \leq d_1(p_X, p_{X'}) + \epsilon \tag{5.62}$$

Let $x^{n'} \in \mathscr{X}^{n'}$. Define:

$$\mathscr{X}_1 \triangleq \{x | p_{x^{n'}}(x) \leq p_X\}$$
$$\mathscr{X}_2 \triangleq \{x | p_{x^{n'}}(x) > p_X\}$$
$$\mathscr{P}_x \triangleq \{i | x^{n'}(i) = x\}$$

If $x \in \mathscr{X}_1, \mathscr{Q}_x \triangleq \mathscr{P}_x$

If $x \in \mathscr{X}_2$, define $\mathscr{Q}_x \subset \mathscr{P}_x$ such that $|\mathscr{Q}_x| = n' p_X(x)$

Consider a $u^{n'}$ which satisfies the following:

1. If $i \in \mathscr{Q}_x$ for some x, $u^{n'}(i) \triangleq x$
2. If $i \notin \mathscr{Q}_x$ for any x, define $u^{n'}(i)$ in such a way that the empirical distribution of $u^{n'}$ is p_X. This is possible by definitions of $\mathscr{P}_x$ and $\mathscr{Q}_x$ and the relations between the cardinalities of these sets. $u^{n'}$ can be thought of as a function, $u^{n'}(x^{n'})$ of $x^{n'}$.

$$\begin{aligned}
&\frac{1}{n'}d_H^{n'}(x^{n'},u^{n'})\\
&=\frac{1}{n'}\sum_{x\in\mathcal{X}}|n'p_{x^{n'}}(x)-n'p_X(x)|\\
&=\sum_{x\in\mathcal{X}}|p_{x^{n'}}(x)-p_X(x)|\\
&\le\sum_{x\in\mathcal{X}}[|p_{x^{n'}}(x)-p_{X'}(x)|+|p_{X'}(x)-p_X(x)|]\\
&=\sum_{x\in\mathcal{X}}[|p_{x^{n'}}(x)-p_{X'}(x)|]+\sum_{x\in\mathcal{X}}[|p_{X'}(x)-p_X(x)|]\\
&=\sum_{x\in\mathcal{X}}[|p_{x^{n'}}(x)-p_{X'}(x)|]+d_1(p_X,p_{X'})
\end{aligned}\tag{5.63}$$

Thus,

$$\begin{aligned}
&\Pr\left(\frac{1}{n'}d_H^{n'}(X'^{n'},u^{n'}(X'^{n'}))>d_1(p_X,p_{X'})+\epsilon\right)\\
&\le\Pr\left(\sum_{x\in\mathcal{X}}[|p_{X'^{n'}}(x)-p_{X'}(x)|]\le\epsilon\right)\\
&\to 1 \text{ as } n\to\infty \text{ by weak law of large numbers}
\end{aligned}\tag{5.64}$$

□

Definition 5.42. If the input to the jump source-code s' is i.i.d. X' sequence, for block-length n', we get a joint random-variable $X'^{n'}V^{n'}Y^{n'}$ on $\mathcal{X}^{n'}\times\mathcal{U}^{n'}\times\mathcal{Y}^{n'}$ with the corresponding joint probability distribution $p_J=p_{X'^{n'}V^{n'}Y^{n'}}$ given by

$$p_J^{n'}(x^{n'},u^{n'},y^{n'})=p_{X'^{n'}}(x^{n'})k'^{n'}(u^{n'}|x^{n'})q_{s^{n'}}(y^{n'}|u^{n'})\tag{5.65}$$

$V^{n'}$ is the marginal random variable on $\mathcal{U}^{n'}$.

Next, we want to prove that $V^{n'}$ has the same distribution as the uniform X source of block-length n', $U^{n'}$. For this, we need a few definitions and lemmas.

Definition 5.43 (Permutations). Let $\pi^{n'}$ be a permutation of $(1,2,\ldots,n')$. For $i\in\{1,2,\ldots,n'\}$, $\pi^{n'}(i)$ is the image of i under the permutation $\pi^{n'}$. The set of all permutations of $(1,2,\ldots,n')$ is denoted by $\Pi^{n'}$. For $x^{n'}\in\mathcal{X}^{n'},y^{n'}\in\mathcal{Y}^{n'}$, define,

$$\pi^{n'}(x^{n'})\triangleq(x^{n'}(\pi^{n'}(1)),x^{n'}(\pi^{n'}(2)),\ldots,x^{n'}(\pi^{n'}(n')))\tag{5.66}$$

$$\pi^{n'}(y^{n'})\triangleq(y^{n'}(\pi^{n'}(1)),y^{n'}(\pi^{n'}(2)),\ldots,y^{n'}(\pi^{n'}(n')))\tag{5.67}$$

For a set $\mathcal{A}$, $\pi^{n'}A\triangleq\{\pi^{n'}a|a\in\mathcal{A}\}$.

Lemma 5.18. *If $x^{n'}, x'^{n'} \in \mathcal{X}^{n'}$, $x^{n'} \neq x'^{n'}$, then $\pi^{n'}(x^{n'}) \neq \pi^{n'}(x'^{n'})$*

Proof. $x^{n'} \neq x'^{n'} \Rightarrow \exists i$ such that $x^{n'}(i) \neq x'^{n'}(i) \Rightarrow x^{n'}(\pi^{n'}(i)) \neq x'^{n'}(\pi^{n'}(i)) \Rightarrow \pi^{n'}(x^{n'}) \neq \pi^{n'}(x'^{n'})$. □

Lemma 5.19. $|\xi_\epsilon^{n'}(x^{n'})|$ *depends only on the type of* $x^{n'}$.

Proof. Let $x^{n'}$ and $x'^{n'}$ have the same type. Thus, $x'^{n'} = \pi^{n'} x^{n'}$ for some permutation $\pi^{n'}$. For $u^{n'} \in \mathcal{U}^{n'}$, $d_H^{n'}(u^{n'}, x^{n'}) = d_H^{n'}(\pi^{n'} u^{n'}, \pi^{n'} x^{n'})$. Thus, $\pi^{n'} \xi_\epsilon^{n'}(x^{n'}) \subset \xi_\epsilon^{n'}(x'^{n'})$. By Lemma 5.18, $|\pi^{n'} \xi_\epsilon^{n'}(x^{n'})| = |\xi_\epsilon^{n'}(x^{n'})|$. Thus, $|\xi_\epsilon^{n'}(x^{n'})| \leq |\xi_\epsilon^{n'}(x'^{n'})|$. The same argument with $x'^{n'}$ interchanged with $x^{n'}$ proves that $|\xi_\epsilon^{n'}(x'^{n'})| \leq |\xi_\epsilon^{n'}(x^{n'})|$. Thus, $|\xi_\epsilon^{n'}(x'^{n'})| = |\xi_\epsilon^{n'}(x^{n'})|$. □

Definition 5.44. $|\xi_\epsilon^{n'}(q)| \triangleq |\xi_\epsilon^{n'}(x^{n'})|$ if $x^{n'}$ has type q, which by the above lemma, depends only on the type of $x^{n'}$.

Definition 5.45. For $u^{n'} \in \mathcal{U}^{n'}$, $\eta_{\epsilon,q}^{n'}(u^{n'}) \triangleq \{x^{n'} | x^{n'} \in \mathcal{X}^{n'}, x^{n'} \text{ has type } q, u^{n'} \in \xi_\epsilon^{n'}(x^{n'})\}$.

Lemma 5.20. $|\eta_{\epsilon,q}^{n'}(u^{n'})|$ *is independent of the particular* $u^{n'} \in \mathcal{U}^{n'}$.

Proof. All sequences $u^{n'} \in \mathcal{U}^{n'}$ have the same type p_X. The proof of this lemma follows exactly as the proof of Lemma 5.20. □

Definition 5.46. $|\eta_\epsilon^{n'}(q)| \triangleq |\eta_{\epsilon,q}^{n'}(u^{n'})|$ for $u^{n'} \in \mathcal{U}^{n'}$, which, by the above lemma, is independent of $u^{n'}$.

Lemma 5.21. *The random variable $V^{n'}$, defined in Definition 5.42 has the same distribution as $U^{n'}$, that is, uniform on $\mathcal{U}^{n'}$.*

Proof. For $x^{n'} \in \mathcal{X}^{n'}$, $p_{X'^{n'}}(x^{n'})$ depends only on the type q of $x^{n'}$. $p_{X'^{n'}}(x^{n'})$ is denoted by $p_{X'^{n'}}(q)$.

$$\begin{aligned} p_{V^{n'}}(u^{n'}) &= \sum_{x^{n'} \in \mathcal{X}^{n'}} p_{X'^{n'}}(x^{n'}) k'^{n'}(u^{n'} | x^{n'}) \\ &= \sum_{x^{n'} \in \mathcal{X}^{n'}, \xi_\epsilon^{n'}(x^{n'}) \neq \phi} p_{X'^{n'}}(x^{n'}) k'^{n'}(u^{n'} | x^{n'}) + \\ &\qquad \sum_{x^{n'} \in \mathcal{X}^{n'}, \xi_\epsilon^{n'}(x^{n'}) = \phi} p_{X'^{n'}}(x^{n'}) k'^{n'}(u^{n'} | x^{n'}) \end{aligned}$$

$$= \sum_q \sum_{\substack{x^{n'} \in \mathscr{X}^{n'}, \xi_\epsilon^{n'}(x^{n'}) \neq \phi \\ x^{n'} \text{ has type } q}} p_{X'^{n'}}(x^{n'}) k'^{n'}(u^{n'}|x^{n'}) +$$

$$\sum_{x^{n'} \in \mathscr{X}^{n'}, \xi_\epsilon^{n'}(x^{n'}) = \phi} p_{X'^{n'}}(x^{n'}) k'^{n'}(u^{n'}|x^{n'})$$

$$= \sum_q p_{X'^{n'}}(q) \frac{|\eta_\epsilon^{n'}(q)|}{|\xi_\epsilon^{n'}(q)|} + \sum_{x^{n'} \in \mathscr{X}^{n'}, \xi_\epsilon^{n'}(x^{n'}) = \phi} p_{X'^{n'}}(x^{n'}) \frac{1}{|\mathscr{U}^{n'}|}$$

which is independent of $u^{n'}$ (5.68)

Thus, $p_{V^{n'}}$ is uniform on $\mathscr{U}^{n'}$. □

This construction ends here.

Construction 5.2. Let X' be arbitrary and X satisfies that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Recall that n_0 is the least positive integer for which $n_0 p_X(x)$ is integer $\forall x \in \mathscr{X}$. Recall that $n' = n_0 n$.

Consider the joint random variable $X'^{n'} V^{n'}$ on $\mathscr{X}^{n'} \times \mathscr{U}^{n'}$ with corresponding probability distribution $p_{X'^{n'} V^{n'}}$ given by

$$p_{X'^{n'} V^{n'}}(x^{n'}, u^{n'}) = p_{X'^{n'}}(x^{n'}) k'^{n'}(u^{n'}|x^{n'}) \tag{5.69}$$

Recall Lemma 5.21 that $p_{V^{n'}}$ is uniform on $\mathscr{U}^{n'}$.

Definition 5.47. $p_{X'^{n'} V^{n'}}$ can be factored the other way:

$$p_{X'^{n'} V^{n'}}(x^{n'}, u^{n'}) = p_{V^{n'}}(u^{n'}) l'^{n'}(u^{n'}|x^{n'}) \tag{5.70}$$

Let $t = < t^n >_1^\infty$ be a rate R source-code used for source-coding i.i.d. X' source.

Definition 5.48 (Construction of source-code t' from source-code t). Consider the transition probability $r'^{n'} : \mathscr{U}^{n'} \to \mathscr{P}(\mathscr{Y}^{n'})$ defined as $r'^{n'} = l'^{n'} \circ q_t^{n'}$. Recall Definition 5.7 for the definition of the transition probability q_t corresponding to a source-code t.

There exists a rate R source-code $t' = < t'^{n'} >_1^\infty$ with input space $< \mathscr{U}^{n'} >_1^\infty$ and output space $< \mathscr{Y}^{n'} >_1^\infty$ such that $q_{t'}^{n'} = r'^{n'}$. The construction of such a source-code t is analogous to the construction described in Definition 5.41 and hence, omitted.

Lemma 5.22. *Source-code t' has rate R*

Proof. The proof is analogous to the proof of Lemma 5.16, and hence, omitted. □

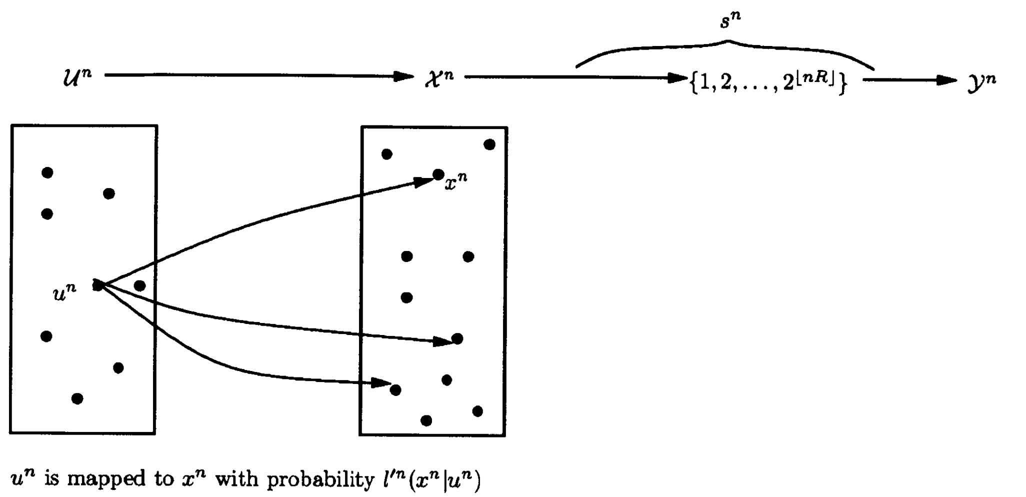


Figure 5.2. Construction of source-code t' from source-code t

Definition 5.49. If the input to the source-code t' is uniform X sequence, for block-length n', we get a joint random-variable $U^{n'}T^{n'}Y^{n'}$ on $\mathcal{U}^{n'} \times \mathcal{X}^{n'} \times \mathcal{Y}^{n'}$ with the corresponding joint probability distribution $p_K = p_{U^{n'}T^{n'}Y^{n'}}$ given by

$$p_K^{n'}(u^{n'}, x^{n'}, y^{n'}) = p_{U^{n'}}(u^{n'}) l'^{n'}(x^{n'}|u^{n'}) q_{t_{n'}}(y^{n'}|x^{n'}) \tag{5.71}$$

$T^{n'}$ is the marginal random variable on $\mathcal{X}^{n'}$.

Lemma 5.23. *$T^{n'}$ has the same distribution as $X'^{n'}$, that is, $p_{T^{n'}} = p_{X^{n'}}$*

Proof. This follows from Definition 5.47 □

See Figure 5.2.

This construction ends here.

Lemma 5.24. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x \in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R source-code for which expected distortion D is achievable when source-coding the uniform X source, then, there exists a rate R source-code for which expected distortion $D + D_{\max} d_1(p_X, p_{X'}) + \delta$ is achievable when source-coding the i.i.d. X' source.*

Proof. Let s be a rate R source code using which, expected distortion D is achievable when source-coding the uniform X source. Recall the construction of jump source code s' in Construction 5.1 which is used to source-code the i.i.d. X' source. First, we get a bound on achievable expected distortion for jump source code s' when coding the i.i.d. X' source.

Recall Definition 5.42 for the definition of p_J.

$$\frac{1}{n'}d^{n'}(x^{n'},y^{n'}) \leq \frac{1}{n'}d_H^{n'}(x^{n'},u^{n'})D_{\max} + \frac{1}{n'}d^{n}(u^{n'},y^{n'}) \text{ by Lemma 5.15}$$

$$\Rightarrow \sum_{x^{n'},u^{n'},y^{n'}} p_J(x^{n'},u^{n'},y^{n'})\left[\frac{1}{n'}d^{n'}(x^{n'},y^{n'})\right] \leq$$
$$\sum_{x^{n'},u^{n'},y^{n'}} p_J(x^{n'},u^{n'},y^{n'})\left[\frac{1}{n'}d_H^{n'}(x^{n'},u^{n'})D_{\max}\right]$$
$$+ \sum_{x^{n'},u^{n'},y^{n'}} p_J(x^{n'},u^{n'},y^{n'})\left[\frac{1}{n'}d^{n}(u^{n'},y^{n'})\right] \quad (5.72)$$

That is,

$$E\left[\frac{1}{n'}d^{n'}(X'^{n'},Y^{n'})\right] \leq E\left[\frac{1}{n'}d_H^{n'}(X'^{n'},V^{n'})D_{\max}\right] + E\left[\frac{1}{n'}d^{n'}(V^{n'},Y^{n'})\right]$$
$$= E\left[\frac{1}{n'}d_H^{n'}(X'^{n'},V^{n'})\right]D_{\max} + E\left[\frac{1}{n'}d^{n'}(V^{n'},Y^{n'})\right]$$
$$\leq \Pr(\xi_\epsilon^{n'}(X'^{n'}) \neq \phi)(d_1(p_X,p_{X'})+\epsilon)D_{\max} +$$
$$\Pr(\xi_\epsilon^{n'}(X'^{n'}) = \phi)D_{\max} +$$
$$E\left[\frac{1}{n'}d^{n'}(V^{n'},Y^{n'})\right]$$
$$\leq (d_1(p_X,p_{X'})+\epsilon)D_{\max} + E\left[\frac{1}{n'}d^{n'}(V^{n'},Y^{n'})\right] +$$
$$\Pr(\xi_\epsilon^{n'}(X'^{n'}) = \phi)D_{\max} \quad (5.73)$$

Thus,

$$\limsup_{n\to\infty} E\left[\frac{1}{n'}d^{n'}(X'^{n'},Y^{n'})\right] \leq$$
$$\limsup_{n\to\infty}(d_1(p_X,p_{X'})+\epsilon)D_{\max} +$$
$$\limsup_{n\to\infty} E\left[\frac{1}{n'}d^{n'}(V^{n'},Y^{n'})\right] +$$
$$\limsup_{n\to\infty} \Pr(\xi_\epsilon^{n'}(X'^{n'}) = \phi)D_{\max}$$
$$(5.74)$$

$\limsup_{n\to\infty} E\left[\frac{1}{n'}d^{n'}(X'^{n'},Y^{n'})\right]$ is the expected distortion produced by the jump source-code s' for the i.i.d. X' source. $p_{V^{n'}}$ is uniform and thus, $\limsup_{n\to\infty} E\left[\frac{1}{n'}d^{n'}(V^{n'},Y^{n'})\right]$ is the expected distortion produced by the source-code s for the uniform X source. $\Pr(\xi_\epsilon^{n'}(X'^{n'} = \phi) \to 0$ as $n \to \infty$ by Lemma 5.17. It follows that if s is a rate R source-code for which expected distortion D is achievable when source-coding the uniform X source, then, s' is a jump source-code for which expected distortion $D + D_{\max}(d_1(p_X, p_{X'}) + \epsilon)$ is achievable, when source-coding the i.i.d. X source.The source-code s' can be interpolated to get a rate R source-code s'' for which expected distortion $D + D_{max}(d_1(p_X, p_{X'}) + \epsilon)$ is achievable when source-coding the i.i.d. X' source. Such a source-code s'' exists for each $\epsilon > 0$. By choosing ϵ such that $D_{\max}\epsilon = \delta$, the lemma follows. □

Note 5.47. *This brings us to the final reason for the introduction of jump source-codes for source-coding i.i.d. X source. In the above lemma, we derive relations between $R_U^E(D)$ and $R_{X'}^E(D)$. This is done by the use of Construction 5.1 Construction 5.1 constructs a jump source-code for source-coding i.i.d. X' source given a source-code for source-coding the uniform X source. A direct construction of a source-code for the i.i.d. X' source from a source-code for the uniform X source is not possible because a source-code for the uniform X source is defined only for certain block-lengths. This also justifies, why we do not need to define jump source-codes for the uniform X source. Given a source-code for the i.i.d. X' source, we can directly construct a source-code for the uniform X source by use of Construction 5.2: this is a source-code for the uniform X source, not a jump source-code. Two other reasons were given previously for defining jump source codes for source-coding i.i.d. X source. Firstly, to prove convexity and continuity of $R_X^E(D)$ and $R_X^P(D)$, we have to first prove the same for the jump rate-distortion function. Secondly, to prove that rate-distortion functions for i.i.d. sources with* lim inf *and* lim sup *definitions are the same, we have to go through jump rate-distortion function. These reasons do not exist for the uniform X source because these results are not proved directly for the uniform X source: they are proved indirectly by proving equality with the corresponding rate-distortion functions for the i.i.d. X source and then invoking these results for the i.i.d. X source.*

Lemma 5.25. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x \in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R source-code for which expected distortion D is achievable when source-coding the i.i.d. X' source, then, there exists a rate R source-code for which expected distortion $D + D_{\max}d_1(p_X, p_{X'}) + \delta$ is achievable when source-coding the uniform X source.*

Proof. Let t be a rate R source-code using which, expected distortion D is achievable when source-coding the i.i.d. X' source. Recall the construction of the source-code t' in Construction 5.2 which is used to source-code the uniform X source. We get a bound on the achievable expected distortion for the source-code t' when source-coding the uniform X source.

Recall Definition 5.49 for the definition of p_K.

$$\frac{1}{n'}d^{n'}(u^{n'}, y^{n'}) \le \frac{1}{n'}d_H^{n'}(u^{n'}, x^{n'})D_{\max} + \frac{1}{n'}d^{n}(x^{n'}, y^{n'}) \text{ by Lemma 5.15}$$
$$\Rightarrow \sum_{u^{n'}, x^{n'}, y^{n'}} p_K(u^{n'}, x^{n'}, y^{n'}) \left[\frac{1}{n'}d^{n'}(u^{n'}, y^{n'})\right] \le$$

$$\sum_{u^{n'},x^{n'},y^{n'}} p_K(u^{n'},x^{n'},y^{n'}) \left[\frac{1}{n'} d_H^{n'}(u^{n'},x^{n'}) D_{\max}\right]$$
$$+ \sum_{u^{n'},x^{n'},y^{n'}} p_K(u^{n'},x^{n'},y^{n'}) \left[\frac{1}{n'} d^{n}(x^{n'},y^{n'})\right] \quad (5.75)$$

That is,

$$\begin{aligned}
E\left[\frac{1}{n'} d^{n'}(U^{n'},Y^{n'})\right] &\le E\left[\frac{1}{n'} d_H^{n'}(U^{n'},T^{n'}) D_{\max}\right] + E\left[\frac{1}{n'} d^{n'}(T^{n'},Y^{n'})\right] \\
&= E\left[\frac{1}{n'} d_H^{n'}(U^{n'},T^{n'})\right] D_{\max} + E\left[\frac{1}{n'} d^{n'}(T^{n'},Y^{n'})\right] \\
&= E\left[\frac{1}{n'} d_H^{n'}(T^{n'},U^{n'})\right] D_{\max} + E\left[\frac{1}{n'} d^{n'}(T^{n'},Y^{n'})\right] \\
&\le \Pr(\xi_\epsilon^{n'}(T^{n'}) \ne \phi)(d_1(p_X,p_{X'})+\epsilon) D_{\max} + \\
&\qquad \Pr(\xi_\epsilon^{n'}(T^{n'}) = \phi) D_{\max} + \\
&\qquad E\left[\frac{1}{n'} d^{n'}(T^{n'},Y^{n'})\right] \\
&\le (d_1(p_X,p_{X'})+\epsilon) D_{\max} + E\left[\frac{1}{n'} d^{n'}(T^{n'},Y^{n'})\right] + \\
&\qquad \Pr(\xi_\epsilon^{n'}(T^{n'}) = \phi) D_{\max}
\end{aligned} \quad (5.76)$$

Thus,

$$\begin{aligned}
\limsup_{n\to\infty} E\left[\frac{1}{n'} d^{n'}(U^{n'},Y^{n'})\right] &\le \\
&\limsup_{n\to\infty} (d_1(p_X,p_{X'})+\epsilon) D_{\max} + \\
&\qquad \limsup_{n\to\infty} E\left[\frac{1}{n'} d^{n'}(T^{n'},Y^{n'})\right] + \\
&\qquad \limsup_{n\to\infty} \Pr(\xi_\epsilon^{n'}(T^{n'}) = \phi) D_{\max}
\end{aligned} \quad (5.77)$$

$\limsup_{n\to\infty} E\left[\frac{1}{n'} d^{n'}(U^{n'},Y^{n'})\right]$ is the expected distortion produced by the source-code t' for the uniform X source. Recall that $p_{T^{n'}}$ is the same distribution as $p_{X'^{n'}}$ and thus,

$$\limsup_{n\to\infty} E\left[\frac{1}{n'} d^{n'}(T^{n'},Y^{n'})\right] \quad (5.78)$$

is the expected distortion produced by the jump source-code $< t^{n'} >_1^\infty$ for the i.i.d. X' source. The expected distortion produced by the jump source-code $< t^{n'} >_1^\infty$ is less than or equal to the expected distortion produced by the source-code $t =< t^n >_1^\infty$. $\Pr(\xi_\epsilon^{n'}(T^{n'} = \phi) \to 0$ as $n \to \infty$ since $T^{n'}$ has the same distribution as $X'^{n'}$ and by Lemma 5.17. It follows that if t is a rate R source-code for which expected distortion D is achievable when source-coding the i.i.d. X' source, then, t' is a rate R source-code for which expected distortion $D + D_{\max}(d_1(p_X, p_{X'} + \epsilon)$ is achievable, when source-coding the uniform X source. Such a source-code t' exists $\forall \epsilon > 0$. By choosing ϵ such that $D_{max}\epsilon = \delta$, the lemma follows. □

Lemma 5.26. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x \in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R source-code for which probability of excess distortion D is achievable when source-coding the uniform X source, then, there exists a rate R source-code for which probability of excess distortion $D + D_{\max}d_1(p_X, p_{X'}) + \delta$ is achievable when source-coding the i.i.d. X' source.*

Proof. Let s be a rate R source-code using which, probability of excess distortion D is achievable when source-coding the uniform X source. Recall the construction of jump source-code s' in Construction 5.1 which is used to source-code the i.i.d. X' source. First, we get a bound on the achievable probability of excess distortion for the jump source-code s' when source-coding the i.i.d. X' source.

