Channel Coding in the Presence of Side Information

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

Channel Coding in the Presence of Side Information

Guy Keshet

TraceSpan Communications Ltd.
Ra'anana 43663, Israel
guy@tracespan.com

Yossef Steinberg

Technion — Israel Institute of Technology
Haifa 32000, Israel
ysteinbe@ee.technion.ac.il

Neri Merhav

Technion — Israel Institute of Technology
Haifa 32000, Israel
merhav@ee.technion.ac.il

now

the essence of knowledge

Boston – Delft

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is G. Keshet, Y. Steinberg and N. Merhav, Channel Coding in the Presence of Side Information, Foundations and Trends® in Communications and Information Theory, vol 4, no 6, pp 445–586, 2007

ISBN: 978-1-60198-048-9
© 2008 G. Keshet, Y. Steinberg and N. Merhav

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

Foundations and Trends® in Communications and Information Theory
Volume 4 Issue 6, 2007

Editorial Board

Editor-in-Chief:
Sergio Verdú
Depart of Electrical Engineering
Princeton University
Princeton, New Jersey 08544

Editors

Venkat Anantharam (UC. Berkeley)
Ezio Biglieri (U. Torino)
Giuseppe Caire (U. Sounthern California)
Roger Cheng (U. Hong Kong)
K.C. Chen (Taipei)
Daniel Costello (U. Notre Dame)
Thomas Cover (Stanford)
Anthony Ephremides (U. Maryland)
Andrea Goldsmith (Stanford)
Dave Forney (MIT)
Georgios Giannakis (U. Minnesota)
Joachim Hagenauer (TU Munich)
Te Sun Han (Tokyo)
Babak Hassibi (Caltech)
Michael Honig (Northwestern)
Johannes Huber (Erlangen)
Hideki Imai (Tokyo)
Rodney Kennedy (Canberra)
Sanjeev Kulkarni (Princeton)
Amos Lapidoth (ETH Zurich)
Bob McEliece (Caltech)
Neri Merhav (Technion)
David Neuhoff (U. Michigan)
Alon Orlitsky (UC. San Diego)
Vincent Poor (Princeton)
Kannan Ramchandran (UC. Berkeley)
Bixio Rimoldi (EPFL)
Shlomo Shamai (Technion)
Amin Shokrollahi (EPFL)
Gadiel Seroussi (MSRI)
Wojciech Szpankowski (Purdue)
Vahid Tarokh (Harvard)
David Tse (UC. Berkeley)
Ruediger Urbanke (EPFL)
Steve Wicker (Cornell)
Raymond Yeung (Hong Kong)
Bin Yu (UC. Berkeley)

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

Editorial Scope

Foundations and Trends® in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians
Foundations and Trends® in Communications and Information Theory, 2007, Volume 4, 6 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

Foundations and Trends® in
Communications and Information Theory
Vol. 4, No. 6 (2007) 445–586
© 2008 G. Keshet, Y. Steinberg and N. Merhav
DOI: 10.1561/0100000025

now
the essence of knowledge

Channel Coding in the Presence of Side Information

Guy Keshet[1], Yossef Steinberg[2] and Neri Merhav[3]

[1] *TraceSpan Communications Ltd., P.O. Box 2444, 10 Hayetzira St., Ra'anana 43663, Israel, guy@tracespan.com*
[2] *Department of Electrical Engineering, Technion — Israel Institute of Technology, Haifa 32000, Israel, ysteinbe@ee.technion.ac.il*
[3] *Department of Electrical Engineering, Technion — Israel Institute of Technology, Haifa 32000, Israel, merhav@ee.technion.ac.il*

Abstract

In this survey we review the concepts and methods of communication systems equipped with side information. We focus on the channel coding problem, where side information is available to the transmitter in either a causal or non-causal manner, and we also consider the source coding problem with side information at the receiver.

We first summarize the main results for channels with causal/non-causal side information and the associated capacity formulas. Next, we consider specific channel models, such as Costa's dirty-paper model, the AWGN channel model with fading and the modulo additive noise channel. Further, we provide applications to the models considered here, in particular, we present the watermarking problem and the Gaussian MIMO broadcast channel. We also consider algorithms for the calculation of the channel's capacity, and practical coding schemes for the communication systems explored in this survey. Finally, we study several

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

related information-theoretic problems and present both the Wyner–Ziv and the Slepian–Wolf problems. The source coding problems and the channel coding problems, are presented in a unified version and the duality between the problems is presented. We also present extensions for the MAC and broadcast channel models, to the case where they are controlled by a state process, and consider several hybrid models, e.g., joint source–channel coding for the Wyner–Ziv source and the Gel'fand–Pinsker channel, and the achievable tradeoff between the message and the state information rates.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

Contents

1 Introduction **1**

2 Notation and Model Formulation **7**

3 Fundamental Results **11**

3.1 Causal Side Information 11
3.2 Non-Causal Side Information 18
3.3 Relations between the Causal and Non-Causal Cases 21
3.4 Modifications and Extensions 24
3.5 Rate-Limited Side Information 29

4 Specific Channel Models **37**

4.1 The Dirty-Paper Channel 37
4.2 The AWGN Channel with Fading 42
4.3 Modulo Additive Noise Channels 45

5 Applications **51**

5.1 The Gaussian Vector Broadcast Channel 51
5.2 Watermarking 56

6 Other Related Problems **69**

6.1 The Slepian–Wolf Problem 70
6.2 The Wyner–Ziv Problem 72

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

6.3 Source Coding and Channel Coding with Side Information 75
6.4 Joint Source–Channel Coding for the Wyner–Ziv Source 82
6.5 Channel Capacity and State Estimation 85
6.6 Multiple User Channels with Side Information 92

7 Algorithms 111

7.1 Numerical Computation of the Channel Capacity 111
7.2 Coding Schemes 117

8 Conclusion 133

References 135

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

1

Introduction

This survey gives an overview of results pertaining to the capacity of a channel whose conditional output probability distribution depends on a state process, and where the channel state information (CSI) signal (also referred to as "side information") is available at the transmitter (CSIT) or at the receiver (CSIR) or at both ends. These channels have been widely studied over the years and can serve for modeling in a wide range of problems, depending on some assumptions regarding the channel state and on the availability and quality (clean or noisy) of the side information at the transmitter and/or the receiver. For CSI available at the transmitter, we will distinguish between channels where the CSIT is causal and channels where it is non-causal. In the causal case, the transmission, at every time instant, depends only on the past and present CSI, whereas in the non-causal case, the transmitter knows in advance the realization of the entire state sequence from the beginning to the end of the block. The causal CSIT channel model was introduced in 1958 by Shannon [96], who also found its capacity. The non-causal CSIT channel model was introduced in 1974 by Kusnetsov

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

and Tysbakov [73], and its capacity was found in 1980 by Gel'fand and Pinsker [54]. Regarding CSIR, we will not distinguish between causal and non-causal CSI at the receiver, since the receiver waits until the end of the block anyway, before decoding. Channels with CSIR are studied in [12, 60, 93].

One interesting example which can be modeled as a channel with non-causal transmitter CSI, is a computer memory with defective cells [59, 60, 71, 72, 73, 112, 113]. In this example, the process of storing to this memory suffers from random errors caused by noise. A computer memory may also have some cells which are defected, e.g., a memory cell whose stored value seems to be fixed regardless of the input, e.g., the cell is "stuck at one" or "stuck at zero." The location and values (defect information) of the defective cells may be found by storing the all-one bit pattern (or the all-zero bit pattern) in the memory, reading the contents of the memory and comparing it with the stored pattern. This process may be repeated several times in order to exclude the effect of random errors. If this process is not repeated, we can refer to the defect information received from this process as a noisy version of the defect information. By knowing in advance the defect information, we can design codes that are more efficient than the usual error correcting codes. In this example, the defect information plays the role of the channel states and transmitter's CSI.