Recall Definition 5.42 that $p_J = p_{X'^{n'}V^{n'}Y^{n'}}$ is the joint distribution on $\mathcal{X}^{n'} \times \mathcal{U}^{n'} \times \mathcal{Y}^{n'}$ induced by $k'^{n'}$ and the source-code s when the distribution on $\mathcal{X}^{n'}$ is i.i.d. X'. Recall also that $V^{n'}$ is a uniform random variable on $\mathcal{U}^{n'}$. Also, by Lemma 5.15,

$$\frac{1}{n'}d^{n'}(x^{n'},y^{n'}) \le \frac{1}{n'}d_H^{n'}(x^{n'},u^{n'})D_{\max} + \frac{1}{n'}d^{n}(u^{n'},y^{n'}) \tag{5.79}$$

Thus,

$$\frac{1}{n'}d^{n'}(X'^{n'},Y^{n'}) \le \frac{1}{n'}d_H^{n'}(X'^{n'},V^{n'})D_{\max} + \frac{1}{n'}d^{n'}(V^{n'},Y^{n'}) \tag{5.80}$$

Let probability of excess distortion D be achievable with source-code s for uniform X source. Then,

$$\begin{aligned}
&\Pr\left(\frac{1}{n'}d^{n'}(X'^{n'},Y^{n'}) > D + D_{\max}(d_1(p_X,p_{X'})+\epsilon)\right) \\
&\le \Pr\left(\frac{1}{n'}d_H^{n'}(X'^{n'},V^{n'})D_{\max} + \frac{1}{n'}d^{n'}(V^{n'},Y^{n'}) > D_{\max}(d_1(p_X,p_{X'})+\epsilon)+D\right)
\end{aligned} \tag{5.81}$$

$$\begin{aligned}
&\Pr\left(\frac{1}{n'}d_H^{n'}(X'^{n'},V^{n'}) > d_1(p_X,p_{X'})+\epsilon\right) = \epsilon_1^{n'} \to 0 \text{ as } n' \to \infty \\
&\qquad\qquad\qquad \text{since } \Pr(\xi_\epsilon^{n'}(X'^{n'}) = \phi) \to 0 \text{ as } n \to \infty \\
&\Rightarrow \Pr\left(\frac{1}{n'}d_H^{n'}(X'^{n'},V^{n'})D_{\max} > (d_1(p_X,p_{X'})+\epsilon)D_{\max}\right) = \epsilon_1^{n'} \to 0 \text{ as } n \to \infty
\end{aligned} \tag{5.82}$$

Recall that $p_{V^{n'}} = p_{U^{n'}}$. Also, it is assumed that probability of excess distortion D is achievable with source-code s for uniform X source. That is,

$$\Pr\left(\frac{1}{n'}d^{n'}(V^{n'}, Y^{n'}) > D\right) = \epsilon_2^{n'} \to 0 \text{ as } n \to \infty \tag{5.83}$$

Thus,

$$\begin{aligned}
&\Pr\left(\frac{1}{n'}d_H^{n'}(X'^{n'}, V^{n'})D_{\max} + \frac{1}{n'}d^{n'}(V^{n'}, Y^{n'}) \le D_{\max}(d_1(p_X, p_{X'}) + \epsilon) + D\right)\\
&\ge \Pr\left(\frac{1}{n'}d_H^{n'}(X'^{n'}, V^{n'})D_{\max} \le D_{\max}(d_1(p_X, p_{X'}) + \epsilon) \text{ and } \frac{1}{n'}d^{n'}(U^{n'}, Y^{n'}) \le D\right)\\
&= \Pr\left(\left\{\frac{1}{n'}d_H^{n'}(X'^{n'}, V^{n'})D_{\max} \le D_{\max}(d_1(p_X, p_{X'}) + \epsilon)\right\} \cap \left\{\frac{1}{n'}d^{n'}(V^{n'}, Y^{n'}) \le D\right\}\right)\\
&\ge 1 - (\epsilon_1^{n'} + \epsilon_2^{n'}) \text{ by Lemma 5.5}
\end{aligned} \tag{5.84}$$

Thus,

$$\begin{aligned}
&\Pr\left(\frac{1}{n'}d_H^{n'}(X'^{n'}, V^{n'})D_{\max} + \frac{1}{n'}d^{n'}(V^{n'}, Y^{n'}) > D + D_{\max}(d_1(p_X, p_{X'}) + \epsilon)\right)\\
&\qquad \le \epsilon_1^{n'} + \epsilon_2^{n'} \to 0 \text{ as } n \to \infty
\end{aligned} \tag{5.85}$$

Thus,

$$\Pr\left(\frac{1}{n'}d^{n'}(X'^{n'}, Y^{n'}) > D_{\max}(d_1(p_X, p_{X'}) + \epsilon) + D\right) \to 0 \text{ as } n \to \infty \tag{5.86}$$

Thus, probability of excess distortion $D + D_{\max}(d_1(p_X, p_{X'}) + \epsilon)$ is achievable for coding i.i.d. X' source with rate R jump source-code s'. For each $\epsilon' > 0$, the jump source-code s' can be interpolated to form a rate R source-code s'' using which probability of excess distortion $D + D_{\max}(d_1(p_X, p_{X'}) + \epsilon) + \epsilon'$ is achievable when source-coding the i.i.d. X' source. $\epsilon > 0$ and $\epsilon' > 0$ are arbitrary. Choosing ϵ and ϵ' such that $D_{\max}\epsilon + \epsilon' = \delta$, the lemma follows. □

Lemma 5.27. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R sourc-code for which probability of excess distortion D is achievable when source-coding the i.i.d. X' source, then, there exists a rate R source-code for which probability of excess distortion $D + D_{\max}d_1(p_X, p_{X'}) + \delta$ is achievable when source-coding the uniform X source.*

Proof. Let t be a rate R source-code using which, probability of excess distortion D is achievable when source-coding the i.i.d. X' source. Recall the construction of source-code t' in Construction 5.2 which is used to source-code the uniform X source. We get a bound on the achievable probability of excess distortion for the source-code t' when sourc-coding the uniform X source.

Recall Definition 5.49 that $p_K = p_{U^{n'}T^{n'}Y^{n'}}$ is the joint distribution on $\mathcal{U}^{n'} \times \mathcal{X}^{n'} \times \mathcal{Y}^{n'}$ induced by $l'^{n'}$ and the source-code t when the distribution on $\mathcal{U}^{n'}$ is uniform. Recall also that $T^{n'}$ has the same distribution as i.i.d. X' source of block-length n', $X'^{n'}$. Also, by Lemma 5.15,

$$\frac{1}{n'}d^{n'}(u^{n'},y^{n'}) \leq \frac{1}{n'}d_H^{n'}(u^{n'},x^{n'})D_{\max} + \frac{1}{n'}d^{n}(x^{n'},y^{n'}) \tag{5.87}$$

Thus,

$$\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) \leq \frac{1}{n'}d_H^{n'}(U^{n'},T^{n'})D_{\max} + \frac{1}{n'}d^{n'}(T^{n'},Y^{n'}) \tag{5.88}$$

Let probability of excess distortion D be achievable with source-code t for the i.i.d. X' source. Then,

$$\begin{aligned}&\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D + D_{\max}(d_1(p_X,p_{X'})+\epsilon)\right)\\ &\leq \Pr\left(\frac{1}{n'}d_H^{n'}(U^{n'},T^{n'})D_{\max} + \frac{1}{n'}d^{n'}(T^{n'},Y^{n'}) > D_{\max}(d_1(p_X,p_{X'})+\epsilon)+D\right)\end{aligned} \tag{5.89}$$

$$\begin{aligned}&\Pr\left(\frac{1}{n'}d_H^{n'}(U^{n'},T^{n'}) > d_1(p_X,p_{X'})+\epsilon\right)\\ &= \Pr\left(\frac{1}{n'}d_H^{n'}(T^{n'},U^{n'}) > d_1(p_X,p_{X'})+\epsilon\right) = \epsilon_1^{n'} \to 0 \text{ as } n' \to \infty\\ &\qquad\qquad \text{since } \Pr(\xi_\epsilon^{n'}(T^{n'}) = \phi) \to 0 \text{ as } n \to \infty\\ &\Rightarrow \Pr\left(\frac{1}{n'}d_H^{n'}(U^{n'},T^{n'})D_{\max} > (d_1(p_X,p_{X'})+\epsilon)D_{\max}\right) = \epsilon_1^{n'} \to 0 \text{ as } n \to \infty\end{aligned} \tag{5.90}$$

Recall that $p_{T^{n'}} = p_{X'^{n'}}$. Also, it is assumed that probability of excess distortion D is achievable with source-code t for i.i.d. X' source . That is,

$$\Pr\left(\frac{1}{n'}d^{n'}(T^{n'},Y^{n'}) > D\right) = \epsilon_2^{n'} \to 0 \text{ as } n \to \infty \tag{5.91}$$

Thus,

$$\begin{aligned}&\Pr\left(\frac{1}{n'}d_H^{n'}(U^{n'},T^{n'})D_{\max} + \frac{1}{n'}d^{n'}(T^{n'},Y^{n'}) \leq D_{\max}(d_1(p_X,p_{X'})+\epsilon)+D\right)\\ &\geq \Pr\left(\frac{1}{n'}d_H^{n'}(U^{n'},T^{n'})D_{\max} \leq D_{\max}(d_1(p_X,p_{X'})+\epsilon) \text{ and } \frac{1}{n'}d^{n'}(T^{n'},Y^{n'}) \leq D\right)\\ &= \Pr\left(\left\{\frac{1}{n'}d_H^{n'}(U^{n'},T^{n'})D_{\max} \leq D_{\max}(d_1(p_X,p_{X'})+\epsilon)\right\} \cap \left\{\frac{1}{n'}d^{n'}(T^{n'},Y^{n'}) \leq D\right\}\right)\end{aligned}$$

$$\geq 1-(\epsilon_1^{n'}+\epsilon_2^{n'}) \text{ by Lemma 5.5} \quad (5.92)$$

Thus,

$$\Pr\left(\frac{1}{n'}d_H^{n'}(U^{n'},T^{n'})D_{\max}+\frac{1}{n'}d^{n'}(T^{n'},Y^{n'})>D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)\right) \leq \epsilon_1^{n'}+\epsilon_2^{n'}\to 0 \text{ as } n\to\infty \quad (5.93)$$

Thus,

$$\Pr\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'})>D_{\max}(d_1(p_X,p_{X'})+\epsilon)+D\right)\to 0 \text{ as } n\to\infty \quad (5.94)$$

Thus, probability of excess distortion $D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)$ is achievable for coding uniform X source with rate R source-code t'. Such a source-code t' exists $\forall\epsilon>0$. Choosing ϵ such that $D_{\max}\epsilon=\delta$, the lemma follows. □

Lemma 5.28. *Let X satisfy that $p_X(x)$ is rational $\forall X$. For $D\in(0,\infty), R_X^E(D)=R_U^E(D)$ and $R_X^P(D)=R_U^P(D)$. In particular, $R_U^E(D)$ and $R_U^P(D)$ are convex and continuous functions of D for $D\in(0,\infty)$.*

Proof. First, we prove that $R_X^E(D)=R_U^E(D)$.

By taking the distribution of X', the same as the distribution of X in Lemma 5.24, it follows, that $\forall\delta>0$,

$$R_X^E(D+\delta)\leq R_U^E(D) \quad (5.95)$$

By taking the distribution of X', the same as the distribution of X in Lemma 5.25, it follows that $\forall\delta>0$,

$$R_U^E(D+\delta)\leq R_X^E(D) \quad (5.96)$$

From the above two equations,

$$R_X^E(D+\delta)\leq R_U^E(D)\leq R_X^E(D-\delta)\forall 0<\delta<D \quad (5.97)$$

From Lemma 5.4, $R_X^E(D)$ is a continuous function of D for $D\in(0,\infty)$. By taking $\delta\to 0$ in the above equation, it follows that $\forall D\in(0,\infty), R_X^E(D)=R_U^E(D)$. Since $R_X^E(D)$ is convex and continuous for $D\in(0,\infty)$, so is $R_U^E(D)$.

Next, we prove that $R_X^P(D)=R_U^P(D)$.

By taking the distribution of X', the same as the distribution of X in Lemma 5.26, it follows that $\forall\delta>0$,

$$R_X^P(D+\delta)\leq R_U^P(D) \quad (5.98)$$

By taking the distribution of X', the same as the distribution of X in Lemma 5.27, it follows that $\forall \delta > 0$,

$$R_U^P(D+\delta) \leq R_X^P(D) \tag{5.99}$$

From the above two equations,

$$R_X^P(D+\delta) \leq R_U^P(D) \leq R_X^P(D-\delta) \forall 0 < \delta < D \tag{5.100}$$

From Lemma 5.6, $R_X^P(D)$ is a continuous function of D for $D \in (0,\infty)$. By taking $\delta \to 0$ in the above equation, it follows that $\forall D \in (0,\infty), R_X^P(D) = R_U^P(D)$. Since $R_X^P(D)$ is convex and continuous for $D \in (0,\infty)$, so is $R_U^P(D)$. □

Lemma 5.29. *Given a rate R source-code s for which expected distortion D is inf-achievable when source-coding the i.i.d. X' source. Then, there exists a rate R source-code s_1 for which expected distortion D is inf-achievable when source-coding the i.i.d. X' source, where limit is taken along a sequence where the block-lengths are divisible by n_0 to achieve the required inf-expected distortion.*

Proof. Let $s =< s^n >)_1^\infty$ be a rate R source-code for which expected distortion D is inf-achievable when source-coding the i.i.d. X' source. By Lemma 5.1, there exists a rate R deterministic source-code for which expected distortion D is inf-achievable when source-coding the i.i.d. X' source. Without loss of generality, assume that s is deterministic. That is, there exists a sequence $n_i \nearrow \infty$ such that

$$E\left[\frac{1}{n_i} d^{n_i}(X^{n_i}, Y^{n_i})\right] \leq D + \epsilon_i, \text{ where } \epsilon_i \to 0 \text{ as } n_i \nearrow \infty \tag{5.101}$$

Assume that $n_{i+1} - n_i > n_0$. If this were not the case, consider a subsequence of n_i such that this is the case, and re-label it to call it n_i. Let n_i' denote the least integer $\geq n_i$ such that n_i' is divisible by n_0. Note that $0 \leq n_i' - n_i < n_0$. Note that n_i' are distinct because of the assumption that $n_{i+1} - n_i > n_0$.

Form $s_1^{n_i'}$ from s^{n_i} as s^{kn+r} in the same way that from s^{kn} in the interpolation-code Definition 5.35.

Consider the source-code $s_1 =< s_1^n >_1^\infty$ which is such that $s_1^{n_i'}$ is defined as above and s_1^n is defined arbitrarily for other block lengths. It can be proved that this code satisfies the requirements of the lemma by an argument similar to the argument in the proof of Lemma 5.8. □

Lemma 5.30. *Given a rate R source-code s for which probability of excess distortion D is inf-achievable when source-coding the i.i.d. X' source. Let $\epsilon > 0$. Then, there exists a rate R source-code s_1 for which probability of excess distortion $D + \epsilon$ is inf-achievable when source-coding the i.i.d. X' source, where limit is taken along a sequence where the block-lengths are divisible by n_0 to achieve the required inf probability of excess distortion.*

Proof. Let $s = < s^n >_1^\infty$ be a rate R source-code for which probability of excess distortoin D is inf-achievable when source-coding the i.i.d. X' source. By Lemma 5.1, there exists a rate R deterministic source-code for which probability of excess distortion D is inf-achievable when source-coding the i.i.d. X' source. Without loss of generality, assume that s is deterministic. That is, there exists a sequence $n_i \nearrow \infty$ such that

$$\Pr\left(\frac{1}{n_i} d^{n_i}(X^{n_i}, Y^{n_i}) > D\right) \to 0 \text{ as } n_i \nearrow \infty \tag{5.102}$$

Assume that $n_{i+1} - n_i > n_0$. If this were not the case, consider a subsequence of n_i such that this is the case, and re-label it to call it n_i. Let n'_i denote the least integer $\geq n_i$ such that n'_i is divisible by n_0. Note that $0 \leq n'_i - n_i < n_0$. Note that n'_i are distinct because of the assumption that $n_{i+1} - n_i > n_0$.

Form $s_1^{n'_i}$ from s^{n_i} as s^{kn+r} in the same way that from s^{kn} in the interpolation-code Definition 5.35.

Consider the source-code $s_1 = < s_1^n >_1^\infty$ which is such that $s_1^{n'_i}$ is defined as above and s_1^n is defined arbitrarily for other block lengths. It can be proved by an argument similar to the argument in the proof of Lemma 5.9 that this code satisfies the requirements of the lemma. □

Lemma 5.31. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x \in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R source-code for which expected distortion D is inf-achievable when source-coding the uniform X source, then, there exists a rate R source-code for which expected distortion $D + D_{\max} d_1(p_X, p_{X'}) + \delta$ is inf-achievable when source-coding the i.i.d. X' source.*

Proof. Let s be a rate R source code using which, expected distortion D is inf-achievable when source-coding the uniform X source. Recall the construction of jump source code s' in Construction 5.1 which is used to source-code the i.i.d. X' source.

Rest of the proof is similar to the proof of Lemma 5.24, except that limits are taken along some subsequence $n_i \nearrow \infty$ instead of along n', and that, the interpolation argument is not needed.

□

Lemma 5.32. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x \in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R source-code for which expected distortion D is inf-achievable when source-coding the i.i.d. X' source, then, there exists a rate R source-code for which expected distortion $D + D_{\max} d_1(p_X, p_{X'}) + \delta$ is inf-achievable when source-coding the uniform X source.*

Proof. Let t be a rate R source-code using which, expected distortion D is inf-achievable when source-coding the i.i.d. X' source. That is, there exists a sub-sequence $n_i \nearrow \infty$ such that for the source-code t,

$$\mathrm{E}\left[\frac{1}{n_i}d^{n_i}(X^{n_i}, Y^{n_i})\right] \leq D + \epsilon_i \text{ where } \epsilon_i \to 0 \text{ as } n_i \nearrow \infty \tag{5.103}$$

Construct a source-code t_i from t as s_1 is constructed from s in Lemma 5.29. For the source-code t_1,

$$\mathrm{E}\left[\frac{1}{n'_i}d^{n'_i}(X^{n'_i}, Y^{n'_i})\right] \leq D + \epsilon_i \text{ where } \epsilon_i \to 0 \text{ as } n'_i \nearrow \infty \text{ and } n'_i \text{ is divisible by } n_0 \forall i \tag{5.104}$$

Rest of the argument follows the argument in the proof of Lemma 5.25 by looking at the source-code t'_1 which is constructed from t_1 by using Construction 5.2, except that limits are taken along $n'_i \nearrow \infty$ instead of n'.

The reason why we need to go from t to t_1 in this argument is that if none of the n_i is divisible by n_0, t'^{n_i} is not defined for any n_i, and thus, taking limits along n_i would not make sense for the source-code t', which is the main part of the argument in the proof of Lemma 5.25.

□

Lemma 5.33. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x \in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R source-code for which probability of excess distortion D is inf-achievable when source-coding the uniform X source, then, there exists a rate R source-code for which probability of excess distortion $D + D_{\max}d_1(p_X, p_{X'}) + \delta$ is inf-achievable when source-coding the i.i.d. X' source.*

Proof. Let s be a rate R source-code using which, probability of excess distortion D is inf-achievable when source-coding the uniform X source. Recall the construction of jump source-code s' in Construction 5.1 which is used to source-code the i.i.d. X' source.

Rest of the proof is similar to the proof of Lemma 5.26, except that limits are taken along some subsequence $n_i \nearrow \infty$ instead of along n', and that, the interpolation argument is not needed.

□

Lemma 5.34. *Let X' be arbitrary and X satisfy that $p_X(x)$ is rational $\forall x in \mathcal{X}$. Let $\delta > 0$. If there exists a rate R sourc-code for which probability of excess distortion D is inf-achievable when source-coding the i.i.d. X' source, then, there exists a rate R source-code for which probability of excess distortion $D + D_{\max}d_1(p_X, p_{X'}) + \delta$ is inf-achievable when source-coding the uniform X source.*

Proof. Let t be a rate R source-code using which, probability of excess distortion D is inf-achievable when source-coding the i.i.d. X' source. That is, there exists a sub-sequence $n_i \nearrow$

∞ such that for the source-code t,

$$\Pr\left(\frac{1}{n_i}d^{n_i}(X^{n_i},Y^{n_i}) > D\right) = \epsilon_i \text{ where } \epsilon_i \to 0 \text{ as } n_i \nearrow \infty \tag{5.105}$$

Construct a source-code t_i from t as s_1 is constructed from s in Lemma 5.30. For the source-code t_1,

$$\Pr\left(\frac{1}{n'_i}d^{n'_i}(X^{n'_i},Y^{n'_i})\right\rangle D) = \epsilon_i \text{ where } \epsilon_i \to 0 \text{ as } n'_i \nearrow \infty \text{ and } n'_i \text{ is divisible by } n_0 \forall i \tag{5.106}$$

Rest of the argument follows the argument in the proof of Lemma 5.27 by looking at the source-code t'_1 which is constructed from t_1 by using Construction 5.2, except that limits are taken along $n'_i \nearrow \infty$ instead of n'.

The reason why we need to go from t to t_1 in this argument is that if none of the n_i is divisible by n_0, t'^{n_i} is not defined for any n_i, and thus, taking limits along n_i would not make sense for the source-code t', which is the main part of the argument in the proof of Lemma 5.27.