Another important example for a channel model with transmitter CSI, is the power constrained Gaussian additive noise channel model with additive interference which is known non-causally to the transmitter as side information and is statistically independent of the noise. In this example, the channel output is given by

$$Y_i = X_i + S_i + Z_i, \quad i = 1, 2, \ldots, N, \tag{1.1}$$

where $S^N = (S_1, \ldots, S_N)$ is the i.i.d. side information sequence (the interference) distributed as $S^N \sim \mathcal{N}(0, QI)$ (I being the identity matrix), Q is the side information variance, and $Z^N = (Z_1, \ldots, Z_N)$ is the i.i.d. noise sequence distributed as $Z^N \sim \mathcal{N}(0, BI)$, B being the noise variance. Based on the message to be sent and on the interference samples $S_1, \ldots, S_N$, the encoder sends a codeword $X^N = (X_1, \ldots, X_N)$,

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

which must satisfy the power constraint

$$\frac{1}{N}\sum_{n=1}^{N} E(X_n^2) \leq \Gamma, \tag{1.2}$$

where $\Gamma > 0$ is a given constant. This channel with non-causal CSIT, is known as Costa's channel [30], which has recently received much attention, as it has been proven useful for modeling in various communication problems, among them precoding for intersymbol interference (ISI) channel, digital watermarking, and various broadcasting schemes. Again, this problem can be divided to the causal and non-causal CSIT scenarios. When the CSIT is non-causal, this problem is known as writing on dirty paper (WDP), or the dirty-paper problem. When the CSIT is causal, the problem is known as writing on dirty tape (WDT), or the dirty-tape problem [14]. The dirty-tape setting is often used to describe cases where we restrict the encoder to be causal in order to reduce the implementation complexity compared to dirty-paper implementation.

Costa showed, for the dirty-paper setting, that the capacity of this channel is the same as if the interference was not present or, equivalently, if it was also known at the decoder and could be subtracted off, i.e., the capacity is given by

$$C = \frac{1}{2}\log\left(1 + \frac{\Gamma}{B}\right). \tag{1.3}$$

This surprising result is another reason why this problem has received so much attention.

Yet another interesting example for a channel model with transmitter CSI, is digital watermarking [19, 21, 20, 22, 23, 28, 36, 43, 44, 45, 75, 76, 78, 83, 86, 100, 99]. Digital watermarking is the process of embedding a message within a host signal to form a composite (watermarked) signal. The embedding must not cause a noticeable distortion relative to the host signal. On the other hand, the embedding should be robust to attacks on the watermarked signal. In some applications, these attacks are the result of standard signal processing operations. In other cases they are malicious. The digital watermarking problem can be modeled as a channel whose conditional output probability depends on a state process, and where the transmitter has channel state information. Here the input power constraint is replaced by a constraint on the

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

distortion between the channel input and the host signal. This expresses the requirement that the embedding does not cause a noticeable distortion to the host signal. The dirty-paper problem may be used to model watermarking in a Gaussian environment. We can model watermarking as a communication system in which the transmitter, which must satisfy a distortion constraint with regard to the host signal, sends a watermark message through a noisy channel with the host signal playing the role of state. The state (the host signal), is available to the transmitter, and therefore it can be used by the transmitter as side information, just like in the dirty-paper problem.

The purpose of this survey is to give an overview of the subject of coding for channels with side information, both from the theoretical point of view (capacity, fundamental limits, duality with source coding), and the practical point of view, including coding and decoding techniques, structures of classes of codes, application aspects, etc. A fundamental ingredient in the solutions of the channel coding problems presented in this survey, and the dual source coding problems (e.g., the Slepian–Wolf and the Wyner–Ziv problems [98, 124]), is the "binning" concept. In the context of channel coding with side information, a binning scheme divides a set of codewords into subsets or "bins," such that the codewords in each bin are as far apart as possible. These bins are constructed at random and each bin is assigned to a different message index. When given with a message to be transmitted in a channel coding scenario, we use only the codewords from the bin with the same index as our message, and we choose a codeword which is the closest to (or jointly typical with) the side information vector. Since the messages to be transmitted over the channel are presented by bin indices, it is desired that codewords in each bin are as far apart as possible, to cause minimal rate loss, but not too far, so that every realization of the side information will have a "matching" codeword in the bin. Therefore, in channel coding applications, the codewords in each bin should constitute a good source code, where here the "source" is the channel state. In source coding problems, a codeword that is jointly typical with the source sequence is found, among all the codewords, and the bin index in which it resides is sent to the decoder. The decoder looks in this particular bin for a codeword that is jointly typical with the

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

side information. The codewords in each bin should be close to each other — to reduce the overall rate — but not too close, so that only one codeword in the particular bin will be jointly typical with the side information. Therefore, in the context of source coding, the codewords in each bin should constitute a good channel code.

The outline of this survey is as follows. In Section 2, we specify the notation conventions that will be used and we formalize the model of a channel with CSIT. Section 3 describes the main theoretical results pertaining to the capacity of channels with side information. In Section 4, we describe some specific channel models for various problems. In Section 5, we present several applications for the models which were presented in the preceding sections. Section 6 presents related problems which are linked to the problem of coding for a channel with side information. These problems include source coding dual, which is the Wyner–Ziv problem, the Gaussian vector broadcast channel and multiuser channels. In Section 7, we describe algorithms for computation of the capacity of channels that were presented in the previous sections, as well as several coding techniques for these channels. Section 8 concludes the survey.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

Documents
Document Sharing Hub

Full document details: IEEE Channel Coding in the Presence of Side Information

2

Notation and Model Formulation

Throughout this survey, scalar random variables will be denoted by capital letters, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and their sample values, which will be denoted with the same symbols superscripted by the dimension. Thus, for example, W^k will denote a random k-vector $(W_1, \ldots, W_k)$, and $w^k = (w_1, \ldots, w_k)$ will be a specific vector value in $\mathcal{W}^k$, the kth Cartesian power of $\mathcal{W}$. The notations w_i^j and W_i^j, where i and j are integers and $i \leq j$, will designate segments $(w_i, \ldots, w_j)$ and $(W_i, \ldots, W_j)$, respectively, where for $i = 1$, the subscript will be omitted (as above). For $i > j$, w_i^j (or W_i^j) will be understood as the null string. Sequences without specifying indices are denoted by $\{\cdot\}$. The cardinality of a set $\mathcal{W}$ will be denoted by $|\mathcal{W}|$. The empty set will be denoted by $\emptyset$. $\mathbb{R}$ stands for the set of real numbers, $\mathbb{C}$ for the set of complex numbers, and $\mathbb{R}_+$ for the set of positive reals.

Sources and channels will be denoted generically by the letter P subscripted by the name of the random variable (RV) and its conditioning, if applicable, e.g., $P_U(u)$ is the probability function of U at the point $U = u$, $P_{Z|S}(z|s)$ is the conditional probability of $Z = z$ given $S = s$,

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

and so on. Whenever clear from the context, these subscripts will be omitted. A sequence of finite-dimensional distributions will be denoted by bold letters, for example, $\mathbf{\Lambda} = \{P_{W^N}(w^N)\}_{N=1}^{\infty}$. $U(a,b)$ stands for the uniform distribution on the interval (a,b). Information theoretic quantities like entropies, divergences, and mutual informations will be denoted following the usual conventions of the information theory literature, e.g., $H(U^N)$, $I(Z^n;W^k)$, $D(P_{Y|XS}\|P_Y)$, and so on.

We consider the channel depicted in Figure 2.1, whose input, state, and output, at time index n, are $X_n \in \mathcal{X}$, $S_n \in \mathcal{S}$, and $Y_n \in \mathcal{Y}$, respectively, where $\mathcal{X},\mathcal{Y},\mathcal{S}$ are the corresponding alphabets. The general results given in this survey apply to finite alphabets, but a number of extensions are described that apply to, for example, continuous alphabet and input constraints, etc. Unless otherwise specified, we assume throughout that the channel and state are memoryless, i.e.,

$$P_{Y^N|X^N,S^N}(y^N|x^N,s^N) = \prod_{n=1}^{N} P_{Y|X,S}(y_n|x_n,s_n). \tag{2.1}$$

$$P_{S^N}(s^N) = \prod_{n=1}^{N} P_S(s_n). \tag{2.2}$$

Many variants of this channel model have been studied. In some of these variants, the state sequence is generated in a different manner. The first example is the compound channel, in which the channel

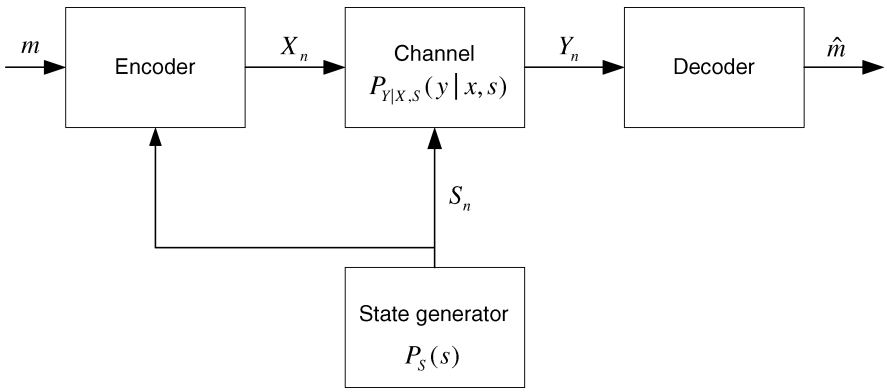

Fig. 2.1 CSIT channel model.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

state is fixed during the course of transmission. Another example is the arbitrary varying channel (AVC), in which, the states vary arbitrarily from symbol to symbol during the transmission (the states are not generated stochastically, but in an arbitrary deterministic manner). Usually, in AVC channel models, the state sequence is unknown to the transmitter [74]. Ahlswede [1] analyzed the case where this sequence is known to the transmitter. Winshtok and Steinberg [121] considered the problem of joint source–channel coding for the arbitrary varying source transmitted over an AVC where the state sequence is known to the transmitter. This problem combines the model considered by Ahlswede for an AVC and the problem of an arbitrary varying Wyner–Ziv source. Other examples include channels in which the state sequence is a more general stochastic process (e.g., the Gilbert–Elliot channel) or some hybrid situations like a compound Gilbert–Elliot channel. More details on many of these channel models can be found in [74].

The channel input may be subjected to a transmission-cost constraint

$$E\left\{\sum_{i=1}^{N}\phi(X_i)\right\} \leq N\Gamma, \tag{2.3}$$

where ϕ is a given function from $\mathcal{X}$ to $\mathbb{R}^+$ and $\Gamma \geq 0$ is a prescribed constant. In this problem, the transmitter sends a random message M to the receiver in N uses of the channel. The random message M is uniformly distributed in $\mathcal{M} = \{1, \ldots, 2^{NR}\}$, R being the code rate. The transmitter is also provided with causal or non-causal channel state information (CSI).

An $(N, 2^{NR})$ code consists of the following:

1. A set of N encoding functions $f_n : \mathcal{M} \times \mathcal{S}^n \to \mathcal{X}$, for $n = 1, \ldots, N$, such that $x_n = f_n(m, s^n)$ and Equation (2.3) is satisfied, where m ranges over $\mathcal{M}$, and s^n is the vector of the channel states realizations up to time n. For the non-causal case, a mapping $f : \mathcal{M} \times \mathcal{S}^N \to \mathcal{X}^N$, such that $x^N = f(m, s^N)$, or equivalently, $x_n = f_n(m, s^N)$, $n = 1, \ldots, N$, and again Equation (2.3) is satisfied.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

2. A decoding function $g_n : \mathcal{Y}^N \to \mathcal{M}$, such that the decoded message is $\hat{m} = g(y^N)$.
   The average probability of error is given by

$$P_e = \frac{1}{2^{NR}} \sum_{m=1}^{2^{NR}} \sum_{\{y^N : g(y^N) \neq m\}} \sum_{s^N} P_{S^N}(s^N) P_{Y^N|X^N,S^N}(y^N|x^N, s^N), \tag{2.4}$$

   where x^N depends on the inputs to the encoder in a manner that depends on whether it is the causal or non-causal case, as described above.

We define an achievable rate and capacity in the following way:

Definition 2.1. A rate R is said to be achievable if for every $\epsilon > 0$, there exists an integer $N_0 = N_0(\epsilon)$ such that for all $N > N_0$, there exists an $(N, 2^{NR})$ code with probability of error $P_e \leq \epsilon$.

Definition 2.2. The capacity is the supremum of all achievable rates.

This model, definitions, and notations will be used throughout the survey. Next, we will present the fundamental results in this survey.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

3

Fundamental Results

3.1 Causal Side Information

The causal side information case was first investigated by Shannon [96] in 1958. Shannon showed that the capacity of the channel, depicted in Figure 2.1, is equal to the capacity of an ordinary discrete memoryless channel (DMC), with the same output alphabet and an input alphabet of size $|\mathcal{X}|^{|\mathcal{S}|}$. The input letters of the new channel consist of all mappings from $\mathcal{S}$ to $\mathcal{X}$. Any coded communication system for the equivalent channel, without side information, can be translated into a coded communication system for the original channel with the same probability of error. As we shall see later, the converse implies also the other direction, and hence from this point on we will refer to these two channels as equivalent. The equivalent channel and its relation to the original channel are depicted in Figure 3.1.

The equivalent channel input variable is denoted by T, and is called *strategy*. Every input letter t (*strategy*) of this channel is a particular function from the state alphabet $\mathcal{S}$ to the input alphabet $\mathcal{X}$ of the original channel (i.e., $t : \mathcal{S} \to \mathcal{X}$). The alphabet of T, which will be denoted by $\mathcal{T}$, consists of all $|\mathcal{X}|^{|\mathcal{S}|}$ distinct functions from the state

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

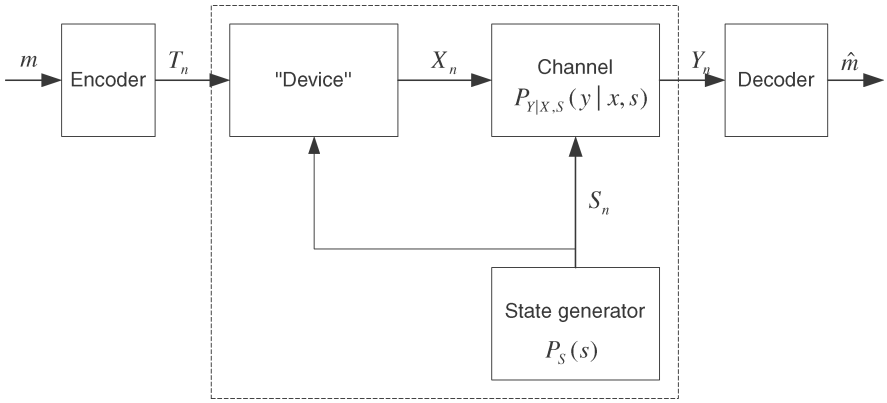

Fig. 3.1 Shannon's equivalent channel.