□

Lemma 5.35. *Let X satisfy that $p_X(x)$ is rational $\forall X$. For $D \in (0,\infty)$, $R^E_X(D,\inf) = R^E_U(D,\inf)$ and $R^P_X(D,\inf) = R^P_U(D,\inf)$. In particular, $R^E_U(D,\inf)$ and $R^P_U(D,\inf)$ are convex and continuous functions of D for $D \in (0,\infty)$.*

Proof. The proof is analogous to the proof of Lemma 5.28. □

Next, we want to prove the equality of the rate-distortion functions with the expected distortion and the probability of excess distortion definitions for the uniform X source.

Equality of the rate-distortion function for the uniform X source with the expected and the probability of excess distortion definitions: $R^E_U(D) = R^P_U(D)$

We prove that $R^E_U(D) = R^P_U(D)$. This is the bridge between our results for the i.i.d. X and the uniform X sources. This result is interesting in its own right, and so is the proof technique.

Lemma 5.36. *Let X satisfy that $p_X(x)$ is rational $\forall x$. Then, for $D \in (0,\infty)$, $R^E_U(D) = R^P_U(D)$*

Proof. First, we prove that $R^E_U(D) \le R^P_U(D)$. The idea of the proof is that the probability of excess distortion criterion is "stronger" than the expected probability of error criterion. That is, if a particular probability of excess distortion level is achievable for some source, the same expected distortion is also achievable by the same source-code for the same source. A rigorous proof is the following:

Recall the definitions of n_0 and n'.

Let probability of excess distortion D be achievable for the uniform X source with source-code $s =< s^{n'} >_1^\infty$. Then, for the source-code s,

$$\Pr\left[\frac{1}{n'}d^{n'}(U^{n'}, Y^{n'}) > D\right] = \epsilon^{n'} \to 0 \text{ as } n \to \infty \tag{5.107}$$

It follows from the above equation that

$$E\left[\frac{1}{n'}d^{n'}(U^{n'}, Y^{n'})\right] \le (1-\epsilon^{n'})D + \epsilon^{n'} D_{\max} \to D \text{ as } n \to \infty \tag{5.108}$$

Thus, the expected distortion D is achievable for the uniform X source by use of the same source-code, and in particular, by a source-code of the same rate. It follows that $R_U^E(D) \le R_U^P(D)$.

Next, we prove that $R_U^P(D) \le R_U^E(D)$.

Let expected distortion D be achievable by a rate R source-code $t =< t^{n'} >_1^\infty$ when encoding the uniform X source. By Lemma 5.1, there exists a rate R deterministic source-code $s =< s^{n'} >_1^\infty =< e^{n'}, f^{n'} >_1^\infty$ using which expected distortion D is achievable for the uniform X source. That is, for the source-code s,

$$\begin{aligned}
&\limsup_{n\to\infty} \sum_{u^{n'} \in \mathcal{U}^{n'}} p_{U^{n'}}(u^{n'}) D_{u^{n'}} \le D \text{ where } D_{u^{n'}} \triangleq \frac{1}{n'} d^{n'}(u^{n'}, f^{n'} \circ e^{n'}(u^{n'})) \\
&\Rightarrow \frac{1}{|\mathcal{U}^{n'}|} \sum_{u^{n'} \in \mathcal{U}^{n'}} D_{u^{n'}} \le D + \epsilon^{n'} \text{ where } \epsilon^{n'} \to 0 \text{ as } n \to \infty
\end{aligned} \tag{5.109}$$

Let $\epsilon > 0$, $0 < \delta < 1$ be such that $(1-\delta)|\mathcal{U}^{n'}|$ out of the $|\mathcal{U}^{n'}|$ many $D_{u^{n'}}$ are $\ge D + \epsilon$. We find a relation between ϵ and δ below. From the above equation, it follows that

$$\begin{aligned}
&(1-\delta)(D+\epsilon) \le D + \epsilon^{n'} \\
&\Rightarrow \delta \ge \frac{\epsilon - \epsilon^{n'}}{D+\epsilon} \ge \frac{\epsilon}{2(D+\epsilon)} \text{ for } n' \text{ sufficiently large since } \epsilon^{n'} \to 0 \text{ as } n \to \infty
\end{aligned} \tag{5.110}$$

Thus, for sufficiently large n', possibly depending on ϵ, $\ge \delta|\mathcal{U}^{n'}| = \dfrac{\epsilon|\mathcal{U}^{n'}|}{2(D+\epsilon)}$ out of the $|\mathcal{U}^{n'}|$ many $u^{n'}$ satisfy $D_{u^{n'}} \le D + \epsilon$. *Note that this is true for all $\epsilon > 0$ and that, δ is independent of n.*

Let $\mathcal{U}_\delta^{n'}$ be a set of cardinality $\delta|\mathcal{U}^{n'}|$ such that for sufficiently large n', $D_{u^{n'}} \leq D+\epsilon \forall u^{n'} \in \mathcal{U}_\delta^{n'}$. Note that $p_{U^{n'}}(\mathcal{U}_\delta^{n'}) = \delta$, by definition. Let $U_\delta^{n'}$ denote the source which puts a uniform distribution on $\mathcal{U}_\delta^{n'}$. Denote $\mathcal{U}_\delta =< \mathcal{U}_\delta^{n'} >_1^\infty$. Denote $U_\delta =< U_\delta^{n'} >_1^\infty$.

The source-code $s =< s^{n'} >_1^\infty =< e^{n'}, f^{n'} >_1^\infty$ can be used to code the source U_δ by just restricting the source-code input space to $\mathcal{U}_\delta$. s still has rate R and when coding the source U_δ. With source-code s, probability of excess distortion $D+\epsilon$ is achievable for the source U_δ. In fact, the stronger condition is true that $\frac{1}{n'}d^{n'}(u_\delta^{n'}, f_\delta^{n'} \circ e_\delta^{n'}(u_\delta^{n'})) \leq D+\epsilon \ \forall u_\delta^{n'} \in \mathcal{U}_\delta^{n'}$. Thus, it follows that $R_V^P(D+\epsilon) \leq R_U^E(D)$ for any source $V =< V^{n'} >_1^\infty$ on $\mathcal{U}_\delta =< \mathcal{U}_\delta^{n'} >_1^\infty$. In particular, $R_{U_\delta}^P(D+\epsilon) \leq R_U^E(D)$.

We now cover the whole space $\mathcal{U}^n$ by polynomially many "copies" of $\mathcal{U}_\delta^n$ and consequently, $\forall \beta > 0$, construct a rate $R+\beta$ source-code for which the probability of excess distortion $D+\epsilon$ achievable when coding the uniform X source, U. Precisely, this is done as follows:

Recall Definition 5.43 for the definition of permutations.

$\pi^{n'}\mathcal{U}_\delta^{n'} \triangleq \{u^{n'} \in \mathcal{U}^{n'} : u^{n'} = \pi^{n'}u_\delta^{n'} \text{ for some } u_\delta^{n'} \in \mathcal{U}_\delta^{n'}\}$. $\pi^{n'}\mathcal{U}_\delta^{n'}$ is, what we call, a copy of $\mathcal{U}_\delta^{n'}$, got by permuting the entries of each element of $\mathcal{U}_\delta^{n'}$ by the permutation $\pi^{n'}$.

Note that $\cup_{\pi^{n'} \in \Pi^{n'}} \pi^n \mathcal{U}_\delta^{n'} = \mathcal{U}^{n'}$. Let $\mathcal{A}^{n'}$ be a subset of $\Pi^{n'}$ of smallest cardinality such that $\cup_{\pi^{n'} \in \mathcal{A}^{n'}} \pi^{n'}\mathcal{U}_\delta^{n'} = \mathcal{U}^{n'}$. We will prove that

$$|\mathcal{A}^{n'}| \leq \frac{n' \log(|\mathcal{X}|)}{\log \frac{1}{1-\delta}} \tag{5.111}$$

Assuming this, we first finish the proof that $R_U^P(D) \leq R_U^E(D)$.

For $\pi^{n'} \in \mathcal{A}^{n'}$, let $\mathcal{B}_{\pi^n}$ be disjoint sets such that $\mathcal{B}_{\pi^n} \subset \pi^n \mathcal{U}_\delta^n$ and $\cup_{\pi^n \in \mathcal{A}} B_{\pi^n} = \mathcal{U}$.

For $\pi^{n'} \in \mathcal{A}^{n'}$, let $\mathcal{K}_{\pi^{n'}}$ denote arbitrary sets, each of cardinality $2^{\lfloor n'R \rfloor}$. Let

$$i_{\pi^{n'}} : \{1, 2, \ldots, 2^{\lfloor n'R \rfloor}\} \to \mathcal{K}_{\pi^{n'}} \tag{5.112}$$

be arbitrary *bijective* maps.

Define maps $(\pi^{n'}e^{n'}, \pi^{n'}f^{n'})$ as follows:

$$\begin{aligned} \pi^{n'}e^{n'} &: \mathcal{B}_{\pi^{n'}} \to \mathcal{K}_{\pi^{n'}} \\ \pi^{n'}f^{n'} &: \mathcal{K}_{\pi^{n'}} \to \mathcal{Y}^{n'} \end{aligned} \tag{5.113}$$

$$\pi^{n'}e^{n'}(\pi^{n'}u_\delta^{n'}) = i_{\pi^{n'}}(e^{n'}(u_\delta^{n'})), \text{ for } \pi^{n'}u_\delta^{n'} \in \mathcal{B}_{\pi^{n'}}$$

$$\pi^{n'} f^{n'}(k_{\pi^{n'}}) = f^{n'}(i^{-1}_{\pi^{n'}}(k_{\pi^{n'}})) \tag{5.114}$$

Define source-code $s_* =< s_*^{n'} >_1^{\infty} =< e_*^{n'}, f_*^{n'} >_1^{\infty}$ as follows.

$$\begin{aligned} e_*^{n'} &: \mathcal{U}^{n'} \to \cup_{\pi^{n'} \in \mathcal{A}^{n'}} \mathcal{K}_{\pi^{n'}} \\ f_*^{n'} &: \cup_{\pi^{n'} \in \mathcal{A}^{n'}} \mathcal{K}_{\pi^{n'}} \to \mathcal{Y}^{n'} \end{aligned} \tag{5.115}$$

$$\begin{aligned} e_*^{n'}(u^{n'}) &= \pi^{n'} e^{n'}(u^{n'}) \text{ if } u^{n'} \in \mathcal{B}_{\pi^{n'}} \\ f_*^{n'}(k) &= \pi^{n'} f^{n'}(k) \text{ if } k \in K_{\pi^{n'}} \end{aligned} \tag{5.116}$$

Let $\beta > 0$. For sufficiently large n',

$$\begin{aligned} |\cup_{\pi^{n'} \in \mathcal{A}^{n'}} \mathcal{K}_{\pi^{n'}}| = |\mathcal{A}^{n'}||\mathcal{K}_{\pi^{n'}}| &= \frac{n' \log(|\mathcal{X}|)}{\log \frac{1}{1-\delta}} 2^{\lfloor n'R \rfloor} \\ &\leq 2^{\lfloor n'(R+\beta) \rfloor} \text{ for } n' > z \text{ for some integer } z \end{aligned} \tag{5.117}$$

The source-code s_* can be suitably modified for block-lengths $n' < z$ in an arbitrary way such that the source-code s_* has rate $R+\beta$. For $n' > z$, it follows by construction that $\forall u^{n'} \in \mathcal{U}^{n'}$,

$$\frac{1}{n'} d^{n'}(u^{n'}, f_*^{n'}(e_*^{n'}(u^{n'}))) \leq D + \epsilon \tag{5.118}$$

It follows that for an arbitrary source $V =< V^{n'} >_1^{\infty}$ on $\mathcal{U}$, probability of excess distortion $D + \epsilon$ is achievable with rate $R + \delta$ source-code s_*. Thus, $R_V^P(D+\epsilon) \leq R_U^E(D) + \beta$. This is true for all $\beta > 0$. Thus, $R_V^P(D+\epsilon) \leq R_U^E(D)$.

In particular, $R_U^P(D+\epsilon) \leq R_U^E(D)$. By Lemma 5.28, $R_U^P(D)$ is continuous for $D \in (0,\infty)$. It follows that $R_U^P(D) \leq R_U^E(D)$ for $D \in (0,\infty)$.

It remains to prove that

$$|\mathcal{A}| \leq \frac{n \log(|\mathcal{X}|)}{\log \frac{1}{1-\delta}} \tag{5.119}$$

This is proved below.

Let $u^{n'}, u'^{n'} \in \mathcal{U}^{n'}$.

First, we want to calculate

$$\Pr(u^{n'} = p^{n'} u'^{n'}) \tag{5.120}$$

Fix $u'^{n'}$. The above probability is independent of $u^{n'}$ by symmetry. An elaborate argument which shows how this symmetry works is the following:

Denote $\mathscr{B}_{u^{n'}} \triangleq \{\pi^{n'} \in \Pi^{n'} \mid u^{n'} = P^{n'} u'^{n'}\}$. Let $u_1^{n'}, u_2^{n'} \in \mathscr{U}^{n'}$. Thus, there are the corresponding sets $\mathscr{B}_{u_1^{n'}}$ and $\mathscr{B}_{u_2^{n'}}$. $u_2^{n'} = \pi_0^{n'} u_1^{n'}$. Thus, if $u'^{n'} = \pi^{n'} u_2^{n'}$, then, $u'^{n'} = \pi^{n'} \pi_0^{n'} u_1^{n'}$. It follows that $\{\pi^{n'} \pi_0^{n'} \mid \pi^{n'} \in \mathscr{B}_{u_2^{n'}}\} \subset \mathscr{B}_{u_1^{n'}}$. If $\pi_1^{n'} \neq \pi_2^{n'}$, then, $\pi_1^{n'} \pi_0^{n'} \neq \pi_2^{n'} \pi_0^{n'}$. It follows that $|B_{u_1^{n'}}| \geq |B_{u_2^{n'}}|$. By interchanging $u_1^{n'}$ and $u_2^{n'}$, it follows that $|B_{u_1^{n'}}| \leq |B_{u_2^{n'}}|$. Thus, $|B_{u_1^{n'}}| = |B_{u_2^{n'}}|$. It follows that $\Pr(u^{n'} = P^{n'} u'^{n'})$ is independent of $u^{n'}$. Thus,

$$\Pr(u^{n'} = P^{n'} u'^{n'}) = \frac{1}{|\mathscr{U}^{n'}|} \tag{5.121}$$

Let $u^{n'}, u'^{n'}, u''^{n'} \in \mathscr{U}^{n'}$. From Lemma 5.18, it follows that $\pi^{n'} u'^{n'} = \pi^{n'} u''^{n'} \Rightarrow u'^{n'} = u''^{n'}$. Thus,

$$\Pr(u^{n'} \in P^{n'} \mathscr{U}_\delta^{n'}) = \frac{|\mathscr{U}_\delta^{n'}|}{|\mathscr{U}^{n'}|} = \delta \tag{5.122}$$

Let $P_1^{n'}, P_2^{n'}, \ldots, P_t^{n'}$ be independent, uniform random variables on $\Pi^{n'}$. Then,

$$\begin{aligned}
&\Pr(u^{n'} \notin P_1^{n'} \mathscr{U}_\delta^{n'} \cup P_2^{n'} \mathscr{U}_\delta^{n'} \cup \cdots \cup P_t^{n'} \mathscr{U}_\delta^{n'}) \\
&= \Pr(\{u^{n'} \notin P_1^{n} \mathscr{U}_\delta^{n'}\} \text{ and } \{u^{n'} \notin P_2^{n} \mathscr{U}_\delta^{n'}\} \text{ and } \ldots \text{ and } \{u^{n'} \notin P_t^{n'} \mathscr{U}_\delta^{n'}\}) \\
&= \Pr(\{u^{n'} \notin P_1^{n'} \mathscr{U}_\delta^{n'}\} \cap \{u^{n'} \notin P_2^{n'} \mathscr{U}_\delta^{n'}\} \cap \ldots \cap \{u^{n'} \notin P_t^{n'} \mathscr{U}_\delta^{n'}\}) \\
&= \Pr(u^{n'} \notin P^{n'} \mathscr{U}_\delta^{n'}) \times \Pr(u^{n'} \notin P^{n'} \mathscr{U}_\delta^{n'}) \times \cdots \times \Pr(u^{n'} \notin P^{n'} \mathscr{U}_\delta^{n'}) \\
&= (1-\delta)^t
\end{aligned} \tag{5.123}$$

$$\begin{aligned}
&\Pr(P_1^{n'} \mathscr{U}_\delta^{n'} \cup P_2^{n'} \mathscr{U}_\delta^{n'} \cup \cdots \cup P_t^{n'} \mathscr{U}_\delta^{n'} \neq \mathscr{U}^{n'}) \\
&= \Pr(\exists u^{n'} \in \mathscr{U}^{n'} \text{ such that } u^{n'} \notin P_1^{n'} \mathscr{U}_\delta^{n'} \cup P_2^{n'} \mathscr{U}_\delta^{n'} \cup \cdots \cup P_t^{n'} \mathscr{U}_\delta^{n'}) \\
&= \Pr(\cup_{u^{n'} \in \mathscr{U}^{n'}} \{u^{n'} \notin P_1^{n'} \mathscr{U}_\delta^{n'} \cup P_2^{n'} \mathscr{U}_\delta^{n'} \cup \cdots \cup P_t^{n'} \mathscr{U}_\delta^{n'}\}) \\
&\leq \sum_{u^{n'} \in \mathscr{U}^{n'}} \Pr(\{u^{n'} \notin P_1^{n'} \mathscr{U}_\delta^{n'} \cup P_2^{n'} \mathscr{U}_\delta^{n'} \cup \cdots \cup P_t^{n'} \mathscr{U}_\delta^{n'}\}) \text{ by the union bound} \\
&= |\mathscr{U}^{n'}|(1-\delta)^t \\
&\leq |\mathscr{X}^{n'}|(1-\delta)^t \\
&< 1 \text{ if } t > \frac{n' \log |\mathscr{X}|}{\log \frac{1}{1-\delta}}
\end{aligned} \tag{5.124}$$

It follows that there exists a subset $\mathscr{A}^{n'}$ of the set of permutations $\Pi^{n'}$ of cardinality $\leq \frac{n' \log|\mathscr{X}|}{\log \frac{1}{1-\delta}}$, such that $\cup_{\pi^{n'} \in \mathscr{A}^{n'}} \pi^{n'} \mathscr{U}_\delta^{n'} = \mathscr{U}^{n'}$. This completes the proof. □

We now integrate the above results to prove the equality of all rate-distortion functions when X satisfies that $p_X(x)$ is rational $\forall x \in \mathscr{X}$

Equality of all possible rate-distortion functions for sources which satisfy $p_X(x)$ is rational $\forall x \in \mathscr{X}$

The following theorem proves equality of all possible rate distortion functions for the i.i.d. and uniform X sources when X satisfies that $p_X(x)$ is rational $\forall x \in \mathscr{X}$.

Theorem 5.37. *For X such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Then, for $D \in (0,\infty)$, $R_X^E(D) = R_X^P(D) = R_X^E(D,j) = R_X^P(D,j) = R_X^E(D,\text{inf}) = R_X^P(D,\text{inf}) = R_U^E(D) = R_U^P(D) = R_U^E(D,\text{inf}) = R_U^P(D,\text{inf})$.*

Proof. This follows from Lemmas 5.10, 5.14, 5.28, 5.35, and 5.36. □

Note 5.48. *Note that we do not directly prove that $R_X^E(D) = R_X^P(D)$. We prove $R_U^E(D) = R_U^P(D)$, $R_X^E(D) = R_U^E(D)$, and $R_X^P(D) = R_U^P(D)$, in order to prove that $R_X^E(D) = R_X^P(D)$. This can be thought of as one reason for the introduction of the uniform X source. However, this is not the most important reason for the introduction of the uniform X source. As stated before, the main reason for the introduction of the uniform X source is that we do not know, how to prove Theorems 5.46 and 5.47, directly for the i.i.d. X source.*

Finally, we prove the equality of the rate-distortion functions for arbitrary X.

Equality of all possible rate-distortion functions for the i.i.d. X source: proofs that $R_X^E(D) = R_X^P(D) = R_X^E(D,\text{inf}) = R_X^P(D,\text{inf})$

Theorem 5.37 proves equality of all possible rate-distortion functions for the i.i.d. X source when X is such that $p_X(x)$ is rational $\forall x$. Note that uniform X sources are undefined for arbitrary X. For this, we will carry out limiting arguments with random variables $X_n \to X$ where X_n is such that $p_{X_n}(x)$ is rational $\forall x \in \mathscr{X}$. This is the final step in this subsection, after which, we will move on to channels. The limiting arguments will be useful in their own right, when we consider channels.

Lemma 5.38. *For arbitrary X, for $D \in (0,\infty)$, $0 < \epsilon < D$,*

$$R_X^E(D) - R_X^E(D+\epsilon) \leq \frac{\epsilon \log|\mathscr{X}|}{D} \tag{5.125}$$

$$R_X^P(D) - R_X^P(D+\epsilon) \le \frac{\epsilon \log|\mathscr{X}|}{D} \tag{5.126}$$

(5.127)

For X such that $p_X(x)$ is rational $\forall x$, $D \in (0,\infty)$, $0 < \epsilon < D$,

$$R_U^E(D) - R_U^E(D+\epsilon) \le \frac{\epsilon \log|\mathscr{X}|}{D} \tag{5.128}$$

$$R_U^P(D) - R_U^P(D+\epsilon) \le \frac{\epsilon \log|\mathscr{X}|}{D} \tag{5.129}$$

Proof. By Lemma 5.4, $R_X^E(D)$ is convex for $D \in [0,\infty)$. It follows that

$$\frac{R_X^E(D) - R_X^E(D+\epsilon)}{\epsilon} \le \frac{R_X^E(0) - R_X^E(D)}{D} \le \frac{\log|\mathscr{X}|}{D}$$

$$\Rightarrow R_X^E(D) - R_X^E(D+\epsilon) \le \frac{\epsilon \log|\mathscr{X}|}{D} \tag{5.130}$$

This proves the first statement in the lemma.

By Lemma 5.6, $R_X^P(D)$ is convex for $D \in [0,\infty)$. The second statement in the lemma follows exactly as above.

For X such that $p_X(x)$ is rational $\forall x$, by Lemma 5.28, for $D \in (0,\infty)$, $R_U^E(D) = R_X^E(D)$ and $R_U^P(D) = R_X^P(D)$. The third and fourth statements in the lemma, now follow by using the first and second statements in the lemma which have been proved above. □

Lemma 5.39. *Let X be an* arbitrary *random variable on $\mathscr{X}$. Let X_n, $1 \le n \le \infty$ be a collection of random-variables on $\mathscr{X}$ such that $p_{X_n}(x)$ is rational $\forall x \in \mathscr{X}, \forall X_n$. Also, let $X_n \to X$ in distribution. That is, $\forall x \in X$, $\lim_{n\to\infty} p_{X_n}(x) = p_X(x)$. U_n is the uniform X_n source. Let $\delta_n > 0$ and let $\delta_n \to 0$ as $n \to \infty$ Define $\epsilon_n \triangleq D_{\max} d_1(p_X, p_{X_n}) + \delta_n$. Then, for $D \in (0,\infty)$,*

$$\lim_{n\to\infty} R_{U_n}^E(D+\epsilon_n) = \lim_{n\to\infty} R_{X_n}^E(D+\epsilon_n) = \lim_{n\to\infty} R_{U_n}^P(D+\epsilon_n) = \lim_{n\to\infty} R_{X_n}^P(D+\epsilon_n) = R_X^E(D) \tag{5.131}$$

Proof. Note that $\epsilon_n \to 0$ as $n \to \infty$.

By Lemma 5.24, and by an argument similar to that used in Lemma 5.28, for all $delta_n > 0$,

$$R_X^E(D + D_{\max} d_1(p_X, p_{X_n}) + \delta_n) \le R_{U_n}^E(D)$$

$$\Rightarrow R_X^E(D) \le R_{U_n}^E(D - (D_{\max} d_1(p_X, p_{X_n}) + \delta_n)) \tag{5.132}$$

By Lemma 5.25, and by an argument similar to that used in Lemma 5.28, $\forall \delta_n > 0$,

$$R^E_{U_n}(D + D_{\max} d_1(p_X, p_{X_n}) + \delta_n) \leq R^E_X(D) \tag{5.133}$$

Thus,

$$\begin{aligned}
&R^E_{U_n}(D + \epsilon_n) \leq R^E_X(D) \leq R^E_{U_n}(D - \epsilon_n) \\
&\Rightarrow 0 \leq R^E_X(D) - R^E_{U_n}(D + \epsilon_n) \leq R^E_{U_n}(D - \epsilon_n) - R^E_{U_n}(D + \epsilon_n) \\
&\Rightarrow 0 \leq R^E_X(D) - R^E_{U_n}(D + \epsilon_n) \leq [R^E_{U_n}(D - \epsilon_n) - R^E_{U_n}(D)] + [R^E_{U_n}(D) - R^E_{U_n}(D + \epsilon_n)] \\
&\Rightarrow 0 \leq R^E_X(D) - R^E_{U_n}(D + \epsilon_n) \leq \frac{\delta_n \log|\mathcal{X}|}{D - \epsilon_n} + \frac{\epsilon_n \log|\mathcal{X}|}{D} \text{ by Lemma 5.38}
\end{aligned} \tag{5.134}$$

Taking limit as $n \to \infty$, and recalling that $\epsilon_n \to 0$ as $n \to \infty$, it follows that

$$\lim_{n\to\infty} R^E_{U_n}(D + \epsilon_n) = R^E_X(D) \tag{5.135}$$

By Theorem 5.37, $R^E_{U_n}(D+\epsilon_n) = R^E_{X_n}(D+\epsilon_n) = R^P_{U_n}(D+\epsilon_n) = R^P_{X_n}(D+\epsilon_n)$. It thus follows that

$$\begin{aligned}
\lim_{n\to\infty} R^E_{U_n}(D + \epsilon_n) &= \lim_{n\to\infty} R^E_{X_n}(D + \epsilon_n) = \\
&\lim_{n\to\infty} R^P_{U_n}(D + \epsilon_n) = \lim_{n\to\infty} R^P_{X_n}(D + \epsilon_n) = R^E_X(D)
\end{aligned} \tag{5.136}$$

The lemma follows. □

Lemma 5.40. *Let X be an* arbitrary *random variable on $\mathcal{X}$. Let $X_n, 1 \leq n \leq \infty$ be a collection of random-variables on $\mathcal{X}$ such that $p_{X_n}(x)$ is rational $\forall x \in \mathcal{X}, \forall X_n$. Also, let $X_n \to X$ in distribution. That is, $\forall x \in X$, $\lim_{n\to\infty} p_{X_n}(x) = p_X(x)$. U_n is the uniform X_n source. Let $\delta_n > 0$ and let $\delta_n \to 0$ as $n \to \infty$. Define $\epsilon_n \triangleq D_{\max} d_1(p_X, p_{X_n}) + \delta_n$. Then, for $D \in (0, \infty)$,*

$$\begin{aligned}
\lim_{n\to\infty} R^E_{U_n}(D + \epsilon_n) &= \lim_{n\to\infty} R^E_{X_n}(D + \epsilon_n) = \\
&\lim_{n\to\infty} R^P_{U_n}(D + \epsilon_n) = \lim_{n\to\infty} R^P_{X_n}(D + \epsilon_n) = R^P_X(D)
\end{aligned} \tag{5.137}$$

Proof. Note that $\epsilon_n \to 0$ as $n \to \infty$.

By Lemma 5.26, and by an argument similar to that used in Lemma 5.28, for all $\delta_n > 0$,

$$\begin{aligned}
&R^P_X(D + D_{\max} d_1(p_X, p_{X_n}) + \delta_n) \leq R^P_{U_n}(D) \\
&\Rightarrow R^P_X(D) \leq R^P_{U_n}(D - (D_{\max} d_1(p_X, p_{X_n}) + \epsilon_n))
\end{aligned} \tag{5.138}$$

By Lemma 5.27, and by an argument similar to that used in Lemma 5.28, $\forall \delta_n > 0$,

$$R^P_{U_n}(D + D_{\max} d_1(p_X, p_{X_n}) + \delta_n) \leq R^P_X(D) \tag{5.139}$$

Thus,

$$\begin{aligned}
&R^P_{U_n}(D+\epsilon_n) \leq R^P_X(D) \leq R^P_{U_n}(D-\epsilon_n) \\
&\Rightarrow 0 \leq R^P_X(D) - R^P_{U_n}(D+\epsilon_n) \leq R^P_{U_n}(D-\epsilon_n) - R^P_{U_n}(D+\epsilon_n) \\
&\Rightarrow 0 \leq R^P_X(D) - R^P_{U_n}(D+\epsilon_n) \leq [R^P_{U_n}(D-\epsilon_n) - R^P_{U_n}(D)] + [R^P_{U_n}(D) - R^P_{U_n}(D+\epsilon_n)] \\
&\Rightarrow 0 \leq R^P_X(D) - R^P_{U_n}(D+\epsilon_n) \leq \frac{\epsilon_n \log|\mathscr{X}|}{D-\epsilon_n} + \frac{\epsilon_n \log|\mathscr{X}|}{D} \text{ by Lemma 5.38}
\end{aligned} \tag{5.140}$$

Taking limit as $n \to \infty$, and recalling that $\epsilon_n \to 0$ as $n \to \infty$, it follows that

$$\lim_{n\to\infty} R^P_{U_n}(D+\epsilon_n) = R^P_X(D) \tag{5.141}$$

By Theorem 5.37, $R^P_{U_n}(D+\epsilon_n) = R^P_{X_n}(D+\epsilon_n) = R^E_{U_n}(D+\epsilon_n) = R^E_{X_n}(D+\epsilon_n)$. It thus follows that

$$\begin{aligned}
&\lim_{n\to\infty} R^P_{U_n}(D+\epsilon_n) = \lim_{n\to\infty} R^P_{X_n}(D+\epsilon_n) = \\
&\qquad \lim_{n\to\infty} R^E_{U_n}(D+\epsilon_n) = \lim_{n\to\infty} R^E_{X_n}(D+\epsilon_n) = R^P_X(D)
\end{aligned} \tag{5.142}$$

The lemma follows.

□

Lemma 5.41. *For arbitrary X, for $D \in (0,\infty)$, $R^P_X(D) = R^E_X(D)$.*

Proof. This follows from Lemmas 5.39 and 5.40. □

The following is the theorem which proves the equality of all possible rate-distortion functions for the i.i.d. X source when X is arbitrary.

Theorem 5.42. *For arbitrary X, for $D \in (0,\infty)$, $R^E_X(D) = R^E_X(D,j) = R^E_X(D,\inf) = R^P_X(D) = R^P_X(D,j) = R^P_X(D,\inf)$*

Proof. This follows from Lemmas 5.10, 5.14, and 5.41. □

This ends this section. To re-capitulate, we proved the equality of various rate-distortion functions for the i.i.d. X and the uniform X sources. Many of the results and proof techniques are interesting in their own right. More importantly, we will require these results, in particular, the equality of the rate-distortion functions for the i.i.d. X and uniform X sources when proving the desired result of the equality of the pseudo universal channel capacity of the set of channels $\mathscr{C}_{X,D}$ and the rate-distortion function for the i.i.d. X source under the expected and the probability of excess distortion definitions. Before we do that, we first define, rigorously, the set of channels and prove various results concerning channels.

5.5 The channel-coding problem

In this section, we discuss the channel-coding problem. We define channels and jump channels. Then, we define what it means for a channel or a jump channel to *pseudo*-directly communicate a source to within a particular distortion level. This is followed by the definition of the pseudo-universal capacity of the set of channels which pseudo-directly communicate i.i.d. X source and uniform X source to within particular distortion levels. Finally, we derive relations between the various pseudo-universal capacities defined in this section.

5.5.1 Channels

We will consider 3 sets of channels:

1. Channels with input space $< \mathscr{X}^n >_1^\infty$ and output space $< \mathscr{Y}^n >_1^\infty$. We will see, what it means to communicate i.i.d X source pseudo-directly over a channel to within a particular distortion level.

2. For some k, channels with input space $< \mathscr{X}^{kn} >_1^\infty$ and output space $< \mathscr{Y}^{kn} >_1^\infty$. These are jump channels with jump k.

3. Channels with input space $< \mathscr{U}^{n'} >_1^\infty$ and output space $< \mathscr{Y}^{n'} >_1^\infty$. Recall that $n' = n_0 n$ where n_0 is the least positive integer for which $n_0 p_X(x)$ is an integer $\forall x \in \mathscr{X}$ for some random variable X on $\mathscr{X}$. Thus, for channels of this kind, one should think of an underlying random variable X which is such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. This random-variable will come into the picture when defining pseudo-direct communication of the uniform X source over a channel to within a particular distortion level.

Definition 5.50 (Channels with input space $< \mathscr{X}^n >_1^\infty$ and output space $< \mathscr{Y}^n >_1^\infty$). A channel is a sequence $c =< c^n >_1^\infty$ where $c^n : \mathscr{X}^n \to \mathscr{P}(\mathscr{Y}^n)$ is a transition probability/stochastic kernel. This should be interpreted as follows. When the block-length is n, the channel acts as c^n. For $x^n \in \mathscr{X}^n, y^n \in \mathscr{Y}^n$, $c^n(y^n|x^n)$ is the probability that the channel output is y^n given that the channel input is x^n.

Note 5.49. The definition of transition probability corresponding to a source-code s, q_s, defined in Definition 5.7 is exactly the same as that of a channel defined above.

Definition 5.51 (Channels with input space $< \mathcal{X}^{kn} >_1^\infty$ and output space $< \mathcal{Y}^{kn} >_1^\infty$: jump channels with jump k). A channel is a sequence $c =< c^{kn} >_1^\infty$ where $c^{kn} : \mathcal{X}^{kn} \to \mathcal{P}(\mathcal{Y}^{kn})$ is a transition probability/stochastic kernel. This should be interpreted as follows. When the block-length is n, the channel acts as c^{kn}. For $x^{kn} \in \mathcal{X}^{kn}, y^{kn} \in \mathcal{Y}^{kn}$, $c^{kn}(y^{kn}|x^{kn})$ is the probability that the channel output is y^{kn} given that the channel input is x^{kn}. These will be called jump channels with jump k.

Note 5.50. The definition of transition probability corresponding to a jump source-code s, q_s, defined in Definition 5.10 is exactly the same as that of a channel defined above.

Definition 5.52 (Channels with input space $< \mathcal{U}^{n'} >_1^\infty$ and output space $< \mathcal{Y}^{n'} >_1^\infty$). A channel is a sequence $c =< c^{n'} >_1^\infty$ where $c^{n'} : \mathcal{U}^{n'} \to \mathcal{P}(\mathcal{Y}^{n'})$ is a transition probability/stochastic kernel. This should be interpreted as follows. When the block-length is n', the channel acts as $c^{n'}$. For $u^{n'} \in \mathcal{U}^{n'}, y^{n'} \in \mathcal{Y}^{n'}$, $c^{n'}(y^{n'}|x^{n'})$ is the probability that the channel output is $y^{n'}$ given that the channel input is $u^{n'}$.

Note 5.51. The definition of transition probability corresponding to a source-code s, q_s, defined in Definition 5.13 is exactly the same as that of a channel defined above.

5.5.2 Channels which communicate sources to within various particular distortion levels

In this subsection, we define channels which pseudo-directly communicate i.i.d. X source to within a distortion level D and channels which pseudo-directly communicate uniform X source to within a distortion level D.

Definition 5.53 ($\mathcal{C}_{X,D}$). Let X be arbitrary. Consider a channel $c =< c^n >_1^\infty$ with input space $< \mathcal{X}^n >_1^\infty$ and output space $< \mathcal{Y}^n >_1^\infty$. Let the input to this channel be the i.i.d. X source. That is, when the block-length is n, the input to c^n is X^n. The channel produces an output Y^n. This leads to a joint random variable X^nY^n on the input-output space. The channel c is said to pseudo-directly communicate i.i.d. X source to within a distortion level D if

$$\lim_{n\to\infty} p_{X^nY^n}\left(\frac{1}{n}d^n(X^n,Y^n) > D\right) = 0 \tag{5.143}$$

The set of all channels which pseudo-directly communicate i.i.d. X source to within a distortion level D is denoted by $\mathcal{C}_{X,D}$.

Note 5.52. Let c be a channel with input space $< \mathcal{X}^n >_1^\infty$ and output space $< \mathcal{Y}^n >_1^\infty$. Let s be a source-code such that $q_s = c$. Then, $c \in \mathcal{C}_{X,D}$ if and only if probability of excess distortion D is achievable with source-code s when encoding the i.i.d. X source.

Definition 5.54 ($\mathscr{C}_{X,D,k}$). Let X be arbitrary. Consider a channel $c =< c^{kn} >_1^\infty$ with input space $< \mathscr{X}^{kn} >_1^\infty$ and output space $< \mathscr{Y}^{kn} >_1^\infty$. Let the input to this channel be the i.i.d. X source. That is, when the block-length is n, the input to c^{kn} is X^{kn}. The channel produces an output Y^{kn}. This leads to a joint random variable $X^{kn}Y^{kn}$ on the input-output space. The jump channel c with is said to pseudo-directly communicate i.i.d. X source to within a distortion level D if

$$\lim_{n\to\infty} p_{X^{kn}Y^{kn}}\left(\frac{1}{kn}d^{kn}(X^{kn},Y^{kn}) > D\right) = 0 \tag{5.144}$$

The set of all jump channels *with jump k*which pseudo-directly communicate i.i.d. X source to within a distortion level D is denoted by $\mathscr{C}_{X,D,k}$.

Note 5.53. Let c be a channel with input space $< \mathscr{X}^{kn} >_1^\infty$ and output space $< \mathscr{Y}^{kn} >_1^\infty$. Let s be a source-code such that $q_s = c$. Then, $c \in \mathscr{C}_{X,D,k}$ if and only if probability of excess distortion D is achievable with jump source-code s when encoding the i.i.d. X source.

Definition 5.55 ($\mathscr{C}_{U,D}$). Let X be such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Consider a channel $c =< c^{n'} >_1^\infty$ with input space $< \mathscr{U}^{n'} >_1^\infty$ and output space $< \mathscr{Y}^{n'} >_1^\infty$. Let the input to this channel be the uniform X source. That is, when the block-length is n', the input to $c^{n'}$ is $X^{n'}$. The channel produces an output $Y^{n'}$. This leads to a joint random variable $U^{n'}Y^{n'}$ on the input-output space. The channel c is said to pseudo-directly communicate uniform X source to within a distortion level D if

$$\lim_{n\to\infty} p_{U^{n'}Y^{n'}}\left(\frac{1}{n'}d^{n'}(U^{n'},Y^{n'}) > D\right) = 0 \tag{5.145}$$

The set of all channels which pseudo-directly communicate uniform X source to within a distortion level D is denoted by $\mathscr{C}_{U,D}$.

Note 5.54. Let c be a channel with input space $< \mathscr{U}^{n'} >_1^\infty$ and output space $< \mathscr{Y}^{n'} >_1^\infty$. Let s be a source-code such that $q_s = c$. Then, $c \in \mathscr{C}_{U,D}$ if and only if probability of excess distortion D is achievable with source-code s when encoding the uniform X source.

Note 5.55 (Channel sets and partially known channels). *In Chapter 2, we considered partially known channels $k \in \mathscr{A}$. In this chapter, we will refer to partially known channels simply as channel sets.*

5.5.3 Pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$, $\mathscr{C}_{X,D,j}$, and $\mathscr{C}_{U,D}$

In Chapter 2, we defined the universal capacity of a partially known channel $k \in \mathscr{A}$. In this chapter, instead of the universal capacity, we will consider the pseudo-universal capacity of a partially known channel, or equivalently, a set of channels. The pseudo-universal capacity of a set of channels differs from universal capacity in that we do not ask for a uniformity in the rate at which error probability $\to 0$ as block-length $n \to \infty$ over the set of channels. This will become clearer in the rigorous definition.

We will prove universal source channel separation theorem for rate-distortion by using pseudo-universal capacity instead of universal capacity, in this chapter. We are quite sure that the results can be generalized to the case when we use the universal capacity. The reason we have not proved the results using universal capacity is that this is the way we ended up proving the results. For the reader unsatisfied with this explanation, the reader can think of this chapter as an operational view-point of the optimality of digital communication for communication over a *fully* known channel. This is because pseudo-universal capacity and universal capacity are same if the channel is fully known: the uniformity over the set of channels becomes trivial.

We define the pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$, $\mathscr{C}_{X,D,j}$, and $\mathscr{C}_{U,D}$.

Pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$

Recall the definitions of $\mathscr{E}^n_{\mathscr{X}}(R)$ and $\mathscr{F}^n_{\mathscr{Y}}(R)$. In a similar vein, let $\mathscr{F}^n_{\mathscr{X}}(R)$ denote the set of all deterministic functions with domain $\{1,2,\ldots,2^{\lfloor nR \rfloor}\}$ and range $\mathscr{X}^n$ and let $\mathscr{E}^n_{\mathscr{Y}}(R)$ denote the set of all deterministic functions with domain $\mathscr{Y}^n$ and range $\{1,2,\ldots,2^{\lfloor nR \rfloor}\}$.

Definition 5.56 (Encoder-Decoder). A deterministic encoder-decoder pair is a sequence $< e^n, f^n >_1^\infty$ where $e^n \in \mathscr{F}^n_{\mathscr{X}}(R)$ and $f^n \in \mathscr{E}^n_{\mathscr{Y}}(R)$. A random encoder-decoder pair is a probability distribution on $F^n_{\mathscr{X}}(R) \times E^n_{\mathscr{Y}}(R)$, which we denote by $< E^n, F^n >_1^\infty$: $p_{E^n F^n}$ is the probability distribution on $F^n_{\mathscr{X}}(R) \times E^n_{\mathscr{Y}}(R)$.

Note that in Section 2.7.3, we had defined a random channel code consisting of a random channel encoder and a random channel decoder as transition probabilities. The above definition is an equivalent definition of a channel code as a probability distribution on the set of deterministic channel encoders and decoders.

Let $c =< c^n >_1^\infty$ be a channel with input space $< \mathscr{X}^n >_1^\infty$ and output space $< \mathscr{Y}^n >_1^\infty$. The composition of encoder, channel and decoder, which we denote by $< D^n \circ c^n \circ E^n >_1^\infty$ is a transition probability/stochastic kernel: $D^n \circ c^n \circ E^n : \{1,2,\ldots,2^{\lfloor nR \rfloor}\} \to \mathscr{P}(\{1,2,\ldots,2^{\lfloor nR \rfloor}\})$. For $a, b \in \{1,2,\ldots,2^{\lfloor nR \rfloor}\}$,

$$D^n \circ c^n \circ E^n(b|a) = \sum_{\{x^n \in \mathscr{X}^n, y^n \in \mathscr{Y}^n\}} \sum_{\{e^n, d^n : e^n(a) = x^n, d^n(y^n) = b\}} \tag{5.146}$$

$$c^n(y^n|x^n) p_{E^n F^n}(e^n, f^n) \tag{5.147}$$

Let $\mathscr{M}^n_R \triangleq \{1,2,\ldots,2^{\lfloor nR \rfloor}\}$. Let M^n_R be some distribution on $\mathscr{M}^n_R$. With input M^n_R to the composition of the encoder, channel and decoder $F^n \circ c^n \circ E^n$, there is an output distribution

$\hat{M}_R^n$ on $\{1,2,\ldots,2^{\lfloor nR \rfloor}\}$, which will depend on the channel c, though this dependence on c is not shown in the notation $\hat{M}_R^n$.

Definition 5.57 (Pseudo-universal achievability of rate R over channel set $\mathscr{C}_{X,D}$ with encoder-decoder $< E^n, F^n >_1^\infty$). Rate R is said to be pseudo-universally achievable over the channel set $\mathscr{C}_{X,D}$ by use of encoder-decoder $< E^n, F^n >_1^\infty$ if

$$\sup_{m^n \in \mathscr{M}_R^n} \Pr(\hat{M}_R^n \neq M_R^n \mid M_R^n = m^n) \to 0 \text{ as } n \to \infty \forall c \in \mathscr{C}_{X,D} \tag{5.148}$$

As we said in Section 2.11, the above definition is independent of the particular distribution M_R^n; it only depends on the channel transition probability.

Note 5.56. The reader is urged to compare the above definition of pseudo-universal achievability of rate R with Definition 2.31 of the universal achievability of rate R, and note the absence of uniformity requirement over the particular channel at the rate at which error probability $\to 0$ as block-length $n \to \infty$ in the definition of pseudo-universal achievability of rate R as opposed to the definition of universal achievability of rate R

Definition 5.58 (Pseudo-universal achievability of rate R over channel set $\mathscr{C}_{X,D}$). Rate R is said to be pseudo-universally achievable over the channel set $\mathscr{C}_{X,D}$ if rate R is pseudo-universally achievable over $\mathscr{C}_{X,D}$ with some encoder-decoder pair $< E^n, F^n >_1^\infty$.

Definition 5.59 (Pseudo-universal capacity of $\mathscr{C}_{X,D}$, $pC_{rc}(\mathscr{C}_{X,D})$). The supremum of all pseudo-universally achievable rates over $\mathscr{C}_{X,D}$ is the pseudo-universal capacity of $\mathscr{C}_{X,D}$, and is denoted by $pC_{rc}(\mathscr{C}_{X,D})$.

Pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D,k}$

Encoders and decoders are defined as in Definition 5.56, except that they are defined only for block-lengths kn.

The composition of encoder, channel and decoder is $< F^{kn} \circ c^{kn} \circ E^{kn} >_1^\infty$.

Definition 5.60 (Pseudo-universal achievability of rate R over channel set $\mathscr{C}_{X,D,k}$ with encoder-decoder $< E^{kn}, F^{kn} >_1^\infty$). This is defined analogously to Definition 5.57, expect that limits are taken along block-lengths kn. Rate R is said to be pseudo-universally achievable over the channel set $\mathscr{C}_{X,D,k}$ by use of encoder-decoder $< E^{kn}, F^{kn} >_1^\infty$ if

$$\sup_{m^{kn} \in \mathscr{M}_R^{kn}} \Pr(\hat{M}_R^{kn} \neq M_R^{kn} \mid M_R^{kn} = m^{kn}) \to 0 \text{ as } n \to \infty \forall c \in \mathscr{C}_{X,D,k} \tag{5.149}$$

Definition 5.61 (Pseudo-universal achievability of rate R over channel set $\mathscr{C}_{X,D,k}$). This is defined analogously to Definition 5.5.3, expect that limits are taken along block-lengths kn. Rate R is said to be pseudo-universally achievable over the channel set $\mathscr{C}_{X,D,k}$ if rate R is pseudo-universally achievable over $\mathscr{C}_{X,D}$ with some encoder-decoder pair $< E^n, F^n >_1^\infty$.

Definition 5.62 (Pseudo-universal capacity of $\mathscr{C}_{X,D,k}$, $pC_{rc}(\mathscr{C}_{X,D,k})$). This is defined analogously to Definition 5.59, expect that limits are taken along block-lengths kn. The supremum of all pseudo-universally achievable rates over $\mathscr{C}_{X,D,k}$ is the pseudo-universal capacity of $\mathscr{C}_{X,D,k}$, and is denoted by $pC_{rc}(\mathscr{C}_{X,D,k})$.