set to the input alphabet. To achieve capacity, it is enough to use at most $|\mathcal{Y}|$ of them, which is the number of output letters (see [35] Section 8.3). The "device" shown in Figure 3.1 takes T_n and S_n as the inputs and produces $X_n = T_n(S_n)$ for a function T_n and state S_n. The equivalent channel is characterized by a set of conditional output probability functions $\{P_{Y|T}(y|t) : t \in \mathcal{T}, y \in \mathcal{Y}\}$ such that

$$P_{Y^N|T^N}(y^N|t^N) = \prod_{n=1}^{N} P_{Y|T}(y_n|t_n), \tag{3.1}$$

where

$$P_{Y|T}(y|t) = \sum_{s \in \mathcal{S}} P_S(s) P_{Y|X,S}(y|t(s), s). \tag{3.2}$$

Shannon used strategies in which the input to the channel depends only on the current state of the channel and not on previous states. He showed that using only this type of strategies, the capacity of the equivalent channel is equal to the capacity of the original channel. The capacity of this channel is therefore the capacity of the equivalent DMC

$$C = \max_{P_T(\cdot)} I(T;Y), \tag{3.3}$$

where $P_T(\cdot)$ is a probability distribution of T, which is independent of the state S and the maximization is taken over all joint distributions

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

satisfying

$$P_{T,S,X,Y}(t,s,x,y) = P_S(s)P_T(t)\delta(x,t(s))P_{Y|X,S}(y|x,s), \tag{3.4}$$

where $\delta(x,t(s)) = 1$ for $x = t(s)$, and $\delta(x,t(s)) = 0$ otherwise.

We can rewrite the capacity formula for the causal case (3.3), and present it in terms of auxiliary random variable (as it is being presented in some papers):

$$C = \max_{P_U(\cdot), f:\mathcal{U}\times\mathcal{S}\to\mathcal{X}} I(U;Y), \tag{3.5}$$

where U is an auxiliary random variable and the joint distributions of the random variables S,U,X,Y is given by

$$P_{S,U,X,Y}(s,u,x,y) = P_S(s)P_U(u)\delta(x,f(u,s))P_{Y|X,S}(y|x,s). \tag{3.6}$$

We can think of this auxiliary random variable as indexing the set of all functions from $\mathcal{S}$ to $\mathcal{X}$. There are $|\mathcal{X}|^{|\mathcal{S}|}$ possible functions, but we only need to use at most $|\mathcal{Y}|$ in the causal case.

In order to prove the equivalence of the two channels, we first note that from (3.2), it can easily be seen that each code for the equivalent channel can be translated to a code for the original channel, and vice versa, with the same probability of error. The translation of codes consists merely of using the input $x_n = t_n(s_n)$ for a function t_n and state s_n. Next, we prove that the rate of the original channel cannot be larger than the capacity of the equivalent channel, thus establishing the converse part of (3.3). Let $U_n = (M, S^{n-1})$, and let C be given by (3.5), we can write:

$$\begin{aligned} NR - H(M|Y^N) &= H(M) - H(M|Y^N) \\ &= I(M;Y^N) \\ &= \sum_{n=1}^{N} I(M;Y_n|Y^{n-1}) \\ &\le \sum_{n=1}^{N} I(M,Y^{n-1};Y_n) \end{aligned}$$

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

$$\begin{aligned} &\leq \sum_{n=1}^{N} I(M, S^{n-1}; Y_n) \\ &= \sum_{n=1}^{N} I(U_n; Y_n) \\ &\leq NC, \end{aligned} \tag{3.7}$$

where the second inequality follows from the data processing inequality using the Markov chain $(M, Y^{n-1}) \to (M, S^{n-1}) \to Y_n$ and the last inequality follows since U_n is independent of S_n. If the message is recovered reliably, then $H(M|Y^N)$ must be small (Fano's inequality). Thus, the rate R cannot be larger than C.

The capacity in Equation (3.3) is expressed in terms of strategies. This might pose some conceptual and practical problems for code construction, especially for large $|\mathcal{S}|$. Large alphabets make the implementation of the encoder and decoder more difficult.

The encoder, in this problem, is able to produce the channel input X_n, when given with $U_n = (M, S^{n-1})$ and S_n, as one should expect in a causal CSIT model.

Shannon's capacity formula can be extended [5] to the case where the alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ are the real line and where the transmitter is subjected to an average power constraint

$$E\{\phi(X)\} \leq \Gamma, \tag{3.8}$$

and the capacity is given by

$$C = \sup_{P_T(\cdot): E\{\phi(T(S))\} \leq \Gamma} I(T; Y), \tag{3.9}$$

where the expectation is relative to the product distribution $P_{S,T}(s,t) = P_S(s)P_T(t)$, $t \in \mathcal{T}$, $s \in \mathcal{S}$. In this case, the supremum in (3.9) is over all the distributions $P_T(\cdot)$ of the continuous functions $t : \mathcal{S} \to \mathcal{X}$, where the alphabets $\mathcal{X}, \mathcal{S}$ are the real line.

Shannon's results have been used to compute the capacity of a discrete modulo-additive noise channel [50] with causal side information at the transmitter, which represents a specific channel whose capacity was determined using (3.3). These results were also used to

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

bound the capacity for the dirty-tape problem [49], which is the causal counterpart of the dirty-paper problem. These problems will be discussed in Section 4.

Shannon's model was extended by Salehi [93] to the case where the transmitter has access to one noisy version of the state information, W_n, and the receiver has access to another version, V_n. We will see later that this model is not really more general than Shannon's model. This communication system is depicted in Figure 3.2.

Denote by $\mathcal{W}$ and $\mathcal{V}$ the alphabets of W_n and V_n, respectively, and let $\mathcal{T}$ be the set of all $|\mathcal{X}|^{|\mathcal{W}|}$ possible functions from $\mathcal{W}$ to $\mathcal{X}$. The variables T_n, S_n, W_n, V_n, and Y_n, $n = 1, \ldots, N$, take values according to the joint distribution

$$P_{T,S,W,V,Y}(t,s,w,v,y) = P_T(t)P_{S,W,V}(s,w,v)P_{Y|X,S}(y|x = t(w),s), \tag{3.10}$$

for some arbitrary distribution $P_T(t)$, where T is independent of W and V. The capacity of the channel is given by

$$C = \max I(T;Y|V), \tag{3.11}$$

where the maximum is taken over all joint probability distributions as in (3.10).

Salehi proved his capacity formula (3.11) directly in [93]. Caire and Shamai [12] showed that this result follows from Shannon's result

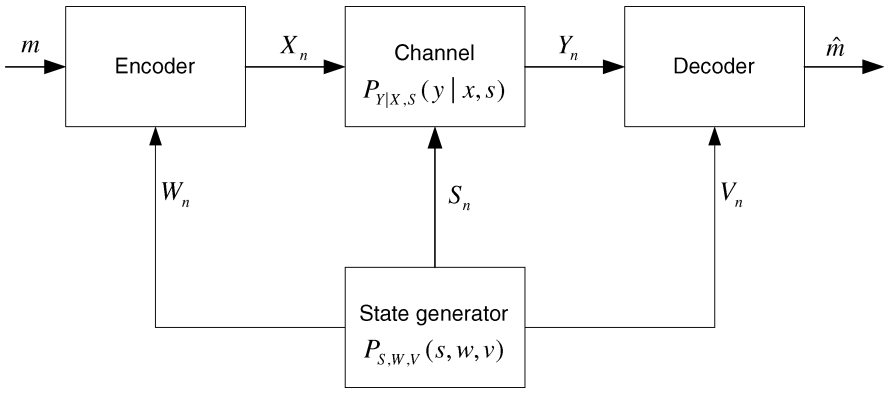

Fig. 3.2 Transmitter and receiver side information channel model.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

for the capacity of a channel with perfect causal transmitter side information by a simple argument: They considered the receiver side information $\{V_n\}$ as an additional channel output in an equivalent channel. For this channel, the conditional probability distribution of the output (Y,V) is given by

$$\begin{aligned} P'_{Y,V|X,W}(y,v|x,w) &= \sum_s P_{Y,V|X,W,S}(y,v|x,w,s)P_{S|X,W}(s|x,w) \\ &= \sum_s P_{Y|X,S}(y|x,s)P_{S,W,V}(s,w,v)/P_W(w), \quad (3.12) \end{aligned}$$

and $P_W(w) = \sum_{s,v} P_{S,V,W}(s,v,w)$.

This equivalent channel is clearly of the type studied by Shannon, with causal transmitter side information and no receiver side information. The capacity of this channel is therefore $C = \max_{P_T(t)} I(T;Y,V)$. But since V is independent of T we have $I(T;V) = 0$, so that (3.11) follows immediately. It follows then that, as argued earlier, Salehi's model is not really more general than Shannon's original model.

The capacity of this channel is expressed in terms of strategies. We mentioned earlier that this might pose some conceptual and practical problems for code construction, especially for large $|\mathcal{W}|$, that will require very large codebooks. Caire and Shamai [12] found a special case in which the capacity can be expressed without using strategies. In this case, the CSIT is a deterministic function of the CSIR. Let $W_n = g(V_n)$, where $g(\cdot)$ is a deterministic function from $\mathcal{V}$ to $\mathcal{W}$. Then, the channel capacity is given by

$$\begin{aligned} C &= \sum_{w\in\mathcal{W}} p(w) \max_{p(x|w)} I(X;Y|V,W=w) \\ &= \max I(X;Y|V,W). \quad (3.13) \end{aligned}$$

This capacity can be achieved by a multiplexed multiple codebook scheme. For each value of $w \in \mathcal{W}$, a codebook of length $(p(w)-\delta)N$, where δ is a small positive number, and rate slightly less than $I(X;Y|V,w)$ is generated i.i.d. according to the probability distribution $p(x|w)$. For the message m, a set of $|\mathcal{W}|$ codewords is selected, one for each codebook. At time index n, if $W_n = w$, the transmitter sends the first not yet transmitted symbol of the wth codeword. Then the

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

codewords are multiplexed according to the transmitter side information sequence W^N. If $g(\cdot)$ is deterministic, the receiver can demultiplex the received sequence before decoding since it can perfectly recover W^N from V^N. After demultiplexing, the $|\mathcal{W}|$ codewords are independently decoded. The case where W_n is a deterministic function of V_n, can describe a scenario in which the transmitter's side information is obtained via an error-free low rate causal feedback channel from the receiver to the transmitter. Heegard and El Gamal [60] studied a related problem, in which the CSIT and CSIR are subject to rate constraints R_e and R_d, respectively. We will examine this problem is Section 3.2.

A special case of Caire and Shamai's result, is the case where the side information available to the encoder is equal to the side information available to the decoder, i.e., $W_n = V_n$. We will then denote this as Z_n, i.e., $W_n \triangleq V_n = Z_n$. In this special case, $\mathcal{T}$ denotes the set of all possible functions from $\mathcal{Z}$ to $\mathcal{X}$. The capacity in this case is given by

$$C = \max I(X;Y|Z), \tag{3.14}$$

where the maximum is taken over all joint probability distributions of the form:

$$P_{S,Z,X,Y}(s,z,x,y) = P_{S,Z}(s,z)P_{X|Z}(x|z)P_{Y|X,S}(y|x,s). \tag{3.15}$$

The case where the same side information is available to both the encoder and decoder has been used in [5] to model *private* information embedding which has a wide variety of applications, such as digital watermarking, data hiding, and steganography. Information embedding can be viewed as a problem of channel coding with side information. We refer to the information embedding case where both the encoder and decoder have the same side information signal as *private* information embedding, and to the case where only the encoder has side information as *public* information embedding.

We mentioned, in Section 2, that we assume that the states of the channel are generated by a memoryless source of a given distribution (2.2). In other models, the state sequence is not memoryless. We will elaborate on this in Section 6.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

3.2 Non-Causal Side Information

In this section, we consider the case of non-causal side information. This model was first investigated by Kuznetsov and Tsybakov [73] in 1974, who considered the problem of coding for a computer memory with defective cells. Here, the positions of the defective cells and their type, e.g., stuck at zero or stuck at one cells, serve as the channel state information, and they are known non-causally to the encoder.

Kuznetsov and Tsybakov presented coding techniques for this channel, but they have not determined the capacity. The capacity was found in 1980 by Gel'fand and Pinsker [54] and also by Heegard and El Gamal [60]. As in Section 2, the channel is memoryless and the output probability is given by (2.1). The capacity of this channel is given by

$$C = \max_{P_{U,X|S}} [I(U;Y) - I(U;S)], \tag{3.16}$$

where U is an auxiliary random variable with cardinality $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}| + 1$, and the maximization is subjected to the constraint that $U \to (X,S) \to Y$ forms a Markov chain. The concept of auxiliary RV was first used by Cover [31] for the broadcast channel.

Denoting $R_{GP}(P_{U,X|S}) = I(U;Y) - I(U;S)$, this function has the following two properties:

Lemma 3.1 ([54]). (i) For fixed $P_{U|S}(u|s)$, $R_{GP}(P_{U,X|S})$ is a convex functional of the distribution $P_{X|U,S}(\cdot|u,s)$. (ii) For fixed $P_{X|U,S}$, $R_{GP}(P_{U,X|S})$ is a concave functional of the distribution $P_{U|S}(\cdot|s)$.

By Lemma 3.1, the optimum $P_{X|U,S}(\cdot|u,s)$ (in the sense of (3.16)) puts all its mass on a single x for every u,s, and so, there exists a deterministic mapping $f : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ such that $P_{X|U,S}(x|u,s) = 1$ if and only if $x = f(u,s)$. This remains true [27], even when the transmitter is subjected to an average power constraint (2.3) as we shall prove in Section 3.4.

We will now outline the proof of the direct part of this capacity formula. To this end, we will need to use a random binning technique, which is also being used in the dual problem of source coding with side information [35]. For each message m, generate 2^{NR_0}

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

codewords (forming a bin) $\{u^N(m,1),\ldots,u^N(m,2^{NR_0})\}$[1] i.i.d. according to the distribution $P_U(u)$. Given the message m and the state sequence s^N, the encoder seeks a codeword in bin m that is jointly typical with s^N, say $u^N(m,j)$. If multiple such codewords in bin m exist, choose the one with the smallest j. If no such j exists, then declare an encoding error. The encoder then creates the input to the channel as $x_i = f(u_i(m,j),s_i)$, $i=1,\ldots,N$. The decoder finds an $\hat{m}$ and a $\hat{j}$ such that $u^N(\hat{m},\hat{j})$ is jointly typical with the channel output sequence y^N, the decoded message in this case will be $\hat{m}$. If multiple or no such pair $\left(\hat{m},\hat{j}\right)$ exists, then declare a decoding error. The probability of encoding failure goes to zero as long as $R_0 > I(U;S)$ and the probability of decoding failure goes to zero as long as $R + R_0 < I(U;Y)$. Thus, the overall probability of error goes to zero as long as $R < I(U;Y) - I(U;S)$.