Pseudo-universal capacity of the set of channels $\mathscr{C}_{U,D}$

Recall the definitions of $\mathscr{E}^{n'}_{\mathscr{U}}(R)$ and $\mathscr{F}^{n'}_{\mathscr{U}}(R)$. A deterministic encoder-decoder pair is a sequence $< e^{n'}, f^{n'} >_1^\infty$ where $e^{n'} \in \mathscr{F}^{n'}_{\mathscr{U}}(R)$ and $f^{n'} \in \mathscr{E}^{n'}_{\mathscr{U}}(R)$. A random encoder-decoder pair is a probability distribution on $F^{n'}_{\mathscr{X}}(R) \times E^{n'}_{\mathscr{X}}(R)$, which we denote by $< E^{n'}, F^{n'} >_1^\infty$: $p_{E^{n'}F^{n'}}$ is the probability distribution on $F^{n'}_{\mathscr{X}}(R) \times E^{n'}_{\mathscr{X}}(R)$.

Let $c =< c^{n'} >_1^\infty$ be a channel with input space $< \mathscr{X}^{n'} >_1^\infty$ and output space $< \mathscr{Y}^{n'} >_1^\infty$. The composition of encoder, channel and decoder, which we denote by $< D^{n'} \circ c^{n'} \circ E^{n'} >_1^\infty$ is a transition probability/stochastic kernel: $D^{n'} \circ c^{n'} \circ E^{n'} : \{1,2,\ldots,2^{\lfloor n'R \rfloor}\} \to \mathscr{P}(\{1,2,\ldots,2^{\lfloor n'R \rfloor}\})$. For $a, b \in \{1,2,\ldots,2^{\lfloor n'R \rfloor}\}$,

$$D^{n'} \circ c^{n'} \circ E^{n'}(b|a) = \sum_{\{x^{n'} \in \mathscr{X}^{n'}, y^{n'} \in \mathscr{Y}^{n'}\}} \sum_{\{e^{n'}, d^{n'} : e^{n'}(a) = x^{n'}, d^{n'}(y^{n'}) = b\}} c^{n'}(y^{n'}|x^{n'}) p_{E^{n'}F^{n'}}(e^{n'}, f^{n'}) \tag{5.150}$$

As before, $\mathscr{M}^{n'}_R$ denotes the set $\{1,2,\ldots,2^{\lfloor n'R \rfloor}\}$. Let $M^{n'}_R$ denote some distribution on $\mathscr{M}^{n'}_R$. With input $M^{n'}_R$ to the composition of the encoder, channel and decoder $F^{n'} \circ c^{n'} \circ E^{n'}$, there is an output distribution $\hat{M}^{n'}_R$ on $\{1,2,\ldots,2^{\lfloor n'R \rfloor}\}$, which will depend on the particular channel, though this dependence on c is not shown in the notation $\hat{M}^{n'}_R$.

Definition 5.63 (Pseudo-universal achievability of rate R over channel set $\mathscr{C}_{U,D}$ with encoder-decoder $< E^{n'}, F^{n'} >_1^\infty$). Rate R is said to be pseudo-universally achievable over the channel set $\mathscr{C}_{U,D}$ by use of encoder-decoder $< E^{n'}, F^{n'} >_1^\infty$ if

$$\sup_{m^{n'} \in \mathscr{M}^{n'}_R} \Pr(\hat{M}^{n'}_R \neq M^{n'}_R \mid M^{n'}_R = m^{n'}) \to 0 \text{ as } n \to \infty \forall c \in \mathscr{C}_{U,D} \tag{5.151}$$

Definition 5.64 (Pseudo-universal achievability of rate R over channel set $\mathscr{C}_{U,D}$). Rate R is said to be pseudo-universally achievable over the channel set $\mathscr{C}_{U,D}$ if rate R is pseudo-universally achievable over $\mathscr{C}_{U,D}$ with some encoder-decoder pair $< E^{n'}, F^{n'} >_1^\infty$.

Definition 5.65 (Pseudo-universal capacity of $\mathscr{C}_{U,D}$, $pC_{rc}(\mathscr{C}_{U,D})$). The supremum of all pseudo-universally achievable rates over $\mathscr{C}_{U,D}$ is the pseudo-universal capacity of $\mathscr{C}_{U,D}$, and is denoted by $pC_{rc}(\mathscr{C}_{U,D})$.

Note 5.57. We will *not* require the notion of jump pseudo-universal capacity for the set of channels $C_{U,D}$.

■ 5.5.4 Relation between the pseudo-universal capacities of the set of channels $\mathscr{C}_{U,D}$, $\mathscr{C}_{X,D}$, and $\mathscr{C}_{X,D,k}$

In this subsection, we prove relations between pseudo-universal capacities of the sets of channels $\mathscr{C}_{U,D}$, $\mathscr{C}_{X,D}$, and $\mathscr{C}_{X,D,k}$.

The results, not stated entirely precisely here, but stated more precisely in the lemmas that follow, are:

1. $pC_{rc}(\mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)}) \leq pC_{rc}(\mathscr{C}_{U,D})$. This is proved by interpreting channels as source-codes and using Construction 5.1 to construct a channel corresponding to every channel $\in \mathscr{C}_{U,D}$.
2. $pC_{rc}(\mathscr{C}_{U,D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)}) \leq pC_{rc}(\mathscr{C}_{X',D})$. This is proved by interpreting channels as source-codes and using Construction 5.2 to construct a channel corresponding to every channel $\in \mathscr{C}_{X,D,n_0}$.

In order to prove the above two results, we first prove that $pC_{rc}(\mathscr{C}_{X,D+\epsilon,k}) \leq pC_{rc}(\mathscr{C}_{X,D})$ $\leq pC_{rc}(\mathscr{C}_{X,D,k})$. This is proved by defining interpolation encoder-decoder $< E'^n, F'^n >_1^\infty$ for the set of channels $\mathscr{C}_{X,D}$ given an encoder-decoder $< E^{kn}, F^{kn} >_1^\infty$ for the set of channels $\mathscr{C}_{X,D,k}$.

Lemma 5.43. $\forall k, \forall \epsilon > 0,\ pC_{rc}(\mathscr{C}_{X,D+\epsilon,k}) \leq pC_{rc}(\mathscr{C}_{X,D}) \leq pC_{rc}(\mathscr{C}_{X,D,k})$

Proof. If rate R is pseudo-universally achievable for the set of channels $\mathscr{C}_{X,D}$ by using an encoder-decoder $< E^n, F^n >_1^\infty$, then rate R is pseudo-universally achievable for the set of channels $\mathscr{C}_{X,D,k}$ by using the subsequence $< E^{kn}, F^{kn} >_1^\infty$ as the encoder-decoder. Thus, $pC_{rc}(\mathscr{C}_{X,D}) \leq pC_{rc}(\mathscr{C}_{X,D,k})$.

Let rate R be pseudo-universally achievable for the set of channels $\mathscr{C}_{X,D+\epsilon,k}$. That is, there exists encoder-decoder $< E^{kn}, F^{kn} >_1^\infty$ such that

$$\Pr(\hat{M}_c^{kn} \neq M^{kn}) \to 0 \text{ as } n \to \infty \forall c \in \mathscr{C}_{X,D+\epsilon,k} \tag{5.152}$$

We will modify encoder-decoder $< E^{kn}, F^{kn} >_1^\infty$ into encoder-decoder $< E'^n, F'^n >_1^\infty$ for the set of channels $\mathscr{C}_{X,D}$. This code will be such that rates $< R$ are pseudo-universally achievable for the set of channels $\mathscr{C}_{X,D}$.

The idea is the following: Let $r \in \{0, 1, 2, \ldots, k-1\}$. For block-length $kn + r$, the channel is c^{kn+r}. Use only the first block of length kn to communicate. Precisely, this is done as follows:

Let $(e^{kn}, f^{kn}) \in \mathscr{F}_{\mathscr{X}}^{kn}(R) \times \mathscr{E}_{\mathscr{Y}}^{kn}(R)$. Let $r \in \{0,2,\ldots,k-1\}$. Let $m \in \{1,2,\ldots,2^{\lfloor knR \rfloor}\}$. Define

$$e'^{kn+r}(m) = (e^{kn}(m), x^r) \tag{5.153}$$

where $x^r \in \mathscr{X}^r$ is arbitrary. For $y^{kn+r} \in \mathscr{Y}^{kn+r}$, define,

$$f'^{kn+r}(y^{kn+r}) = f^{kn}(y^{kn}) \tag{5.154}$$

The domain of the function e'^{kn+r} is the set $\{1,2,\ldots,2^{\lfloor knR \rfloor}\}$ and the range of e'^{kn+r} is the set $\mathscr{X}^{kn+r}$. The domain of the function f'^{kn+r} is $\mathscr{Y}^{kn+r}$ and the range of the function f'^{kn+r} is the set $\{1,2,\ldots,2^{\lfloor knR \rfloor}\}$.

If $(E^{kn}, F^{kn}) \in \mathscr{P}(\mathscr{F}_{\mathscr{X}}^{kn}(R) \times \mathscr{E}_{\mathscr{Y}}^{kn}(R))$, for $r \in \{0,1,2,\ldots,k-1\}$, for sufficiently large n, define $(E'^{kn+r}, F'^{kn+r}) \in \mathscr{P}(\mathscr{F}_{\mathscr{X}}^{kn+r}(R) \times \mathscr{E}_{\mathscr{Y}}^{kn+r}(R-\beta))$, defined by

$$p_{E'^{kn+r}, F'^{kn+r}}(e'^{kn+r}, f'^{kn+r}) = p_{E^{kn}F^{kn}}(e^{kn}, f^{kn}) \tag{5.155}$$

where (e'^{kn+r}, f'^{kn+r}) is constructed from (e^{kn}, f^{kn}) as above.

Given $< E^{kn}, F^{kn} >_1^{\infty}$, we construct a code $< E'^n, F'^n >_1^{\infty}$ as above. $< E'^n, F'^n >_1^{\infty}$ is the interpolation channel-code corresponding to $< E^{kn}, F^{kn} >_1^{\infty}$.

We prove that by using $< E'^n, F'^n >_1^{\infty}$, rate $R-\beta$ is pseudo-universally achievable for set of channels $\mathscr{C}_{X,D}$.

Let $c =< c^n >_1^{\infty} \in \mathscr{C}_{X,D}$. For block-length $kn+r$, $r \in \{0,1,2,\ldots,r-1\}$,

$$\begin{aligned}
&\frac{1}{kn+r} d^{kn+r}(X^{kn+r}, Y^{kn+r}) = \\
&\qquad \frac{1}{kn+r} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) + \\
&\qquad\qquad \frac{1}{kn+r} d^{r}(X^{kn+r}(kn+1..kn+r), Y^{kn+r}(kn+1..kn+r)) \\
&\leq \frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) + \frac{rD_{\max}}{kn+r} \\
&\leq \frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) + \epsilon \text{ for sufficiently large } n
\end{aligned} \tag{5.156}$$

It follows that for channels $\in \mathscr{C}_{X,D}$, for $r \in \{0,1,2,\ldots,k-1\}$,

$$\Pr\left(\frac{1}{kn} d^{kn}(X^{kn+r}(1..kn), Y^{kn+r}(1..kn)) > D + \epsilon\right) \to 0 \text{ as } n \to \infty \tag{5.157}$$

Let block-length be $kn+r$. Let the message set be $\{1,2,\ldots,2^{\lfloor knR \rfloor}\}$. The channel is c^{kn+r}, where $c \in \mathscr{C}_{X,D}$. Let the encoder-decoder be E'^{kn+r}, F'^{kn+r}. It follows by construction that error probability $\to 0$ as block length $\to \infty$.

Note that the cardinality of the message set is $2^{\lfloor knR \rfloor}$. For $\beta > 0$, for sufficiently large n, $2^{\lfloor (kn+r)(R-\beta) \rfloor} \leq 2^{\lfloor knR \rfloor}$. It follows that rates $< (R - \beta)$ are pseudo-universally achievable achievable over the set of channels $\mathscr{C}_{X,D}$. It follows that $pC_{rc}(\mathscr{C}_{X,D+\epsilon}, k) \leq pC_{rc}(\mathscr{C}_{X,D})$.

The lemma follows.

□

Lemma 5.44. *Let X be such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Let X' be arbitrary. Then, $\forall \epsilon > 0$,*

$$pC_{rc}(\mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)}) \leq pC_{rc}(\mathscr{C}_{U,D}) \tag{5.158}$$

Proof. Recall the notation $n' = n_0 n$. Let $c = < c^{n'} >_1^\infty \in \mathscr{C}_{U,D}$. Consider a source-code s such that $q_s = c$. Then, probability of excess distortion D is achievable by the source-code s when encoding the uniform X source. Let $\epsilon > 0$. From Lemma 5.26, by use of Construction 5.1, it follows that there exists a jump source-code s' for which probability of excess distortion $D + D_{\max}(d_1(p_X, p_{X'}) + \epsilon)$ is achievable when encoding the i.i.d. X source. $q_{s'}$ can be thought of as a jump channel c' with jump n_0. Then, $c' \in \mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon),n_0}$.

Denote $\mathscr{C}' = \{c' \mid c \in \mathscr{C}_{U,D}\}$, where c' is constructed, as above, from c.

Note that $c' = q_{s'} = < k'^{n'} \circ q_s^{n'} >_1^\infty = < k'^{n'} \circ c^{n'} >_1^\infty$. Thus, if rate R is pseudo-universally achievable for the set of channels $\mathscr{C}'$ by using a possibly random encoder-decoder sequence $< g^{n'}, h^{n'} >_1^\infty$, rate R is also pseudo-universally achievable for the set of channels $\mathscr{C}_{U,D}$ by using the encoder-decoder sequence $< g^{n'} \circ k'^{n'}, h^{n'} >_1^\infty$. It follows that $pC_{rc}(\mathscr{C}') \leq pC_{rc}(\mathscr{C}_{U,D})$.

Also, $\mathscr{C}' \subset \mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon),n_0}$.

Thus, $pC_{rc}(\mathscr{C}') \geq pC_{rc}(\mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon),n_0})$.

It follows that $pC_{rc}(\mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon),n_0}) \leq pC_{rc}(\mathscr{C}_{U,D})$.

By Lemma 5.43, $pC_{rc}(\mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)}) \leq pC_{rc}(\mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon),n_0})$.

Thus, $pC_{rc}(\mathscr{C}_{X',D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)}) \leq pC_{rc}(\mathscr{C}_{U,D})$ □

Lemma 5.45. *Let X be such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Let X' be arbitrary. Then, $\forall \epsilon > 0$,*

$$pC_{rc}(\mathscr{C}_{U,D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)}) \leq pC_{rc}(\mathscr{C}_{X',D}) \tag{5.159}$$

Proof. Recall the notation $n' = n_0 n$. Let $c = < c^{n'} >_1^\infty \in \mathscr{C}_{X',D,n_0}$. Consider a jump source-code t such that $q_t = c$. Then, probability of excess distortion D is achievable by the jump source-code t when encoding the i.i.d. X' source. Let $\epsilon_1 > 0$. From Lemma 5.27, by use of Construction 5.2, it follows that there exists source-code t' for which probability of excess distortion $D + D_{\max}(d_1(p_X, p_{X'}) + \epsilon_1)$ is achievable when encoding the uniform X source. $q_{t'}$ can be thought of as a channel c'. Then, $c' \in \mathscr{C}_{U,D+D_{\max}(d_1(p_X,p_{X'})+\epsilon_1)}$.

Denote $\mathscr{C}' = \{c' \mid c \in \mathscr{C}_{X',D}\}$, where c' is constructed, as above, from c.

Note that $c' = q_{t'} =< l'^{n'} \circ q_t^{n'} >_1^\infty =< l'^{n'} \circ c^{n'} >_1^\infty$. Thus, if rate R is pseudo-universally achievable for the set of channels $\mathscr{C}'$ by using a possibly random encoder-decoder sequence $< g^{n'}, h^{n'} >_1^\infty$, rate R is also pseudo-universally achievable for the set of channels $\mathscr{C}_{X',D,n_0}$ by using the encoder-decoder sequence $< g^{n'} \circ k'^{n'}, h^{n'} >_1^\infty$. It follows that $pC_{rc}(\mathscr{C}') \leq pC_{rc}(\mathscr{C}_{X',D,n_0})$. By Lemma 5.43, $\forall \epsilon_2 > 0$, $pC_{rc}(\mathscr{C}_{X',D,n_0}) \leq pC_{rc}(\mathscr{C}_{X',D-\epsilon_2})$. Thus, $pC_{rc}(\mathscr{C}') \leq pC_{rc}(\mathscr{C}_{X',D-\epsilon_2})$.

Also, $\mathscr{C}' \subset \mathscr{C}_{U,D+D_{\max}(d_1(p_X,p_{X'})+\epsilon_1)}$. Thus, $pC_{rc}(\mathscr{C}') \geq pC_{rc}(\mathscr{C}_{U,D+D_{\max}(d_1(p_X,p_{X'})+\epsilon_1)})$.

Thus, $\forall \epsilon_1 > 0, \epsilon_2 > 0$, $pC_{rc}(\mathscr{C}_{X',D-\epsilon_2}) \geq pC_{rc}(\mathscr{C}_{U,D+D_{\max}(d_1(p_X,p_{X'})+\epsilon_1)})$. It follows that $\forall \epsilon > 0$, $pC_{rc}(\mathscr{C}_{X',D}) \geq pC_{rc}(\mathscr{C}_{U,D+D_{\max}(d_1(p_X,p_{X'})+\epsilon)})$. □

Note 5.58. *The reason for the introduction of jump channels which communicate i.i.d. X' source is the following. Given a channel which communicates the uniform X source, we can only draw a relation to it for jump channels with jump n_0. This is because channels which communicate uniform X source are defined only for block-lengths which are multiples of n_0. Then, by use of Lemma 5.43, we relate the capacities of channels and jump channels which communicate i.i.d. X source.*

This ends this section. To re-capitulate, we defined the set of channels and jump channels which pseudo-directly communicate the i.i.d. X source to within a distortion D and the set of channels which pseudo-directly communicate the uniform X source to within a distortion D. We proved various relations between the pseudo-universal capacities of these sets of channels. We use these results along with the results from the previous section in the next section to prove the main result: the equality of the pseudo-universal channel capacity of the set of channels $\mathscr{C}_{X,D}$ and the rate-distortion function for the i.i.d. X source under the expected and the probability of excess distortion criteria. As we shall see, we will crucially use the relations between the rate-distortion functions for the i.i.d. and uniform X sources, and also, the pseudo-universal channel capacities of the set of channels which pseudo-directly communicate the i.i.d. X source to within a distortion level D and the set of channels which pseudo-directly communicate the uniform X source to within a distortion level D.

■ 5.6 Relation between pseudo-universal channel capacity and the rate-distortion function: equality of the pseudo universal channel capacity and the rate distortion function

In this section, we prove: $pC_{rc}(\mathscr{C}_{X,D}) = R_X^E(D) = R_X^P(D) = R_X^E(D, \inf) = R_X^P(D, \inf)$. To this end, we first prove that $R_U^P(D) \geq pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^P(D, \inf)$. This proof is independent of all past results in this chapter. Then, we integrate this with the results from the previous two sections relating rate-distortion functions for the i.i.d. X and uniform X sources and the results relating the pseudo-universal channel capacities of the set of channels which pseudo-directly communicate the i.i.d. X source to within a distortion D and the set of channels which pseudo-directly communicate the uniform X source to within a distortion D, in order to prove the desired result.

5.6.1 Proof of $R_U^p(D) \geq pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^p(D,\inf)$

Theorem 5.46. *Let X be such that $p_X(x)$ is rational $\forall x \in \mathscr{X}$. Then, for $D > 0$,*

$$R_U^p(D) \geq pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^p(D,\inf) \tag{5.160}$$

Proof. This proof is independent of all past lemmas and theorems in this chapter, and the proof of $pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^p(D,\inf)$ is essentially the same as Step 1 of Subsection 2.14.9.

This is followed by a proof of $R_U^p(D) \geq pC_{rc}(\mathscr{C}_{U,D})$.

This theorem and its proof is the main idea of this whole chapter; the rest are just mathematical details which are necessary and have some ideas but are not the main idea.

First, we prove that $pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^p(D,\inf)$. Most of the argument of Step 1 of Subsection 2.14.9 is reproduced here with the minor necessary changes.

This is done via parallel random-coding arguments for

- the pseudo-universal capacity of the set of channels $\mathscr{C}_{U,D}$, and
- the rate-distortion source-coding problem of finding the minimum rate needed to compress the uniform X source to within a distortion D under the inf probability of excess distortion criterion.

The random coding arguments are similar, yet different from the ones used in the information theory literature. We want to derive a connection between the above two problems in order to prove the desired result, and we are not interested in simplified functional expressions for the pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$ or simplified expressions for the rate-distortion function $R_U^p(D,\inf)$.

The two problems:

- *The channel-coding problem:* The channel-coding problem is that of computing the pseudo universal capacity of the set of channels $\mathscr{C}_{U,D}$
- *The source-coding problem:* The source-coding problem that we consider is to derive an upper bound on $R_U^p(D,\inf)$, the minimum rate needed to compress the uniform X source to within a distortion level D under the inf-probability of excess distortion criterion

Block length: For both the channel coding and the source coding problems, let the block-length be n'. Towards the end of the argument we will take the limit $n' \to \infty$. Recall that $n' = n_0 n$ is the set of all integers for which the uniform X source makes sense.

Codebook generation:

- *Codebook for the channel-coding problem:* Let communication be desired at rate R. Generate $2^{\lfloor n'R \rfloor}$ sequences independently and uniformly from the set $\mathcal{U}^{n'}$, the set of all sequences $\in \mathcal{X}^n$ which have empirical distribution *precisely* p_X.

  This is the code book $\mathcal{K}^{n'}$. Note that the codewords $\in \mathcal{U}^{n'}$. The encoder is denoted by $< E^{n'} >_1^{\infty}$. Note that the encoder is random.

- *Codebook for the source-coding problem:* Let q be an empirical distribution (type) on $\mathcal{Y}$, that is $q \in \mathcal{P}(\mathcal{Y})$. Let q be an achievable type when the block-length is n'. In other words, $n'q(y)$ is an integer $\forall y \in \mathcal{Y}$. Let $\mathcal{U}_q^{n'} \subset \mathcal{Y}^{n'}$ denote the set of all sequences with empirical distribution, precisely q. Generate $2^{\lfloor n'R \rfloor}$ codewords independently and uniformly from the set $\mathcal{U}_q^{n'}$.

  This is the code book $\mathcal{L}^{n'}$. Note that the codewords $\in \mathcal{U}_q^{n'} \subset \mathcal{Y}^{n'}$. Note that the codebook is random.

Joint typicality:

- *Joint typicality for the channel coding problem:* Sequences $(u^{n'}, y^{n'}) \in$ the channel input-output space $\mathcal{U}^{n'} \times \mathcal{Y}^{n'}$ are said to be jointly typical if

$$\frac{1}{n'} d^{n'}(u^{n'}, y^{n'}) \leq D \tag{5.161}$$

- *Joint typicality for the source coding problem:* Sequences $(u^{n'}, y^{n'}) \in$ the source input - source reconstruction space $\mathcal{U}^{n'} \times \mathcal{Y}^{n'}$ are said to be jointly typical if

$$\frac{1}{n'} d^{n'}(u^{n'}, y^{n'}) \leq D \tag{5.162}$$

Note that the definition of joint-typicality for both the channel-coding and the source-coding problems is the same.

Decoding:

- *Decoding for the channel coding problem:* Let the sequence $y^{n'}$ be received. If there exists *unique* codeword $u^{n'}$ in the code book $\mathcal{K}^{n'}$ for which $(u^{n'}, y^{n'})$ are jointly typical, declare that $u^{n'}$ is transmitted, else declare error. The decoder is denoted by $F^{n'}$. Note that the encoder-decoder $E^{n'}, F^{n'}$ is random

Note 5.59. This decoding rule can be thought of as a variant of minimum distance decoding

- *Encoding for the source coding problem:* Let the sequence $u^{n'} \in \mathcal{U}^{n'}$ needs to be source-coded. If there exists some sequence $y^{n'}$ in the code book $\mathcal{L}^{n'}$ for which $(u^{n'}, y^{n'})$ are jointly typical, encode $u^{n'}$ to one such $y^{n'}$, else declare error. Note that the encoder-decoder is random

Note 5.60. Note that "unique" in the channel coding problem gets converted to "some" in the source coding problem

Some notation:

- *Notation for the channel coding problem:* We will do the analysis assuming that a particular message is transmitted. The message set is

$$\mathcal{M}_R^{n'} = \{1, 2, \ldots, 2^{\lfloor n'R \rfloor}\} \tag{5.163}$$

Assume that message $m_i^{n'} \in \mathcal{M}_R^{n'}$ is transmitted.