The above encoding procedure which finds a jointly typical codeword $u^N(m,j)$ with s^N in bin m, defines a source coding process [35], thus the collection of codewords in each bin plays the role of a source code. The above decoding procedure, finds a codeword from the entire collection of codewords which is jointly typical with y^N, thus the collection of all the codeword in all the bins plays the role of a channel code. We will make use of these observations later on.

Next, we will outline the proof of the converse part of the Gel'fand–Pinsker capacity formula (3.16). We will prove that every code $(N,2^{NR})$ with error probability $P_e \leq \epsilon$ for this channel satisfies $R \leq C$. From Fano's inequality,

$$NR - H(M|Y^N) \leq I(M;Y^N), \tag{3.17}$$

and since the message M is independent of the state sequence S^N, $I(M;S^N) = 0$, we can rewrite (3.17) as

$$NR - H(M|Y^N) \leq I(M;Y^N) - I(M;S^N). \tag{3.18}$$

[1] $u^N(m,j)$ represents a codeword indexed by $m \in \mathcal{M}$, which is the bin index and $j \in \{1,\ldots,2^{NR_0}\}$, which is an index within the bin.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

Using the chain rule

$$I(M;Y^N) = \sum_{i=1}^{N} I(M;Y_i|Y^{i-1})$$
$$= \sum_{i=1}^{N} \left[I(M,S_{i+1}^N;Y_i|Y^{i-1}) - I(S_{i+1}^N;Y_i|M,Y^{i-1}) \right], \quad (3.19)$$

and similarly

$$I(M;S^N) = \sum_{i=1}^{N} I(M;S_i|S_{i+1}^N)$$
$$= \sum_{i=1}^{N} \left[I(M,Y^{i-1};S_i|S_{i+1}^N) - I(Y^{i-1};S_i|M,S_{i+1}^N) \right]. \quad (3.20)$$

We will now show that[2]

$$\sum_{i=1}^{N} I(S_{i+1}^N;Y_i|M,Y^{i-1}) = \sum_{i=1}^{N} I(Y^{i-1};S_i|M,S_{i+1}^N). \quad (3.21)$$

The left-hand side of (3.21) can be written as

$$\sum_{i=1}^{N} I(S_{i+1}^N;Y_i|M,Y^{i-1}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} I(S_j;Y_i|M,Y^{i-1},S_{j+1}^N). \quad (3.22)$$

Similarly the right-hand side can be written as

$$\sum_{i=1}^{N} I(Y^{i-1};S_i|M,S_{i+1}^N) = \sum_{i=1}^{N} \sum_{j=1}^{i-1} I(Y_j;S_i|M,S_{i+1}^N,Y^{j-1})$$
$$= \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} I(Y_j;S_i|M,S_{i+1}^N,Y^{j-1}), \quad (3.23)$$

[2] This equality was first noted in [37].

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

and (3.21) follows. By subtracting (3.20) from (3.19), and by using (3.21) we get

$$NR - H(M|Y^N) \leq \sum_{i=1}^{N} \left[I(M, S_{i+1}^N; Y_i|Y^{i-1}) - I(M, Y^{i-1}; S_i|S_{i+1}^N) \right]. \tag{3.24}$$

Since

$$I(M, Y^{i-1}; S_i|S_{i+1}^N) = I(M, Y^{i-1}, S_{i+1}^N; S_i), \tag{3.25}$$

and

$$I(M, S_{i+1}^N; Y_i|Y^{i-1}) \leq I(M, Y^{i-1}, S_{i+1}^N; Y_i), \tag{3.26}$$

we get

$$\begin{aligned} &NR - H(M|Y^N) \\ &\quad \leq \sum_{i=1}^{N} \left[I(M, Y^{i-1}, S_{i+1}^N; Y_i) - I(M, Y^{i-1}, S_{i+1}^N; S_i) \right]. \end{aligned} \tag{3.27}$$

Letting $U_i = (M, Y^{i-1}, S_{i+1})$ we get

$$\begin{aligned} NR - H(M|Y^N) &\leq \sum_{i=1}^{N} \left[I(U_i; Y_i) - I(U_i; S_i) \right] \\ &\leq N \max_U \left[I(U; Y) - I(U; S) \right]. \end{aligned} \tag{3.28}$$

If the message is recovered reliably, then $H(M|Y^N)$ must be small, i.e., $H(M|Y^N) \leq N\epsilon$. Thus, the rate R cannot be larger than C.

3.3 Relations between the Causal and Non-Causal Cases

The simple relationship between causal and noncausal setup was first pointed out by Erez in his masters thesis [47]. We can rewrite (3.16) and present it in terms of strategies. Equation (3.16) is maximized over $P_{U|S}(u|s)P_{X|U,S}(x|u,s)$ and by Lemma 3.1, as discussed in Section 3.2, the optimal $P_{X|U,S}(x|u,s)$ may take values of 0 or 1 only, and thus, X is a deterministic function of U and S. Therefore, we can extend the input alphabet to the set of all functions $t : \mathcal{S} \to \mathcal{X}$ in order to

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

eliminate $P_{X|U,S}(x|u,s)$ from the problem. This is very important for the numerical calculation of channel capacity. We shall see in Section 7.1 that the new form of the capacity formula will make the numerical algorithms for the computation of the channel capacity more efficient.

The capacity for the non-causal case, in terms of strategies, is given by

$$C = \max_{P_{T|S}(t|s)} [I(T;Y) - I(T;S)], \tag{3.29}$$

where the maximization is taken over all joint distributions satisfying

$$P_{T,S,X,Y}(t,s,x,y) = P_S(s)P_{T|S}(t|s)\delta(x,t(s))P_{Y|X,S}(y|x,s). \tag{3.30}$$

There are $|\mathcal{X}|^{|\mathcal{S}|}$ possible functions, but we only need to use at most $|\mathcal{S}||\mathcal{X}|+1$ to achieve capacity in the non-causal case, and $|\mathcal{Y}|$ in the causal case. The difference between the strategies formulation of the causal and non-causal case lies in the dependency of the mapping (i.e., strategy) T on S. While in the causal case the random strategy T is selected *independently* of S, according to P_T, in the non-causal case the mapping T is selected randomly according to $P_{T|S}(\cdot|s)$. This small but significant difference between the two models accounts for the fact that in the operational definition of the code, the mapping (encoder) depends also on future values of the state sequence.

The causal side information capacity formula, discussed in Section 3.1, can be obtained as a special case of the non-causal one if we restrict U to be independent of S. Let us look at the proof of the direct part in the non-causal case. The encoder, in the causal case, is only given the state sequence until time index i, $i = 1,\ldots,N$, i.e., S^i. Therefore, the entire state sequence S^N and the codeword U^N will be jointly typical if the distribution of the auxiliary random variable U is independent of S. For the converse part, it was shown in [54] that

$$NR - N\epsilon \leq \sum_{i=1}^{N} I(U_i;Y_i) - I(U_i;S_i), \tag{3.31}$$

where U_i is defined as $U_i = (m, Y^{i-1}, S_{i+1}^N)$ and $(m, Y^{i-1}, S_{i+1}^N) \to (X_i, S_i) \to Y_i$ forms a Markov chain. Therefore, when the side information is given in a causal manner, U_i is independent of S_i

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

$((m, Y^{i-1}) \rightarrow (m, S^{i-1}) \rightarrow Y_i$ forms a Markov chain), resulting in Shannon's capacity (3.5).

Another important observation is the following. The last step in Gel'fand and Pinsker's proof of the converse part is

$$\sum_{i=1}^{N} I(U_i;Y_i) - I(U_i;S_i) \leq N \max_i [I(U_i;Y) - I(U_i;S)], \tag{3.32}$$

which proves the existence of a RV U, s.t. $R \leq I(U;Y) - I(U;S)$. If we have a power constraint, we cannot use (3.32). The index i which maximizes $I(U_i;Y) - I(U_i;S)$, does not have to fulfill the power constraint $E[\phi(X_i)] \leq \Gamma$. Therefore, we have to take a different approach to prove the converse. We define a random variable J uniformly distributed in $\{1,2,\ldots,N\}$, $U_J \triangleq (m, Y^{J-1}, S_{J+1}^N)$, and $U \triangleq (U_J, J)$. Now, the proof of the converse part will be complete with these steps:

$$\begin{aligned} R - \epsilon &\leq \frac{1}{N} \sum_{i=1}^{N} I(U_i;Y_i) - I(U_i;S_i) \\ &= I(U_J;Y_J|J) - I(U_J;S_J|J) \\ &= I(U_J,J;Y_J) - I(U_J,J;S_J) \\ &= I(U;Y) - I(U;S), \end{aligned} \tag{3.33}$$

where the second equality follows from the chain rule for mutual information.

We can also link Gel'fand and Pinsker's result to the case where both the transmitter and receiver have the same side information. This is a special case of Equation (3.14), in which both the transmitter and receiver have the same noisy version of the side information $W = V = Z$. In our case, $Z = S$, and the capacity is equal to

$$C = \max_{P_{X|S}(x|s)} I(X;Y|S). \tag{3.34}$$

In general, the capacity in this case is bigger than when we have only perfect non-causal CSIT and no CSIR (Gel'fand and Pinsker's setting), but this capacity can be equal to Gel'fand and Pinsker's capacity formula (3.16), if and only if [88] the distribution achieving the maximum

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

in (3.34) (say $P^*_{X,S,Y}$) can be represented in the form of having the auxiliary variable $U : P_{U,X,S,Y}(u,x,s,y)$ such that:

- $\sum_u P_{U,X,S,Y}(u,x,s,y) = P^*_{X,S,Y}(x,s,y)$,
- the channel input x can be represented as $x = f^*(u,s)$, where $f^*(\cdot,\cdot)$ is the encoding mapping which maximizes (3.16), and
- the following two Markov chains are satisfied: $U \to (X,S) \to Y$ and $S \to Y \to U$.

This is needed because the objective function in (3.34), $I(X;Y|S) = I(U;Y|S)$ from the first Markov chain and the objective function in (3.16), $I(Y;U) - I(U;S) = I(U;Y|S)$ from the second Markov chain. The second Markov chain can be interpreted as follows: all the dependency between the side information and the auxiliary variable U is captured through the channel output Y. So there is no loss in performance in terms of getting information about U, even if we do not have CSIR. These conditions link (3.16) and (3.34). They present another way to determine if (3.16) equals (3.34). The obvious method to determine if they are equal, is to separately calculate the capacity of (3.16) and (3.34). In the dirty-paper problem, as we shall see in Section 4.1, these conditions are met and the capacity with CSIT is equal to the capacity with CSIT and CSIR.

3.4 Modifications and Extensions

Gel'fand and Pinsker's capacity formula (3.16) can be extended to the case where the transmitter is subject to an average (input) cost constraint, i.e.,

$$C = \max_{\substack{P_{U|S}(u|s),\, P_{X|U,S}(x|u,s) \\ E\{\phi(X)\} \le \Gamma}} [I(U;Y) - I(U;S)]. \tag{3.35}$$

The proof of direct part of this capacity formula is a simple extension of from the proof of Gel'fand and Pinsker presented in Section 3.2. The only difference is in the random codebook generation, which should now account for the cost constraint. Given the message m and the state sequence s^N, the encoder seeks a codeword in bin m that is jointly typical with s^N, say $u^N(m,j)$ such that satisfies the average (input)

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

cost constraint (2.3), where $x_i = f(u_i(m,j), s_i)$, $i = 1, \ldots, N$. The proof of the converse part is a simple combination of the proof given by Gel'fand and Pinsker (presented in Section 3.2) and the usual proof of the converse part for channels with input constraints (see e.g., [35] Section 10.2).

By Lemma 3.1, $I(U;Y) - I(U;S)$ is convex in $P_{X|U,S}(x|u,s)$ for all other distributions fixes. Thus, a general $P_{X|U,S}(x|u,s)$, which can be written as a convex combination of $P_{X|U,S}(x|u,s)$'s, will always be dominated by a deterministic $P_{X|U,S}(x|u,s)$. However, a general $P_{X|U,S}(x|u,s)$ that satisfies the average (input) cost constraint might not be a convex combination of $P_{X|U,S}(x|u,s)$'s that also satisfy the average (input) cost constraint. We will now prove that similarly to the case where the transmitter is not subjected to an average (input) cost constraint, the distribution $P_{X|U,S}(x|u,s)$ can be restricted to a deterministic function $x = f(u,s)$ [27].

Without loss of generality, there exists $u_0 \in \mathcal{U}$ such that $0 < P_{X|U,S}(x|u_0,s) < 1$ for some x and s. If no such u_0 existed, then the distribution $P_{X|U,S}(x|u,s)$ puts all its mass on a single x for every u,s, and so, there exists a deterministic mapping $f : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ such that $P_{X|U,S}(x|u,s) = 1$ if and only if $x = f(u,s)$. Let us choose functions $f_1, \ldots, f_N : \mathcal{S} \to \mathcal{X}$ and positive constants $c_1, \ldots, c_N$ with $\sum_{i=1}^{N} c_i = 1$ such that

$$P_{X|U,S}(x|u_0,s) = \sum_{i=1}^{N} c_i \delta(x, f_i(s)), \quad \forall x \in \mathcal{X}, \quad \forall s \in \mathcal{S}. \tag{3.36}$$

Let $\mathcal{U}' = \{u'_1, \ldots, u'_N\}$ and $\widetilde{\mathcal{U}} = \mathcal{U}' \cup \mathcal{U} \backslash \{u_0\}$. Let the RV $\widetilde{U}$ take values in $\widetilde{\mathcal{U}}$ and have joint distribution

$$P_{X|\widetilde{U},S}(x|\widetilde{u},s) = \begin{cases} P_{X|U,S}(x|\widetilde{u},s) & \text{if } \widetilde{u} \in \mathcal{U} \backslash \{u_0\} \\ \delta(x, f_i(s)) & \text{if } \widetilde{u} = u' \in \mathcal{U}', \end{cases} \tag{3.37}$$

and

$$P_{\widetilde{U}|S}(\widetilde{u}|s) = \begin{cases} P_{U|S}(\widetilde{u}|s) & \text{if } \widetilde{u} \in \mathcal{U} \backslash \{u_0\} \\ c_i P_{U|S}(u_0|s) & \text{if } \widetilde{u} = u' \in \mathcal{U}'. \end{cases} \tag{3.38}$$

We first note that the joint distribution on S, X, and Y is the same under the original joint distribution on U, S, X, and Y and under the

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

new joint distribution on $\widetilde{U}$, S, X, and Y. That is,

$$\sum_{\widetilde{u}\in\widetilde{\mathcal{U}}} P_{Y|X,S}(y|x,s)P_{X|\widetilde{U},S}(x|\widetilde{u},s)P_{\widetilde{U}|S}(\widetilde{u}|s)P_S(s)$$

$$= \sum_{u\in\mathcal{U}} P_{Y|X,S}(y|x,s)P_{X|U,S}(x|u,s)P_{U|S}(u|s)P_S(s). \tag{3.39}$$