Let the codeword corresponding to message $m_i^{n'}$ be denoted by $u_c^{n'}$. Let the non-transmitted codewords be denoted by $u'^{n'}_1, u'^{n'}_2, u'^{n'}_{2^{\lfloor n'R \rfloor}-1}$.

$u_c^{n'}$ is a realization of $U_c^{n'}$. By the random code book generation, $U_c^{n'}$ has uniform distribution on $\mathcal{U}^{n'}$.

$u'^{n'}_i$ is a realization of $U'^{n'}_i$, $1 \leq i \leq 2^{\lfloor n'R \rfloor} - 1$. By the random code book generation, $U'^{n'}_i$, $1 \leq i \leq 2^{\lfloor n'R \rfloor} - 1$, has uniform distribution on $\mathcal{U}^{n'}$.

By the random code book generation, the codewords are generated independently of each other, and thus, $U_c^{n'}$, $U'^{n'}_i$, $1 \leq i \leq 2^{\lfloor n'R \rfloor} - 1$ are all independent of each other as random variables.

The action of the channel $c \in \mathcal{C}_{U,D}$ on the transmitted codeword $u_c^{n'}$ produces an output $y^{n'}$.

$y^{n'}$ is the realization of some random variable $Y^{n'}$ which is got by the action of the channel c on $U_c^{n'}$. Note that $Y^{n'}$ will be different for different $c \in \mathcal{C}_{U,D}$. Assume that some particular $c \in \mathcal{C}_{U,D}$ happens, and $Y^{n'}$ is the corresponding channel output random variable. Our argument will hold for all $c \in \mathcal{C}_{U,D}$.

$y^{n'}$ depends on $u_c^{n'}$.

By the codebook generation, the codewords are generated independently of each other, and there is no dependence between $y^{n'}$ and $u'^{n'}_1, u'^{n'}_2, u'^{n'}_{2^{\lfloor n'R \rfloor}-1}$. That is, $y^{n'}$, and $Y^{n'}$ are independent of $U'^{n'}_i$, $1 \leq i \leq 2^{\lfloor n'R \rfloor} - 1$.

- *Notation for the source coding problem:* We will do the analysis assuming that a particular $u^{n'} \in \mathcal{U}^{n'}$ needs to be coded.

  The source is the uniform X source. Thus, $u^{n'}$ is a realization of $U^{n'}$ where $U^{n'}$ has uniform distribution on $\mathcal{U}^{n'}$.

  The codebook is

$$\mathcal{L}^{n'} = \{y_1^{n'}, y_2^{n'}, \ldots, y_{2^{\lfloor n'R \rfloor}}^{n'}\} \tag{5.164}$$

  For all i, $y_i^{n'}$ is a realization of the random variable $V_i^{n'}$. By the random codebook generation, $V_i^{n'}$ is the uniform distribution on the set $\mathcal{U}_q^{n'} \subset \mathcal{Y}^{n'}$ of all sequences with precise type q.

  By the random code book generation, the codewords are generated independently of each other, and thus, $V_i^{n'}$, $1 \leq i \leq 2^{\lfloor n'R \rfloor}$ are independent of each other as random variables.

  Also, the codewords are of course, independent of the source sequence, and thus, $u^{n'}$ and $U^{n'}$ are independent of $V_i^{n'}$, $1 \leq i \leq 2^{\lfloor n'R \rfloor}$.

Analysis:

- *Error analysis for the channel coding problem:* We analyze the probability of *correct* decoding.

  We analyze the probability that a message is correctly received given that a particular message is transmitted. Think of some probability distribution $M^{n'}$ on the message set $\mathcal{M}_R^{n'}$. This probability distribution will *not* matter for the calculation. In fact, the calculation that we do can be done even if there is no probability distribution on the set of messages. We calculate

$$\Pr(\hat{M}_R^{n'} = M_R^{n'} | M^{n'} = m_i^{n'}) \text{ where } m_i^{n'} \in \mathcal{M}_R^{n'} \tag{5.165}$$

  The code book generation is symmetric. For this reason, the above probability will be independent of the particular message $m^{n'} \in \mathcal{M}_R^{n'}$.

  From the decoding rule, it follows that for correct decoding, the following should happen:

  -

$$\frac{1}{n'} d^{n'}(u_c^{n'}, y^{n'}) \leq D \tag{5.166}$$

  -

$$\frac{1}{n'} d^{n'}(u'^{n'}_i, y^{n'}) > D, 1 \leq i \leq 2^{\lfloor n'R \rfloor} - 1 \tag{5.167}$$

Thus, the event of correct decoding is:

$$\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'})\leq D\right\}\cap\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U_i'^{n'},Y^{n'})>D\right\} \tag{5.168}$$

- *Error analysis for the source coding problem:* We analyze the probability of *error*.

  The analysis is done assuming that a particular sequence $u^{n'}\in\mathcal{U}^{n'}$ needs to be source-coded. As we shall see, this error is independent of the particular source sequence because of the same empirical distribution of the source sequences, the symmetric nature of the code book construction, and permutation invariant distortion measure.

  An error happens if there exists no $y^{n'}$ in the code book $\mathcal{L}^{n'}$ such that $(u^{n'},y^{n'})$ are jointly typical, that is, an error happens if

$$\frac{1}{n'}d^{n'}(u^{n'},y^{n'})>D\forall y^{n'}\in\mathcal{L}^{n'} \tag{5.169}$$

  The event of error is

$$\cap_{i=1}^{2^{\lfloor n'R\rfloor}}\left\{\frac{1}{n'}d^{n'}(u^{n'},V_i^{n'})>D\right\} \tag{5.170}$$

  Note 5.61. Note that in the channel coding problem, we analyze the probability of correct decoding and in the source coding problem we analyze the probability of error

  Calculation:

- *Calculation of probability of correct decoding for the channel coding problem:*

  The correct decoding event is:

$$\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'})\leq D\right\}\cap\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U_i'^{n'},Y^{n'})>D\right\} \tag{5.171}$$

  We wish to calculate the probability of the above event.

$$\begin{aligned}
&\Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'})\leq D\right\}\cap\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U_i'^{n'},Y^{n'})>D\right\}\right)\\
&=\Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'})\leq D\right\}\right)+\Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U_i'^{n'},Y^{n'})>D\right\}\right)-\\
&\quad\Pr\left(\left\{\frac{1}{n'}d^{n'}(U_c^{n'},Y^{n'})\leq D\right\}\cup\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U_i'^{n'},Y^{n'})>D\right\}\right)\\
&\geq(1-\omega_{n'})+\Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U_i'^{n'},Y^{n'})>D\right\}\right)-1\\
&=-\omega_{n'}+\Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}-1}\left\{\frac{1}{n'}d^{n'}(U_i'^{n'},Y^{n'})>D\right\}\right)
\end{aligned}$$

$$= -\omega_{n'} + \prod_{i=1}^{2^{\lfloor n'R \rfloor}-1} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U'^{n'}_i, Y^{n'}) > D\right\}\right)$$

(since $U'^{n'}_i, 1 \le i \le 2^{\lfloor n'R \rfloor} - 1$, $Y^{n'}$ are independent random variables)

$$= -\omega_{n'} + \left[\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, Y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

(where $U^{n'}$ has the same distribution as $U'^{n'}_i$ and is independent of $Y^{n'}$)

$$= -\omega_{n'} + \left[\sum_{y^{n'} \in \mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'}) \Pr\left(\frac{1}{n'}d^{n'}(U^{n'}, Y^{n'}) > D \mid Y^{n'} = y^{n'}\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

$$= -\omega_{n'} + \left[\sum_{y^{n'} \in \mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'}) \Pr\left(\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D \mid Y^{n'} = y^{n'}\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

$$= -\omega_{n'} + \left[\sum_{y^{n'} \in \mathcal{Y}^{n'}} p_{Y^{n'}}(y^{n'}) \Pr\left(\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right)\right]^{2^{\lfloor n'R \rfloor}-1}$$

(since $U^{n'}$ and $Y^{n'}$ are independent)

$$\ge -\omega_{n'} + \left[\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1} \tag{5.172}$$

Rate R is achievable if

$$-\omega_{n'} + \left[\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1} \to 1 \text{ as } n' \to \infty \tag{5.173}$$

It is known that $\omega_{n'} \to 0$ as $n' \to \infty$. It follows that rate R is achievable if

$$\left[\inf_{y^{n'} \in \mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R \rfloor}-1} \to 1 \text{ as } n' \to \infty \tag{5.174}$$

- *Calculation of probability of error for the source coding problem:*

  The error event is:

$$\cap_{i=1}^{2^{\lfloor n'R \rfloor}} \left\{\frac{1}{n'}d^{n'}(u^{n'}, V_i^{n'}) > D\right\} \tag{5.175}$$

We wish to calculate the probability of this event.

$$\Pr\left(\cap_{i=1}^{2^{\lfloor n'R\rfloor}} \left\{\frac{1}{n'}d^{n'}(u^{n'}, V_i^{n'}) > D\right\}\right) \tag{5.176}$$

$$= \prod_{i=1}^{2^{\lfloor n'R\rfloor}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_i^{n'}) > D\right\}\right) = \left[\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}} \tag{5.177}$$

where $V^{n'}$ is a random variable which is uniformly distributed on $\mathcal{U}_q^{n'}$ and is independent of $u^{n'}$ for all $u^{n'} \in \mathcal{U}^{n'}$.

The type q with which the codewords are generated can be chosen by us. For block-length n', we can choose the best possible achievable q for which the above error probability is the minimum. Let the set of all possible achievable types q for block-length n' be denoted by $\mathcal{G}^{n'}$. The least possible error probability is given by

$$\left[\inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}} \tag{5.178}$$

To show the above dependence of the distribution of $V^{n'}$ on q, we denote it by $V_q^{n'}$. Thus, the least possible error probability is

$$\left[\inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right)\right]^{2^{\lfloor n'R\rfloor}} \tag{5.179}$$

Since we are using the inf-probability of excess distortion criterion, it follows that rate R is achievable if

$$\left[\inf_{q\in\mathcal{G}^{n'_i}} \Pr\left(\left\{\frac{1}{n'_i}d^{n'_i}(u^{n'_i}, V_q^{n'_i}) > D\right\}\right)\right]^{2^{\lfloor n'_i R\rfloor}} \to 0 \text{ for some } n'_i = n_0 n_i \text{ for some } n_i \to \infty \tag{5.180}$$

Connection between channel coding and source coding:

It turns out that the main calculation we need to do in the channel coding problem is

$$\inf_{y^{n'}\in\mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) \tag{5.181}$$

and the main calculation we need to do in the source coding problem is

$$\inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) \tag{5.182}$$

We will prove that the above two expressions are equal.

We will prove more generally, that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'})>D\right\}\right),\ \text{if } y^{n'} \text{ has type } q \quad (5.183)$$

Let $y^{n'}$ have type q.

First we prove for the channel coding problem that if $y^{n'}$ and $y'^{n'}$ have the same type q, then

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y'^{n'})>D\right\}\right) \quad (5.184)$$

Since $U^{n'}$ is the uniform distribution on $\mathcal{U}^{n'}$, it follows that it is sufficient to prove that the cardinalities of the sets

$$\left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y^{n'})>D\right\} \text{ and } \left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y'^{n'})>D\right\} \quad (5.185)$$

are equal

Since $y^{n'}$ and $y'^{n'}$ have the same type, $y'^{n'}$ is a permutation of $y^{n'}$. Let $y'^{n'}=\pi^{n'}y^{n'}$.

Denote the sets

$$\mathcal{B}_{y^{n'}} \triangleq \left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y^{n'})>D\right\} \quad (5.186)$$

and

$$\mathcal{B}_{y'^{n'}} \triangleq \left\{u^{n'}:\frac{1}{n'}d^{n'}(u^{n'},y'^{n'})>D\right\} \quad (5.187)$$

Let $u^{n'}\in\mathcal{B}_{y^{n'}}$. Since the distortion measure is permutation invariant, $d^{n'}(\pi^{n'}u^{n'},\pi^{n'}y^{n'})=d^{n'}(u^{n'},y^{n'})$. Thus, $\pi^{n'}u^{n'}\in\mathcal{B}_{y'^{n'}}$. If $u^{n'}\neq u'^{n'}$, $\pi^{n'}u^{n'}\neq\pi^{n'}u'^{n'}$. It follows that $|\mathcal{B}_{y'^{n'}}|\geq|\mathcal{B}_{y^{n'}}|$. $y^{n'}$ and $y'^{n'}$ in the above argument can be interchanged. Thus, $|\mathcal{B}_{y^{n'}}|\geq|\mathcal{B}_{y'^{n'}}|$. It follows that $|\mathcal{B}_{y^{n'}}|=|\mathcal{B}_{y'^{n'}}|$. Thus, it follows that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y'^{n'})>D\right\}\right) \quad (5.188)$$

$V_q^{n'}$ denotes the uniform random variable on the set of all sequences of all type q. Let V_q^{n} be independent of $U^{n'}$. It follows, by use of 5.188 that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right)=\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},V_q^{n'})>D\right\}\right) \quad (5.189)$$

Next, we prove for the source-coding problem that if $u^{n'}, u'^{n'} \in \mathcal{U}^{n'}$ (in particular, they have the same type), then

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u'^{n'}, V_q^{n'}) > D\right\}\right) \tag{5.190}$$

Since $V_q^{n'}$ is the uniform distribution on the set of sequences $\mathcal{U}_q^{n'}$ of type q, it follows that it is sufficient to prove that the cardinalities of the sets

$$\left\{y^{n'} : \frac{1}{n'}d^{n'}(u^{n'}, y^{n'}) > D\right\} \text{ and } \left\{y^{n'} : \frac{1}{n'}d^{n'}(u'^{n'}, y^{n'}) > D\right\} \tag{5.191}$$

are equal.

Since $u^{n'}$ and $u'^{n'}$ belong to the set $\mathcal{U}^{n'}$, $u'^{n'}$ is a permutation of $u^{n'}$. Let $u'^{n'} = \pi^{n'} u^{n'}$.

Denote the sets

$$\mathcal{D}_{u^{n'}} \triangleq \left\{y^{n'} : \frac{1}{n'}d^{n'}(u^{n'}, y^{n'}) > D\right\} \tag{5.192}$$

and

$$\mathcal{D}_{u'^{n'}} \triangleq \left\{y^{n'} : \frac{1}{n'}d^{n'}(u'^{n'}, y^{n'}) > D\right\} \tag{5.193}$$

Let $y^{n'} \in \mathcal{D}_{y^{n'}}$. Since the distortion measure is permutation invariant, $d^{n'}(\pi^{n'} u^{n'}, \pi^{n'} y^{n'}) = d^{n'}(u^{n'}, y^{n'})$. Thus, $pi^{n'} y^{n'} \in \mathcal{D}_{u'^{n'}}$. If $y^{n'} \neq y'^{n'}$, $\pi^{n'} y^{n'} \neq \pi^{n'} y'^{n'}$. It follows that $|\mathcal{D}_{u'^{n'}}| \geq |\mathcal{D}_{u^{n'}}|$. $u^{n'}$ and $u'^{n'}$ in the above argument can be interchanged. Thus, $|\mathcal{D}_{u^{n'}}| \geq |\mathcal{D}_{u'^{n'}}|$. It follows that $|\mathcal{D}_{u'^{n'}}| = |\mathcal{D}_{u^{n'}}|$. Thus, it follows that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u'^{n'}, V_q^{n'}) > D\right\}\right) \tag{5.194}$$

$U^{n'}$ denotes the uniform random variable on $\mathcal{U}^{n'}$. Let $U^{n'}$ be independent of $V_q^{n'}$. It follows from 5.194 that

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, V_q^{n'}) > D\right\}\right) \tag{5.195}$$

From (5.190) and (5.195), it follows that if $y^{n'}$ has type q,

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right) \tag{5.196}$$

It follows that

$$\inf_{y^{n'}\in\mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right) = \inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'})>D\right\}\right) \quad (5.197)$$

This proves what we had set out to prove in the connection between source and channel coding.

Denote

$$F^{n'} \triangleq \inf_{y^{n'}\in\mathcal{Y}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'})>D\right\}\right) = \inf_{q\in\mathcal{G}^{n'}} \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'})>D\right\}\right) \quad (5.198)$$

Pseudo-universal capacity of the channel set $\mathcal{C}_{U,D}$ is $\leq$ the rate-distortion function $R_U^P(D,\inf)$

- *Channel coding problem:* From (5.174), it follows that rate R is achievable if

$$[F^{n'}]^{2^{\lfloor n'R\rfloor}-1} \to 1 \text{ as } n' \to \infty \quad (5.199)$$

- *Source coding problem:* From (5.180), it follows that rate R is achievable if

$$[F^{n_i'}]^{2^{\lfloor n_i'R\rfloor}} \to 0 \text{ as } n_i' \to \infty \text{ for some } n_i' = n_0 n_i, \text{ for some } n_i \to \infty \quad (5.200)$$

If rate R is achievable for the channel-coding problem, so is any rate $< R$. Define:

$$\alpha \triangleq \sup\{R|\text{rate } R \text{ is achievable for the channel coding problem by use of the above random-coding method}\} \quad (5.201)$$

Then,

$$\lim_{n_i\to\infty} (F^{n_i})^{2^{\lfloor n_i R'\rfloor}-1} < 1 \;\forall\, R' > \alpha \text{ for some sequence } n_i \to \infty \quad (5.202)$$

Thus,

$$\lim_{n_i\to\infty} (F^{n_i})^{2^{\lfloor n_i R''\rfloor}-1} = 0 \text{ for } R'' > R' \quad (5.203)$$

Note that $R'' > R' > \alpha$, but other than that, R' and R'' are arbitrary. It follows that rates $\leq \alpha$ are achievable for the source coding problem.

Note that the above random-coding method is just one possible method to generate codes for the channel coding problem. In general, it is possible that there exists another coding method which performs better than the above random-coding method, that is, for which rates $> \alpha$ are achievable for the channel coding problem. Thus, what we can claim from the above argument is that rates $< \alpha$ are achievable for the channel-coding problem. Thus, $pC_{rc}(\mathscr{C}_{U,D}) \geq \alpha$. Similarly, the above random-coding method is just one possible method to generate codes for the source coding problem. In general, it is possible that there exists another coding method which performs better than the above random-coding method, that is, for which rates $< \alpha$ are achievable for the source-coding problem when we use the probability of excess distortion criterion with the inf definition. That is, $R_U^P(D, \inf) \leq \alpha$. Thus, $pC_{rc}(\mathscr{C}_{U,D}) \geq \alpha$ and $R_U^P(D, \inf) \leq \alpha$. In particular, $pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^P(D, \inf)$.

Definition 5.66 (α). Note, from the above proof α is equal to the answer to the random-coding argument described in the above proof for both the channel-coding problem of the pseudo-universal capacity of the set of channels $\mathscr{C}_{U,D}$ and the rate-distortion function for compressing the uniform X source under the inf-probability of excess distortion criterion. The answer to these random-coding arguments is then, our definition of α.

Next, we prove that $pC_{rc}(\mathscr{C}_{U,D}) \leq R_U^P(D)$. The main idea of the proof here is to *think of the transition probability corresponding to a "good" source-code as a "bad" channel.*

Let $s = < s^{n'} >_1^\infty$ be a rate $R_U^P(D) + \epsilon$ source-code which compresses the uniform X source, U, to within a probability of excess distortion D. By Lemma 5.2 , we can assume that $s = < s^{n'} >_1^\infty = < e^{n'}, f^{n'} >_1^\infty$ is deterministic. Thus, the cardinality of the image of $f^{n'} \circ e^{n'}$ is $\leq 2^{n'[R_U^P(D)+\epsilon]}$.

Consider the channel $c_\epsilon = < c_\epsilon^n >_1^\infty = < f^{n'} \circ e^{n'} >_1^\infty$. Then, $c_\epsilon \in \mathscr{C}_{U,D}$. Since the cardinality of the image of $f^{n'} \circ e^{n'}$ is $\leq 2^{n'[R_U^P(D)+\epsilon]}$, the capacity of $c_\epsilon \leq R_U^P(D)+\epsilon$ (this is quite intuitive; however, a rigorous proof is proved below, after a few lines).

Such a channel c_ϵ exists for all $\epsilon > 0$. Recall that $c_\epsilon \in \mathscr{C}_{U,D}$. It follows that $pC_{rc}(\mathscr{C}_{U,D}) \leq R_U^P(D) + \epsilon \ \forall \epsilon > 0$. It follows that $pC_{rc}(\mathscr{C}_{U,D}) \leq R_U^P(D)$.

It remains to prove that the capacity of $c_\epsilon \leq R_U^P(D) + \epsilon$. Note that we have defined capacity with the maximal block error probability criterion

$$\sup_{m^{n'} \in \mathscr{M}^{n'}} \Pr(\hat{M}^{n'} \neq M^{n'} \mid M^{n'} = m^{n'}) \to 0 \text{ as } n' \to \infty \tag{5.204}$$

Note also, that the maximal block error probability criterion is a stronger criterion than the average block error probability criterion

$$\Pr(\hat{M}^{n'} \neq M^{n'}) \to 0 \text{ as } n' \to \infty, \text{ where } M^{n'} \text{ is uniform} \tag{5.205}$$

Let rate R be achievable with a possibly random channel-code $< E_0^{n'}, F_0^{n'} >_1^\infty$ over the channel c_ϵ under the maximal block error probability criterion. Then, rate R is also achievable over the channel c_ϵ by using the same channel code $< E^{n'}, F^{n'} >_1^\infty$ under the average block error probability criterion. For the average block error probability criterion, if rate R is achievable by using a random-code, rate R is also achievable by use of a deterministic code. Thus, rate R be achievable over the channel c_ϵ under the average block error probability criterion by use of a deterministic channel code $< e_0^{n'}, f_0^{n'} >_1^\infty$. An exercise in [CK97] tells that there exists a way of throwing away half the codewords such the maximal block error probability is less than or equal to twice the average block error probability for each block length. Let $\xi > 0$. It follows that there exists a deterministic source code $< e_1^{n'}, f_1^{n'} >_1^\infty$ such that rate $R - \xi$ is achievable over the channel c_ϵ under the maximal block error probability criterion by use of the deterministic source code $< e_1^{n'}, f_1^{n'} >_1^\infty$. Now, cardinality of the image of $f^{n'} \circ e^{n'}$ is $\leq 2^{n'[R_U^P(D)+\epsilon]}$, and another way of thinking about this is that when the block-length is n', the output space of the channel c_ϵ^n has cardinality $\leq 2^{n'[R_U^P(D)+\epsilon]}$. With the deterministic code $< e_1^{n'}, f_1^{n'} >_1^\infty$, it follows that the cardinality of the image of $e_1^{n'} \circ c_\epsilon^{n'} \circ f_1^{n'}$ is $\leq 2^{n'[R_U^P(D)+\epsilon]}$. If $R - \xi$ were greater than $R_U^P(D) + \epsilon$, it follows that some message $m^{n'}$ will not have an image in $e_1^{n'} \circ c_\epsilon^{n'} \circ f_1^{n'}$ and the maximal block error probability will be 1. Thus, $R - \xi \leq R_U^P(D) + \epsilon$. $\xi > 0$ is arbitrary. It follows that rates $> R_U^P(D) + \epsilon$ would lead to a maximal block error probability of 1 over the channel c_ϵ. Thus, the capacity of the channel $c_\epsilon \leq R_U^P(D) + \epsilon$. It follows, as we said above, that $pC_{rc}(\mathscr{C}_{U,D}) \leq R_U^P(D)$.

Thus, we have proved that $R_U^P(D) \geq pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^P(D, \inf)$. This completes the proof. □

Next, we go on to prove one of the main theorems of this chapter: the equality of the pseudo-universal capacity and the rate-distortion functions.

5.6.2 Proof of $pC_{rc}(\mathscr{C}_{X,D}) = R_X^E(D) = R_X^P(D) = R_X^E(D, \inf) = R_X^P(D, \inf)$

First, we prove that $pC_{rc}(\mathscr{C}_{U,D}) = R_U^P(D) = R_U^P(D, \inf)$

Theorem 5.47. $pC_{rc}(\mathscr{C}_{U,D}) = R_U^P(D) = R_U^P(D, \inf)$

Proof. By Theorem 5.46, $R_U^P(D) \geq pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^P(D, \inf)$. By Theorem 5.37, $R_U^P(D, \inf) = R_U^P(D)$. It follows that $pC_{rc}(\mathscr{C}_{U,D}) = R_U^P(D) = R_U^P(D, \inf)$. This completes the proof. □

Note 5.62. We have defined the rate-distortion function with both the lim inf and the lim sup definitions. Most literature defines the rate-distortion function with the lim sup definition. One of the reasons why we have defined the rate-distortion function with both the lim inf and the lim sup definitions is for the sake of completeness. Another reason why we have defined the rate-distortion function with the lim inf definition is because the tabular proof in Theorem 5.46 works only with the lim inf definition: we do not know, how to prove $pC_{rc}(\mathscr{C}_{U,D}) = R_U^P(D$ without involving the definition of $R_U^P(D, \inf)$. Then, we use the equality of the rate-distortion functions $R^P_U(D)$ and $R^P_U(D,\inf)$ proved in Theorem 5.37 to prove Theorem 5.47.

Finally, we prove the desired result: $pC_{rc}(\mathscr{C}_{X,D}) = R^E_X(D) = R^P_X(D) = R^E_X(D,\inf) = R^P_X(D,\inf)$. This is done by taking limits along $X_n \to X$ where X_n are random-variables which satisfy $p_{X_n}(x)$ is rational $\forall x \in \mathscr{X}$, and X is arbitrary.

Theorem 5.48. *Let X be an arbitrary random variable on $\mathscr{X}$. For $D \in (0,\infty)$. $pC_{rc}(\mathscr{C}_{X,D}) = R^E_X(D) = R^P_X(D) = R^E_X(D,\inf) = R^P_X(D,\inf)$.*

Proof. Let $X_n, 1 \le n \le \infty$ be a collection of random variables on $\mathscr{X}$ such that $p_{X_n}(x)$ is rational $\forall x \in \mathscr{X}$. Let $X_n \to X$ in distribution. That is, $\forall x \in \mathscr{X}$, $\lim_{n\to\infty} p_{X_n}(x) = p_X(x)$. U_n is the uniform X_n source. Let $\epsilon_n > 0$ and $\epsilon_n \to 0$ as $n \to \infty$. Define $\delta_n \triangleq D_{\max}(d_1(p_X, p_{X_n}) + \epsilon_n)$. First we prove that for $D \in (0,\infty)$,

$$\lim_{n\to\infty} R^E_{X_n}(D+\delta_n) = pC_{rc}(\mathscr{C}_{X,D}) \tag{5.206}$$

From Lemmas 5.44 and 5.45, it follows that

$$pC_{rc}(\mathscr{C}_{U_n,D+\delta_n}) \le pC_{rc}(\mathscr{C}_{X,D}) \le pC_{rc}(\mathscr{C}_{U_n,D-\delta_n}) \tag{5.207}$$

By Theorem 5.47, it follows that

$$R^E_{U_n}(D+\delta_n) \le pC_{rc}(\mathscr{C}_{X,D}) \le R^E_{U_n}(D-\delta_n) \tag{5.208}$$

Rest of the proof of (5.206) follows exactly as in Lemma 5.39 by replacing $R^E_X(D)$ with $pC_{rc}(\mathscr{C}_{X,D})$ in the steps (5.134).

The proof of this theorem, now, follows by use of Lemma 5.39 and Lemma 5.41. □

We use Theorem 5.48 to prove the main result of this chapter: an operational proof of the optimality of digital communication for the problem of pseudo-universal communication with a fidelity criterion.

■ 5.7 An operational view of the optimality of digital communication for pseudo-universal communication with a fidelity criterion, and a discussion of the operational nature of the proof

By using Theorem 5.48, in almost exactly the same way as Step 2 in Section 2.15, we will prove, in this section, the optimality of digital communication for pseudo-universal communication with a fidelity criterion. Because of all the machinery developed in this chapter, the proof will be operational.

As stated in Subsection 5.1.1,

- We use pseudo-universal achievability instead of universal achievability in our proofs. We are quite sure that the proof can be carried out even for universal achievability but I have not carried it out. For the reader unhappy with this explanation, the reader can think of this as an operational view point on the optimality of digital communication for a fully known channel because for a fully known channel, pseudo-universal and universal achievability are the same as argued in Subsection 5.1.1. Similarly, instead of universal communication with a fidelity criterion over a partially known channel $k \in \mathscr{A}$, we will be using pseudo-universal communication with a fidelity criterion over a partially known channel: we do not enforce uniformity in the rate at which the probability of excess distortion $\to 0$ as block-length $\to \infty$ over the particular channel $k \in \mathscr{A}$. Generalization to proving results universal instead of pseudo-universal results operationally is discussed in Section 5.9.

- We do not take into account resource consumption in the system when constructing the digital architecture. We are quite sure that a few more arguments can be made to take into account the resource consumption in the system but I have not carried out these steps. For the reader unhappy with this explanation, the reader should think of it as the way things are done in the usual information theory literature in the discrete case where resource consumption is not considered at all. I should add that I do not agree with this approach in the literature because resource consumption is a very important issue, and when proving optimality of digital communication, one should prove that it can be done with the same or lesser resource consumption as compared to other architectures; in our case in this chapter, I quite strongly believe that it can be done; just that I have not done it. This is commented on vaguely in Section 5.10

With the above constraints, we will carry out the proof of optimality of digital communication for pseudo-universal communication with a fidelity criterion (in other words, an operational view of the pseudo-universal source-channel separation theorem for rate-distortion; the proof is carried out for the i.i.d. X source; of course, essentially the same proof can be carried out for the uniform X source:

In Subsection 5.7.1, we make a formal definition of what it means for a partially known channel to be capable of pseudo-universally communicating a random source to within a certain distortion level. In Sub-section 5.7.2, we state the pseudo-universal source-channel separation theorem for rate-distortion. The operational proof will crucially use Theorem 5.48 and the fact that its proof is operational and by use of this theorem, the final step of the operational proof is carried out in Subsection 5.7.3. This is followed by a discussion of the operational nature of the proof in Subsection 5.7.4.

5.7.1 Capability of a partially known channel to pseudo-universally communicate a random source to within a certain distortion level

Let $k \in \mathscr{A}$ be a partially known channel with input space $\mathscr{I}$ and output space $\mathscr{O}$ as described in Subsection 2.5.2.

First, we describe the point-to-point communication system which communicates i.i.d. X source over a channel $k \in \mathscr{A}$. Recall the action of a a point-to-point communication system in described in Subsection 2.6.6.

The input to the encoder is the i.i.d. X source. Thus, when the block-length is n, the input is the i.i.d. X sequence of length n, X^n. The composition of the encoder, channel, and decoder, produce an output sequence Y^n. This results in a joint random variable X^nY^n on the input-output space $\mathscr{X}^n \times \mathscr{Y}^n$ and the corresponding probability distribution $p_{X^nY^n}$ on $\mathscr{X}^n \times \mathscr{Y}^n$. Note that when we are talking about a partially known channel $p_{X^nY^n}$ will vary depending on the particular $k \in \mathscr{A}$.

Definition 5.67 (A partially known channel which is *capable of* pseudo-universally communicating i.i.d. X source to within a distortion level D). The partially known channel $k \in \mathscr{A}$ is said to be capable of communicating i.i.d. X source to within a distortion D if there exists an encoder-decoder pair $< e^n, f^n >_1^\infty$ independent of the particular $k \in \mathscr{A}$ such that under the joint distribution $p_{X^nY^n}$ as described above,

$$p_{X^nY^n}\left(\frac{1}{n}d^n(X^n, Y^n) > D\right) \to 0 \text{ as } n \to \infty \ \forall\ k \in \mathscr{A} \tag{5.209}$$

The reader should compare this definition with the definition of a partially known channel which is capable of universally communicating i.i.d. X source to within a distortion level D in Definition 2.25: there is no $\omega =< \omega^n >_1^\infty$ sequence in the definition anymore which was introduced to enforce the uniformity over the partially known channel $k \in \mathscr{A}$ in the rate at which the probability of excess distortion $\to 0$ as $n \to \infty$. Now, we do not ask for this uniformity.

A similar definition can be made for the uniform X source which we omit because our main goal is to prove results for the i.i.d. X source; uniform X source is a good intuitive as well as a good mathematical tool towards that goal.

5.7.2 A statement of the pseudo-universal source-channel separation theorem for rate-distortion

Assuming random-coding is permitted, in order to communicate the i.i.d. X source pseudo-universally over a partially known channel to within a particular distortion level, it is sufficient to consider source-channel separation based architectures, that is, architectures which first code (compress) the i.i.d. X source to within the particular distortion level, followed by pseudo-universal reliable communication over the partially known channel. There is sufficiency in the sense if there exists some architecture to communicate the random source to within the required distortion level, pseudo-universally over the partially known channel, there exists a separation architecture which accomplishes the same thing.

The same result holds for the uniform X source.

5.7.3 The final step of the operational proof of the pseudo-universal source-channel separation theorem for rate-distortion for the i.i.d. X source

Let $k \in \mathscr{A}$ be a partially known channel which is capable of pseudo-universally communicating the i.i.d. X source to within a distortion level D. This is accomplished with the help of an encoder-decoder $< e^n, f^n >_1^\infty$. Denote the set of channels

$$\mathscr{C}_{\mathscr{A}} \triangleq \{< e^n \circ k \circ f^n >_1^\infty \mid k \in \mathscr{A}\} \tag{5.210}$$

Note that $\mathscr{C}_{\mathscr{A}}$ is a subset of $\mathscr{C}_{X,D}$. It follows by Theorem 5.48 that the pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$ is $\geq$ the rate-distortion function $R_X^P(D)$.

It now follows that by source-compression followed by pseudo-universal reliable communication, the i.i.d. X source can be communicated universally and reliably over the partially known channel k to within a distortion D. A rough argument is the following: Take the i.i.d. X source. Compress it using a source-encoder $< e_s^n >_1^\infty$ to within a probability of excess distortion D. The output is a rate $R_X^P(D)$ message source. This rate $R_X^P(D)$ message source can now be communicated pseudo-universally and reliably over the channel set $\mathscr{C}_{\mathscr{A}}$ by using a channel encoder-decoder $< E^n, F^n >_1^\infty$. Finally, the output of the channel decoder is source decoded using a source decoder $< f_s^n >_1^\infty$. End-to-end, the i.i.d. X source is pseudo-universally communicated to within a distortion D over the channel set $\mathscr{C}_{\mathscr{A}}$. It follows that by using a source-code $< e_s^n, f_s^n >_1^\infty$ and using a channel code $< e_c^n \circ E^n, F^n \circ f_c^n >_1^\infty$, the i.i.d. X source can be communicated pseudo-universally to within a distortion level D over the partially known channel k, digitally.

A rigorous argument for the above source-coding followed by channel coding is the following:

We said above that the pseudo-universal capacity of the channel set $\mathscr{C}_{\mathscr{A}}$ is $\geq R_X^P(D)$. Assume that the pseudo-universal capacity is strictly $> R_X^P(D)$. Let the universal capacity be $R_X^P(D) + \delta, \delta > 0$.

Let $\epsilon = \frac{\delta}{2}$. By the definition of $R_X^P(D)$, it follows that there exists a rate $R_X^P(D) + \epsilon$ source-code $< e_s^n, f_s^n >_1^\infty$ which compresses the i.i.d. X source to within a probability of excess distortion D.

Let the block-length be n.

The action of e_s^n on X^n produces an output random variable M_R^n on the set $\mathscr{M}_R^n$, where the set $\mathscr{M}_R^n$ is

$$\mathscr{M}_R^n = \{1, 2, \ldots, 2^{\lfloor n(R_X^P(D)+\epsilon)\rfloor}\} \tag{5.211}$$

Since the pseudo-universal capacity of the channel set $\mathscr{C}_{\mathscr{A}}$ is strictly greater than $R_X^P(D) + \epsilon$ by assumption, the message M_R^n can be pseudo-universally and reliably communicated over the partially known channel k in the limit as $n \to \infty$. Finally, the source is re-constructed by using the source decoder f_s^n.

It follows that end-to-end, in this separation based architecture, in the limit as $n \to \infty$, the i.i.d. X source is communicated pseudo-universally to within a distortion level D over the channel set $\mathscr{C}_{\mathscr{A}}$, and hence, over the partially known channel k.

Note that we assumed that the pseudo-universal capacity of the channel set $\mathscr{C}_{\mathscr{A}}$ is strictly $> R_X^P(D)$, whereas it only follows that the pseudo-universal capacity of the channel set $\mathscr{C}_{\mathscr{A}}$ is $\geq R_X^P(D)$. It is unclear what will happen if the capacity of the partially known channel k is precisely $R_X^P(D)$. This "tension" of what happens if the capacity is precisely $R_X^P(D)$ is usual in information theory.

This completes the argument, and thus, rigorously proves the pseudo-universal source-channel separation theorem for rate-distortion when the source is i.i.d. and the distortion metric is additive.

Note 5.63. In order the prove the pseudo-universal source-channel separation theorem for rate-distortion we have only used the fact that the pseudo-universal capacity of the channel set $\mathscr{C}_{X,D}$ is $\geq R_X^P(D)$. We proved in Theorem 5.48 that the universal capacity of the channel set $\mathscr{C}_{X,D}$ is in fact, precisely $R_X^P(D)$. We will use this precise equality to discuss connections between source and channel coding and to give the idea of an alternate proof of the rate-distortion theorem for certain i.i.d. X sources in Section 5.8.

In the next subsection, we comment on the operational nature of the proof of the pseudo-universal source-channel separation theorem for rate-distortion.

■ 5.7.4 Discussion: Operational nature of our proof of the pseudo-universal source-channel separation theorem for rate-distortion, and a comparison with Shannon's proof

Our proof of the pseudo-universal source-channel separation theorem for rate-distortion is operational in the sense that it uses only the definitions of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a certain distortion level. We do not use functional simplifications, for example, mutual information expressions for the channel capacity or the rate-distortion function in the proof.

The proof consists of two steps:

1. Step 1: If there exists some scheme in order to communicate the i.i.d. X source pseudo-universally over the partially known channel k to within a distortion D under the probability of excess distortion criterion, then the pseudo-universal capacity of k is $\geq R_X^p(D)$. One main step in this proof is the proof of Theorem 5.48, which says, in part, that the pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$ is equal to the rate-distortion function for the i.i.d. X source. The whole chapter before the proof of Theorem 5.48 is devoted to proving this theorem rigorously and operationally. This whole big proof is operational in the sense that we only use the definitions of source-codes, channel-codes and the operational meanings of rate-distortion function and channel capacities. We do not rely on mathematical functional simplifications. This main idea of the proof is Theorem 5.46, random-coding arguments carried out in parallel for the source-coding and channel-coding problems where we can "see" that the answers to the two problems is the same without doing much functional simplifications: this is what makes it operational. Mathematically, the main idea is (5.183) which is the same equation as (2.66) and is reproduced below:

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'}) > D\right\}\right), \text{ if } y^{n'} \text{ has type } q \tag{5.212}$$

2. Step 2: If the pseudo-universal capacity of a partially known channel k is $> R_X^p(D)$, then pseudo-universal communication of the i.i.d. X source to within a probability of excess distortion D can be accomplished over the channel k by source compression followed by pseudo-universal reliable communication. Clearly this proof is operational in that we need to know only the operational meaning of $R_X^p(D)$ as the rate at which we can compress the i.i.d. X source to within a distortion D and we only need to know the definition of pseudo-universal capacity as the maximum achievable rate at which pseudo-universal reliable communication is possible over the partially known channel k

I would like to further comment on the operational nature of the proof of Theorem 5.48. First note, that the concept of an operational proof is *not* a precise concept. The question of pseudo-universal capacity of the set of channels $\mathscr{C}_{X,D}$, when posed in mathematical language, is an infinite dimensional optimization program. Similarly, the rate-distortion problem of the minimum rate needed to compress a source to within a certain distortion level is an infinite dimensional optimization program. We need to prove that the answers to both these infinite dimensional optimization programs is the same. A "truly operational" proof will just prove this without any steps. However, no proof can be without any steps; otherwise, the result would be trivial. The main step that we go through is the parallel random-coding argument where we prove that $pC_{rc}(\mathscr{C}_{U,D}) \geq R_U^p(D, \inf)$. We do go through some steps where we use a random-coding argument and make coding schemes for both the channel coding and the source coding problems. We believe that none of these steps make the proof non-operational; however, this is a concept which we cannot make precise.

We would like to compare our proof with that of Shannon, which we believe is non operational. Shannon's proof, again, goes in two steps:

1. Achievability, which is the same as our Step 2: This proof is exactly the same as ours (rather, we should say that our proof is exactly the same as Shannon's!), and is operational

2. Converse, which is the same as our Step 1 (though as we discussed in Subsection 2.14.10, we view it as achievability) : This proof is different from ours. Shannon's proof relies heavily on the information-theoretic definitions of channel capacity as a maximum mutual information and the rate-distortion function as a minimum-mutual information: these mutual information expressions are simplified finite dimensional optimization programs corresponding to the original infinite dimensional optimization programs.

   The following is an outline of Shannon's proof. Shannon's proof required the channel k to be fully known and used the expected distortion criterion instead of the probability of excess distortion criterion. Let the channel k be a discrete memoryless channel. Denote the discrete memoryless channel k by $p_{O|I}$.

   We will not be entirely precise: see [Sha59] for the precise details. The converse consists of two steps:

   - $$H(X^n|Y^n) \geq H(X^n) - nC^I \quad \left[C^I \triangleq \max_{p_I} I(I;O)\right] \tag{5.213}$$

     This says that since the source X^n communicated over a discrete memoryless channel, the entropy of the source at the output cannot fall by "too much". "Too much" is quantified by the information-theoretic capacity C^I of the discrete memoryless channel. The proof of this step uses the definitions of entropy, mutual information, information theoretic capacity, and inequalities concerning entropy and mutual information.

   - $$H(X^n|Y^n) \leq H(X^n) - nR^I(D) \quad \left[R_X^I(D) \triangleq \inf_{\substack{X \sim p_X \\ Ed(X,Y) \leq D}} I(X;Y)\right] \tag{5.214}$$

     This says that since the source X^n has been communicated to within a distortion D under the expected distortion criterion over the channel (by using some architecture which does not matter), the entropy of the source should have fallen by atleast a particular amount, and this particular amount is quantified by the information-theoretic rate-distortion function $R_X^I(D)$, which was defined in the above equation and has also been defined in Definition 2.30. This is proved by using information-theoretic inequalities concerning entropy and mutual information and the convexity of the $R^I(D)$ function.

These two steps imply that

$$C^I \geq R_X^I(D) \tag{5.215}$$

Finally, one invokes the fact that for a discrete memoryless channel, the information-theoretic capacity is the same as operational capacity, and the information-theoretic rate-distortion function for an i.i.d. source is the same as the operational rate-distortion function, and this proves that the capacity of the channel k is greater than or equal to the rate-distortion function $R_X^E(D)$.

Shannon's proof of the converse is a brilliant proof but I have never had much intuition for it, mainly because it goes through first proving the equality of the information theoretic rate-distortion function and the information-theoretic channel capacity, and then invoking the equality of the rate-distortion function and the information-theoretic rate-distortion function, and the equality of channel capacity and the information-theoretic channel capacity. In this sense, it is not operational.

Our proof, which relies only on the operational meanings and uses a random-coding achievability argument, we believe, lends much more insight into the nature of separation, for which, as we said, in Subsection 2.14.10, that the fundamental mathematical reason is (5.183) which is reproduced below:

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'},y^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'},V_q^{n'}) > D\right\}\right), \text{ if } y^{n'} \text{ has type } q \tag{5.216}$$

5.8 Connections between source and channel coding and an alternate proof of the source-channel separation theorem for rate-distortion for those i.i.d. X sources for which $p_X(x)$ is rational $\forall x \in \mathcal{X}$

In this section, we discuss connections between source and channel coding that come out of our work and related to this, we give another proof of the rate-distortion theorem for those i.i.d. X sources for which $p_X(x)$ is rational $\forall x \in \mathcal{X}$, which we believe is more insightful than the original proof of Shannon [Sha59].

5.8.1 Connections between source and channel coding

The connection between source and channel coding was discussed in Subsection 2.14.10. We discuss is further, here. The discussion will be high-level.

We discuss these connections for the uniform X source. Recall that the uniform X source is defined only for those X for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$. On the level of ideas, the proof of $pC_{rc}(\mathscr{C}_{U,D}) = R_U^p(D)$ has the following steps:

- Step 0: Prove that $R_U^p(D) = R_U^p(D, \text{inf})$.: This is proved by using the operational meanings of source codes and the two rate-distortion functions
- Step 1: Prove that $R_U^p(D, \text{inf}) \le \alpha \le pC_{rc}(\mathscr{C}_{U,D})$: We make random coding arguments for both the channel coding problem and rate-distortion source-coding problems: the channel-coding problem is the pseudo-universal capacity of the set of channels $\mathscr{C}_{U,D}$ and the rate-distortion problem is the minimum rate needed compress the uniform X source to within a distortion level D under the inf-probability of excess distortion criterion. We note that the answers to both these random-coding arguments is α. Since the random-coding scheme is just one possible scheme for pseudo-universal communication over the set of channels $\mathscr{C}_{U,D}$, it is potentially possible that a scheme exists which performs better than the random-coding scheme, that is, a scheme for which rates $> \alpha$ might be achievable for the channel coding problem . Thus, $pC_{rc}(\mathscr{C}_{U,D}) \ge \alpha$. Similarly, the random-coding scheme is just one possible scheme for compression of the uniform X source to within a distortion level D under the inf-probability of excess distortion criterion, and it is potentially possible that a scheme exists which performs better than the random-coding scheme, that is, a scheme for which rates α are achievable for the source-coding problem. Thus, $R_U^p(D, \text{inf}) \le \alpha$.

  Thus, $R_U^p(D, \text{inf}) \le \alpha \le pC_{rc}(\mathscr{C}_{U,D})$
- Step 2: Prove that $pC_{rc}(\mathscr{C}_{U,D}) \le R_U^p(D)$: This is done by noting that the transition probability corresponding to a "good source-code" for the problem of compressing the uniform X source to within a distortion D under the probability of excess distortion criterion. is a "bad channel" for the purpose of reliable communication. The transition probability corresponding to this "good source code", when thought as a channel has capacity $R_U^p(D)$, and hence, the infimum capacity of the channel class $\le R_X(D)$.

From these steps, it follows that $R_U^p(D) = \alpha = pC_{rc}(\mathscr{C}_{U,D})$. In particular, $pC_{rc}(\mathscr{C}_{U,D}) = R_U^p(D)$.

If one notes the parallel tabular argument for the channel-coding and source-coding problems, there is a kind of "duality" in the two arguments. These are random coding arguments, and one can, at least to some extent, interpret this as a covering-packing connection, maybe a random-coding covering-packing connection. The proofs here shed light on connection between the two problems. It should be possible to make this connection/duality precise; however, we *do not* know, how to. As we stated in Subsection 2.14.10, the precise mathematical reason for this duality is (5.183), which is reproduced below:

$$\Pr\left(\left\{\frac{1}{n'}d^{n'}(U^{n'}, y^{n'}) > D\right\}\right) = \Pr\left(\left\{\frac{1}{n'}d^{n'}(u^{n'}, V_q^{n'}) > D\right\}\right), \text{ if } y^{n'} \text{ has type } q \quad (5.217)$$

5.8.2 An alternate proof of the rate-distortion theorem for those i.i.d. X sources for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$

The original rate-distortion theorem of Shannon [Sha59], stated as Theorem 2.1 in this thesis, says that a simplified expression for $R_X^E(D)$ is $R_X^I(D)$, where $R_X^I(D)$ is defined in Definition 2.30, and this definition is reproduced below:

$$R_X^I(D) \triangleq \inf_{\{p_{Y|X} : \sum_{x \in \mathscr{X}, y \in \mathscr{Y}} p_X(x) p_{Y|X}(y|x) \leq D\}} I(X;Y) \tag{5.218}$$

We give an alternate proof that a expression for $R_X^P(D)$ is $R_X^I(D)$ for those i.i.d. X sources for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$ which goes through the channel-coding problem and which, I believe is more insightful than the original proof of Shannon for reasons which will follow. Of course, since we have proved the equality of $R_X^P(D)$ and $R_X^E(D)$, this will also prove that a simplified expression for $R_X^E(D)$ is $R_X^I(D)$.

First, let us look at Shannon's argument for the i.i.d. X source, more details of which can be found in [Sha59]:

In a nut-shell, ShannonÕs proof goes as follows: by using the same random coding argument, it is proved that $R_X^E(D) \leq \alpha$. Then one does a computation for α and proves that it is equal to $R_X^I(D)$. This, then, proves that $R_X^E(D) \leq R_X^I(D)$. The proof of $R_X^I(D) \leq R_X^E(D)$ uses information-theoretic inequalities and properties of mutual information. The functional form of $R_X^I(D)$ is thus used crucially in ShannonÕs proof: one has to compute $\alpha = R_X^I(D)$; only then is the proof complete. Unlike Shannon, our proof of converse does not use information-theoretic inequalities or properties of mutual information, but rather, rests on connections between our channel coding problem and the rate-distortion problem. Following is a summary of the steps:

1. Prove $R_X^E(D) \leq \alpha$ (random-coding argument)
2. Compute $\alpha = R_X^I(D)$ (mathematical computation)
3. Thus, $R_X^E(D) \leq R_X^I(D)$
4. Prove $R_X(D) \geq R_X^I(D)$ (information theoretic inequalities)
5. Thus, $R_X^E(D) = R_X^I(D)$

Next, we look at our argument of the proof of the rate-distortion theorem *for those sources X for which $p_X(x)$ is rational* $\forall x \in \mathscr{X}$. Recall that U is the uniform X Source corresponding to X. Some of these steps are the same as the ones of the previous subsection. The steps are:

1. Prove that $R_U^P(D, \inf) \leq \alpha \leq pC_{rc}(\mathscr{C}_{U,D})$ (random-coding arguments)

2. Prove that $pC_{rc}(\mathscr{C}_{U,D}) \leq R_U^P(D)$ (the transition probability corresponding to a "good" source-code is a "bad" channel)
3. Prove that $R_U^P(D) = R_U^P(D, \inf)$ (definition of rate-distortion functions and particular source code constructions)
4. Thus, $R_U^P(D) = \alpha$
5. Prove that $R_U^P(D) = R_X^P(D) = R_X^E(D)$ (definitions of rate-distortion functions and particular code constructions)
6. Thus, $R_X^P(D) = R_X^E(D) = R_U^P(D) = R_U^E(D) = \alpha$
7. Prove that $\alpha = R_X^I(D)$ (mathematical calculation using the method of types)
8. Thus, $R_X^E(D) = R_X^P(D) = R_U^E(D) = R_U^P(D) = R_X^I(D)$

We consider our argument more basic than Shannon's argument for reason that a mathematical calculation is carried out in the last step in our proof as opposed to in Shannon's proof. In our proof, the proof until Step 6 which says, $R_X^P(D) = R_X^E(D) = R_U^P(D) = R_U^E(D) = \alpha$ should be considered operational, where the operational meaning of α is that it is the answer to the random-coding argument. It is only in Step 7 that we make a calculation for α and prove that it is indeed the information-theoretic rate-distortion function $R_X^I(D)$, and this leads to a proof of the rate-distortion theorem. Step 7 is essentially the finally step in the argument. Shannon's proof relies much more crucially on the fact that an expression for α is $R_X^I(D)$: it is needed in Step 2. The converse part in Shannon's proof which proves that $R_X^E(D) \leq R_X^I(D)$ in Step 4 relies very crucially on the expression for $R_X^I(D)$. Shannon's proof thus has a much more crucial dependence on $\alpha = R_X^I(D)$ as compared to our proof where this calculation is needed in the last step. Of course, since the statement of the rate-distortion theorem is a statement about a simplified expression for the rate-distortion function, a calculation has to be made at some step: in our proof, this calculation is made in the last step.

Also note that we prove that $R_U^P(D)$ is equal to α, the answer to the random-coding argument, without explicitly calculating the answer to the random-coding argument: for this reason, as we said above, this proof is operational. This is unlike Shannon in the sense that Shannon also proves that $R_X^E(D) = \alpha$, the answer to the random-coding argument but the proof requires an explicit calculation of the answer to the random-coding argument. *The operational nature of our proof that the rate-distortion function is equal to the random-coding argument suggests a fundamental relationship between the two quantities.*

Note that our proof holds only for those X for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$. We would like to believe that it should be possible to generalize this proof for arbitrary X; however, we are unsure.

5.9 How do we operationally prove the optimality of digital communication for universal communication with a fidelity criterion instead of for pseudo-universal communication with a fidelity criterion?

In this chapter, we have proved operationally, the optimality of digital communication for pseudo-universal communication with a fidelity criterion.

The question arises: can we similarly provide an operational proof of the optimality of digital communication for universal communication with a fidelity criterion?

Let $k \in \mathscr{A}_X$ be a partially known channel which is capable of universally (note, universally, not pseudo-universally) communicating the i.i.d. X source to within a distortion D. Then, we can define the set of channels $\mathscr{C}_{\mathscr{A}_X}$ as in (2.94) which is reproduced below:

$$c \in \{< e^n \circ k^n \circ f^n >_1^\infty | k = < k^n >_1^\infty \in \mathscr{A}\} \triangleq \mathscr{C}_{\mathscr{A}_X} \tag{5.219}$$

The notation has changed from (2.94), where we used the notation $\mathscr{C}_{\mathscr{A}}$ instead of $\mathscr{C}_{\mathscr{A}_X}$: now, we will be dealing with i.i.d. X and uniform X sources together and for that reason, we need a notation to distinguish between the set $\mathscr{A}_X$ and a corresponding set $\mathscr{A}_U$ which will come later.

The set of channels $\mathscr{C}_{\mathscr{A}}$ should be thought of as a set of channels which *directly* (note, directly, not pseudo-directly) communicates the i.i.d. X source to within a distortion D.

We need to carry out proofs for the set of channels $\mathscr{C}_{\mathscr{A}}$ instead of the set of channels $\mathscr{C}_{X,D}$. In order to do that, we will require the introduction of another set of channels $\mathscr{C}_{\mathscr{A}_U}$ for some $l \in \mathscr{A}_U$, where l is a partially known channel which is capable of communicating the uniform X source to within a distortion D. The set of channels $\mathscr{C}_{\mathscr{A}_U}$ directly (note, directly not pseudo-directly) communicates the uniform X source to within a distortion D.

The main idea of the operational proof is Theorem 5.46 and its proof essentially hold for the set of channels $\mathscr{C}_{A_U}$ and using universal capacity instead of pseudo-universal capacity: this was the way did things in the idea-level proof in Section 2.14. What we need is some arguments to relate the universal capacities of channels $\mathscr{C}_{\mathscr{A}_X}$ and $\mathscr{C}_{\mathscr{A}_U}$: of course, there has to be some relation between the set of channels $\mathscr{A}_X$ and the set of channels $\mathscr{A}_U$ to do this.

Mainly, this should require changes in Subsection 5.5.4. Of course, whole of Section 5.5 would need to be re-done; however, the main lemmas which would need re-statement and would need to be re-proved with universal capacity instead of pseudo-universal capacity are those of Subsection 5.5.4.

We have not carried out these steps; however, we believe they can be carried out.

5.10 How do we take into account resource consumption when proving the pseudo-universal source-channel separation theorem for rate-distortion, operationally?

When proving the pseudo-universal source-channel separation theorem for rate-distortion operationally, we did not take into account, resource consumption in the system, unlike what we did in Chapter 2. This is something that we believe can be done by taking into account the interplay between how the channels with the uniform X and the i.i.d. X source inter-play with each other and how the resource consumption is affected in the arguments in this inter-play. This is vague, but we will leave this here.

5.11 Comments and recapitulation

In this chapter, we proved the optimality of digital communication for pseudo-uinversal communication with a fidelity criterion, operationally. By an operational proof, we mean that the proof uses only the operational meanings of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a certain distortion level, that is, uses only the meanings of channel capacity and rate-distortion function as infinite dimensional optimization programs, and not functional simplifications like finite dimensional mutual-information expressions for them. I believe that this operational proof sheds more insight into the nature of separation than the original proof of Shannon which uses equalities and inequalities concerning entropy and mutual information for proving the converse.

An operational proof of universal source-channel separation for rate-distortion for permutation invariant distortion metrics for the uniform X source was provided in Section 2.14, under the technical assumption $R_U^P(D) = R_U^P(D, \inf)$. In this chapter we proved $R_U^P(D) = R_U^P(D, \inf)$ operationally for additive distortion metrics. We believe that this completes the operational story for universal source-channel separation for rate-distortion.

However, the traditional information theory literature uses the i.i.d. X source. For the sake of partial completeness, we proved a *pseudo-universal* source-channel separation theorem for rate-distortion for the i.i.d. X source. Pseudo-universality differs from universality in that we do not require a uniformity in the rate at which the probability of excess distortion $\to 0$ as the block-length $\to \infty$ over the partially known channel $k \in \mathscr{A}$.

The proof also sheds light on the connections between source-coding and channel-coding. We also provide an alternate proof of the rate-distortion theorem for those i.i.d. X sources for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$ which uses functional simplification as the last step, and I believe, is more insightful than the original proof of Shannon. As a part of this proof, we also prove that the rate-distortion function is equal to the answer to the random-coding argument for the source-coding problem, without doing a calculation, either for the rate-distortion function or the answer to the answer to the random-coding argument for the source-coding problem. This is thus an operational proof, and suggests that there might be a fundamental relationship between the rate-distortion function and the answer to the random-coding argument for the source-coding problem.

The proof of $R_U^p(D) = R_U^p(D, \inf)$ has required the assumption of additive distortion metric. We conjecture that an operational proof can also be provided for many permutation invariant distortion metrics true. Similarly, the rigorous proof of the pseudo-universal source-channel separation for rate-distortion for the i.i.d. X source has required the assumption of additive distortion metric. We conjecture that an operational proof can be provided for many permutation invariant distortion measures, too.

We also conjecture that the results can be generalized operationally, to more general sources, for example, many stationary ergodic sources.

Finally, we emphasize that we have proved the optimality of digital communication for pseudo-universal communication with a fidelity criterion and not for universal communication with a fidelity criterion. Also, we have not taken into account the resource consumption in the system. We believe that these steps can be carried out.

5.12 In the next chapter ...

In the next chapter, we recapitulate this thesis and discuss research directions which come out of this thesis.

Chapter 6

Conclusion: Recapitulation and research directions

A thesis is never finished, it is deserted.

-Told to me by Patrick Kreidl, who was probably told this by someone else!

■ 6.1 In this chapter ...

In this chapter, we recapitulate what we have done in this thesis, and discuss research directions which come out of this thesis.

■ 6.2 Recapitulation

This thesis has two flavors:

1. Proving the optimality of communication in certain communication scenarios in the sense of optimality of source-channel separation theorem as stated in Reason 1c in Chapter 1
2. An operational view of rate-distortion theory

One contribution of this thesis is that we proved the optimality of digital communication in the sense of reason 1c in certain communication scenarios. A digital system is one where there is a finite interface (usually, a binary interface) between the source and the channel (point-to-point setting) or between the various sources and the medium (multi-user setting). Digital communication systems are defined precisely in Chapter 1. Optimality of digital communication is in the sense of reason 1c of Chapter 1.

In Chapter 2 we proved the optimality of digital communication in the point-to-point setting: a user wants to communicate a random source to another user over a channel. The channel is assumed to be only partially known in the sense that it might come from a family of transition probabilities. In other words, the setting is universal over the channel. We prove the following high-level statement, which is called the universal source-channel separation theorem for rate-distortion:

Assuming random-coding is permitted, in order to communicate a random source universally over a *partially known* channel to within a particular distortion level, it is sufficient to consider source-channel separation based architectures, that is, architectures which first code (compress) the random-source to within the particular distortion level, followed by universal reliable communication over the partially known channel. There is sufficiency in the sense if there exists some architecture to communicate the random source to within a certain distortion universally over the partially known, and which consumes certain amount of system resources like energy and bandwidth, then there exists a separation based architecture to universally communicate the random source to within the same distortion universally over the partially known channel, and which consumes the same or lesser system resources as the original architecture.

In order to prove precise results, we assume that the distortion measure is additive. We sketch an outline of how the results can be generalized to permutation invariant distortion meaures. Precise assumptions for both the high-level outline and rigorous proofs can be found in Chapter 2. In proving the result, we assume that the source statistics are known. In Chapter 2, we also conjecture that this result can be generalized to the case when the source statistics are not known.

The universal source-channel separation theorem for rate-distortion is generalized to the multi-user scenario in Chapter 3, in the unicast setting. By unicast setting, we mean that the sources that various users want to communicate to each other are independent of each other. Thus, if user i wants to communicate source X_{ij} to user j, the sources X_{ij} are all independent of each other. We prove the optimality of digital communication by proving that it is sufficient to consider architectures where each user compresses its sources to within the corresponding distortion levels and the compressed binary random sources are universally, reliably communicated over the partially known medium.

Chapter 4 discusses partial applicability of the results of the universal source-channel separation theorem for rate-distortion in the multi-user scenario to the wireless scenario. Two of the assumptions which are suspect are:

- The voice signals of various users are independent of each other. As discussed in Chapter 4, this assumptions is true pairwise, but the voice signals of the two users talking to each other are not independent of each other. We still assume that the distortion measure is permutation invariant.
- Distortion measure is permutation invariant: it is unclear, if this is the case for voice

Further discussion of the features of the wireless problem and assumptions that we make are discussed in Chapter 4. With the above assumptions, it follows that restricting attention to digital architectures is optimal for wireless communication, as discussed in Chapter 4.

As discussed in Chapter 1, there are various other factors which should determine which technology is used, and whether, a particular technological should be implemented in the first place or not. The only factor that we have tried to understand in the thesis beyond Chapter 1 is Reason 1c.

The other contribution of this thesis is an operational view of rate-distortion theory. By an operational view of rate-distortion theory, we mean a proof of the separation theorem which uses only the definition of channel capacity as the maximum rate of reliable communication and the rate-distortion function as the minimum rate needed to compress a source to within a certain distortion level. The proof, unlike Shannon's proof, does not use the definition of channel capacity as a maximum mutual information or the rate-distortion function as a minimum mutual information. We believe that our proof offers more insight into the nature of separation than Shannon's. This process is carried out on a high-level for the communication of the uniform X source in Section 2.14. A rigorous proof for the communication of the i.i.d. X source is the subject of Chapter 5. This also leads to connections between source and channel coding, and an alternate proof of the rate-distortion theorem for those i.i.d. X sources for which $p_X(x)$ is rational $\forall x \in \mathscr{X}$.

■ 6.3 Research directions

There are various research directions which come out of this thesis, some of which have been discussed in previous sections:

1. Section 4.3 discusses a two user abstraction of what happens when the sources are correlated. In general, separation does not hold. However, it would be interesting if one could get results for how good does the performance of separation based schemes approach the performance of a general analog scheme, possibly in the spirit of [TCDS]. Also, the problem makes sense not just for two users, but for N users where user i wants to communicate source X_{ij} to user j to within a distortion D_{ij} under a metric d_{ij}, and the distribution of the sources X_{ij}, $1 \le i, j \le N$ may be arbitrary

2. Various research directions come out of Chapter 2, some of which are discussed in Section 2.19:

   (a) We have proved the universal source-channel separation theorem for rate-distortion where universality is over the channel. We believe, as stated briefly in Section 2.19 that the result should be generalizable to universality over the source, that is, the source statistics are unknown.

   (b) Carry out a proof for arbitrary stationary ergodic sources evolving in continuous time, in the methodology of Section 2.14 and discussed further in Section 2.16. Section 2.14 carried out the proof (except for the equality of R_U^P under various definitions), and the full rigorous proof for the i.i.d. X source was carried out in Chapter 5. The operational proof in Chapter 5 is fairly technical and involved; we wonder if we are missing something and the proof can be simpler.

(c) To research, whether universal source-channel separation holds for sub-additive distortion measures, as it does in [Han10] when the setting is not universal.

3. We have used the probability of excess distortion criterion instead of the expected distortion criterion. As we have said in Chapter 2, universal source-channel separation does not hold with the expected distortion criterion. The example in Chapter 2 has a highly non-ergodic channel. Amos Lapidoth pointed out to us, in a note, that universal separation would hold with the expected distortion criterion for memoryless channels. He also provided us with a proof which should possibly be generalizable to indecomposable channels. The proof used the standard information theory machinery of mutual informations.

   We would like to take another route (if possible). The question is: if the expected distortion criterion holds, then does that imply that the probability of excess distortion criterion holds? A mathematical way of formulating this is the following: Let $k \in \mathscr{A}$ be a partially known channel such that each $k \in \mathscr{A}$ satisfies some ergodicity requirements. Suppose the set of channels k is capable of communicating i.i.d. X source to within a distortion D with error $\omega =< \omega^n >_1^\infty$, $\omega^n \to 0$ as $n \to \infty$ *under the expected distortion criterion.* That is, there exist encoder-decoder $< e^n, f^n >_1^\infty$ such that end to end

$$E\left[\frac{1}{n}d^n(X^n, Y^n)\right] \leq \omega^n \ \forall k \in \mathscr{A} \tag{6.1}$$

   Then, is k also capable of communicating i.i.d. X source to within a distortion D with some other error $\omega' =< \omega'^n >_1^\infty$, $\omega'^n \to 0$ as $n \to \infty$ under the expected distortion criterion? That is, do there exist encoder decoder $< e'^n, f'^n >_1^\infty$ such that end to end

$$\Pr\left(\frac{1}{n}d^n(X^n, Y^n) > D\right) < \omega'^n \ \forall k \in \mathscr{A}\ ? \tag{6.2}$$

   If this were the case, we would have reduced the problem with the expected distortion criterion to the problem with the probability of excess distortion criterion. One way to go about this would be to first assume that the set k consists of just one channel. In that case, our guess is that this is true. Then, one would want to carry out the procedure for a more general $\mathscr{A}$.

6.4 One final thought ...

Finally, I would like to add, to tie up with Chapter 1, that what is really needed is a more holistic view, a more human view of the problem, than just a mathematical or engineering understanding of whether separation is optimal or not. Without a human perspective, any technological or mathematical advances / understanding will only bring more negative consequences and misery to this world.

उद्धं, तिरियं अपाचिनं,
यावता जगतो गति
समवेक्खिता व धम्मानं
खन्धानं उदयब्बयं

-इतिवुत्तक ४ . १११

Above, across or back again,
wherever one goes in the world;
let one carefully scrutinize,
the rise and fall of compounded things.

- Itivuttaka 4.111

Bibliography

[AKM] M. Agarwal, S. Kopparty, and S. K. Mitter. The universal capacity of channels with given rate-distortion in the absence of common randomness. In *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing, 2009*, pages 700–707.

[CK97] I. Csiszár and J. Korner. *Information theory: coding theorems for discrete memoryless systems*. Akadémiai Kiadó, 1997.

[Gala] Robert G. Gallager. *Course materials for 6.450 Principles of Digital Communications I, Fall 2006*. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on August 4, 2011. Full URL: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-450-principles-of-digital-communications-i-fall-2006/lecture-notes/.

[Galb] Robert G. Gallager. *Video lectures for 6.450 Principles of Digital Communications I, Fall 2006*. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on August 4, 2011. Full URL: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-450-principles-of-digital-communications-i-fall-2006/video-lectures/.

[Gal68] Robert G. Gallager. *Information theory and reliable communication*. Wiley, January 1968.

[Gal08] Robert G. Gallager. *Principles of digital communications*. Cambridge University Press, 2008.

[Gas02] M. Gastpar. *To code or not to code*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2002.

[Han10] Te Sun Han. *Information-spectrum methods in information theory*. Springer, December 2010.

[Hay05] Steven C. Hayes. *Get out of your mind and into your life*. New Harbinger Publications, Inc., November 2005.

[Jr.a] G. David Forney Jr. *Lecture notes for 6.451 Principles of Digital Communications II, Spring 2005.* MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on August 4, 2011. Full URL: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-451-principles-of-digital-communication-ii-spring-2005/.

[Jr.b] G. David Forney Jr. *Video lectures for 6.451 Principles of Digital Communications II, Spring 2005.* MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on August 4, 2011. Full URL: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-451-principles-of-digital-communication-ii-spring-2005/.

[KEM] R. Koetter, M. Effros, and M. Medard. On a theory of network equivalence. In *Proceedings of the IEEE Information Theory Workshop on Networking and Information Theory, 2009*, pages 326 – 330.

[OPS48] B. M. Oliver, J. R. Pierce, and C. E. Shannon. The philosophy of PCM. *Proceedings of the Institute for Radio Engineers*, Volume 36, Issue 11: pages 1324–1331, November 1948.

[PG77] M. B. Pursley and R. M Gray. Source coding theorems for stationary, continuous-time stochastic processes. *Annals of Probability*, Volume 5, Number 6: pages 966–986, 1977.

[Ron11] John Ronson. *The psychopath test: a journey through the madness industry*. Penguin Group, May 2011.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, Volume 27: pages 379–423 (part 1) and pages 623–656 (part 2), July (part 1) and October (part 2) 1948. Can be found in *Claude Elwood Shannon Collected Papers*, edited by N. J. A. Sloane and Aaron D. Wyner, Wiley-IEEE Press, January 1993.

[Sha59] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *Institute of Radio Engineers, National Convention Record*, Volume 7, Part 4: pages 142–163, March 1959. Can be found in *Claude Elwood Shannon Collected Papers*, edited by N. J. A. Sloane and Aaron D. Wyner, Wiley-IEEE Press, January 1993.

[TCDS] C. Tian, J. Chen, S. Diggavi, and S. Shamai. Optimality and approximate optimality of source-channel separation in networks. *Submitted to IEEE Transactions on Information Theory, arXiv:1004.2648.*

[VH94] S. Verdu and T. S. Han. A general formula for channel capacity. *IEEE Transactions on Information Theory*, Volume 40, Issue 4: pages 1147–1157, July 1994.

[VVS95] S. Vembu, S. Verdu, and Y. Steinberg. The source-channel separation theorem revisited. *IEEE Transactions on Information Theory*, Volume 41, Issue 1: pages 44–54, January 1995.

[WHO] IARC classifies radiofrequency electromagnetic fields as possibly carcinogenic to humans. *World Health Organization/International Agency for Research on Cancer.* Press release. May 31, 2011.

[Wil89] Jan C. Willems. Models for dynamics. *Dynamics Reported*, Volume 2: pages 171–269, April 1989.

[Wil07] Jan C. Willems. The behavioral approach to open and interconnected systems. *IEEE Control Systems Magazine*, Volume 27: pages 46–99, December 2007.

[Ziv72] J. Ziv. Coding of sources with unknown statistics - II: Distortion relative to a fidelity criterion. *IEEE Transactions on Information Theory*, Volume 18, Issue 4: pages 389–394, May 1972.

[Ziv80] J. Ziv. Distortion-rate theory for individual sequences. *IEEE Transactions on Information Theory*, Volume 26, Issue 2: pages 137–143, March 1980.