Both $H(Y)$ and $E\{\phi(X)\}$ are unaffected by the change in distributions. Second, we consider a joint distribution on the variables S, U, and $\widetilde{U}$ defined by

$$P_{\widetilde{U}|S,U}(\widetilde{u}|s,u) = \begin{cases} \delta(\widetilde{u},u) & \text{if } \widetilde{u}\in\mathcal{U}\backslash\{u_0\} \\ c_i\delta(u,u_0) & \text{if } \widetilde{u}=u'\in\mathcal{U}', \end{cases} \tag{3.40}$$

which is consistent with both joint distributions. Under this distribution, the random variables S, U, and $\widetilde{U}$ form a Markov chain. Thus, by the data processing inequality, $I(U;S)\geq I(\widetilde{U};S)$. We finally note that

$$P_{Y|U}(y|u_0) = \sum_{i=1}^{N} c_i P_{Y|\widetilde{U}}(y|u_i'), \tag{3.41}$$

which follows by the definitions of (3.36), (3.37), and (3.38). We can thus show, using the concavity of entropy, that $H(Y|\widetilde{U})\leq H(Y|U)$ and thus $I(\widetilde{U};Y)\geq I(U;Y)$. Therefore, by these three observations $I(\widetilde{U};Y)-I(\widetilde{U};S)\geq I(U;Y)-I(U;S)$ and the average (input) cost constraint is the same under both distributions, and thus the distribution $P_{X|U,S}(x|u,s)$ can be restricted to a deterministic function $x=f(u,s)$ since we can repeat this process until there is no such u_0.

We can also extend (3.29) to the case where the alphabets $\mathcal{X},\mathcal{Y},\mathcal{S}$ are the real line (see [5, 49]) and where the transmitter is subject to an average (input) cost constraint to yield

$$C = \sup_{P_{T|S}(t|s):E\{\phi(T(S))\}\leq\Gamma} [I(T;Y)-I(T;S)], \tag{3.42}$$

where $t\in\mathcal{T}$ is the set of all possible mappings $t:\mathcal{S}\to\mathcal{X}$.

Gel'fand and Pinsker's results were used by Costa [30] to analyze the additive interference channel, which is known also as the dirty-paper channel, as mentioned earlier. In the Introduction, the importance of this result and its relevance to analyzing many communication

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

problems was discussed. This result was also used to analyze digital watermarking and the capacity of a computer memory with defects. Costa's work and other related results will be presented in Section 4.

Cover and Chiang [24, 32] extended Gel'fand and Pinsker's model to the case where both the transmitter and receiver observe different non-perfect CSI sequences, which are correlated to the state sequence. This extension can be used to model a wider range of problems than the previous model. For example, it can model a high-definition television (HDTV) system, where the noisy analog version of the TV signal is the side information to the decoder. In this example, the side information is not really part of the channel, because the analog transmission serves as the side information and there is no state process in the channel.

The model which they used assumes a CSI signal, W_n, available to the transmitter, a correlated CSI signal, V_n, available to the receiver, and a memoryless channel with transition probability $P_{Y|X,W,V}(y|x,w,v)$. We assume that the CSI signals (W_n, V_n) are generated by a memoryless source

$$P_{W^N,V^N}(w^N, v^N) = \prod_{n=1}^{N} P_{W,V}(w_n, v_n). \tag{3.43}$$

The output Y^N has the conditional distribution

$$P_{Y^N|X^N,W^N,V^N}(y^N|x^N, w^N, v^N) = \prod_{n=1}^{N} P_{Y|X,W,V}(y_n|x_n, w_n, v_n). \tag{3.44}$$

For this model, we define a block code of length N as a sequence of N encoding functions $f_n : \mathcal{M} \times \mathcal{W}^N \to \mathcal{X}$, $n = 1, \ldots, N$, such that $x_n = f_n(m, w^N)$, where m ranges over the set of possible source messages $\mathcal{M}$. The decoding function is $g_n : \mathcal{Y}^N \times \mathcal{V}^N \to \mathcal{M}$, such that the decoded message is $\hat{m} = g(y^N, v^N)$. The capacity in this case will be

$$C = \max_{P_{U,X|W}(u,x|w)} [I(U;V,Y) - I(U;W)]. \tag{3.45}$$

Cover and Chiang's result follows directly from Gel'fand and Pinsker's result. Therefore, Cover and Chiang's model is not really a generalization of Gel'fand and Pinsker's original model, just like

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

Salehi's [93] model which is not really a generalization of Shannon's original model.

Another interesting aspect of the Gel'fand–Pinsker channel model presented in Section 2, is that similarly to a DMC with feedback [35], adding feedback to the Gel'fand–Pinsker channel model or to the Shannon channel model does not increase the capacity [82]. As in the classical DMC, feedback can help in simplifying the coding scheme in the Gel'fand–Pinsker channel, so there is no need for a complicated binning scheme.

We can look back at the last few steps in Gel'fand and Pinsker's converse proof [54], and see that it is general enough as to include the feedback. We have defined $U_i = (m, Y^{i-1}, S_{i+1}^N)$ where $U_i \to (X_i, S_i) \to Y_i$ forms a Markov chain. $U_i \to (X_i, S_i) \to Y_i$ is a Markov chain even if we include a feedback in the channel. This is due to the fact that $(m, S^N, Y^{i-1}) \to (X_i, S_i) \to Y_i$ is a Markov chain and therefore, since U_i is a deterministic function of (m, S^N, Y^{i-1}), $U_i \to (X_i, S_i) \to Y_i$ is a Markov chain.

We have mentioned in Section 3.3 that Shannon's causal counterpart of the problem is a special case of the Gel'fand-Pinsker non-causal problem. If we add a feedback to Shannon's model, we arrive at the same conclusion, i.e., feedback does not increase capacity in the causal model. This is due to the fact that the independence between U_i and S_i is maintained in the presence of feedback (see Section 3.3).

When we compare the causal to the non-causal problem, an interesting question is, what is the advantage in terms of capacity of having non-causal side information instead of a causal one. Jafar [61] investigated this problem and his result is stated in the following theorem.

Theorem 3.2. The capacity benefit of non-causal side information over causal side information is bounded as follows:

$$C_{\text{non-causal}}(W,V) - C_{\text{causal}}(W,V) \leq H(W|V). \tag{3.46}$$

Jafar has also expressed the capacity gain that can result from a fixed number of bits of causal or non-causal side information. Denoting by G the side information with fixed number of bits. The capacity advantage

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

of having G at the receiver or at the transmitter is stated in the following theorems.

Theorem 3.3. The maximum possible capacity improvement due to the availability of receiver side information is bounded by the amount of side information itself.

$$C_G - C \leq \lim_{N\to\infty} \frac{1}{N} H(G_1, G_2, \ldots, G_N), \tag{3.47}$$

where C_G is the capacity with the side information G and C is the capacity without the side information.

Theorem 3.4. The maximum possible capacity improvement from a fixed number side information bits (per channel use) is unbounded, even if the side information is causal.

From Theorem 3.3, we can see that if the receiver is provided with one bit of side information per channel use, the capacity benefit cannot be more than one bit.

3.5 Rate-Limited Side Information

While we have assumed thus far that the CSI is delivered to the transmitter and/or receiver, free-of-charge, a more appropriate scenario would describe these deliveries as being carried out across capacity-limited channels, in particular, capacities that are not necessarily higher than the entropy of the state process. Heegard and El Gamal [60] assumed that the CSIT and the CSIR are subjected to rate constraints R_e and R_d, respectively. Therefore, the CSI signals carry only partial information about the underlying state process. Their model is depicted in Figure 3.3.

The corresponding definition of the coding system is as follows:

Definition 3.1. An (N, R, R_e, R_d) code consists of the following four functions:
$J_e : \mathcal{S}^N \to \{1, \ldots, 2^{NR_e}\}$.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

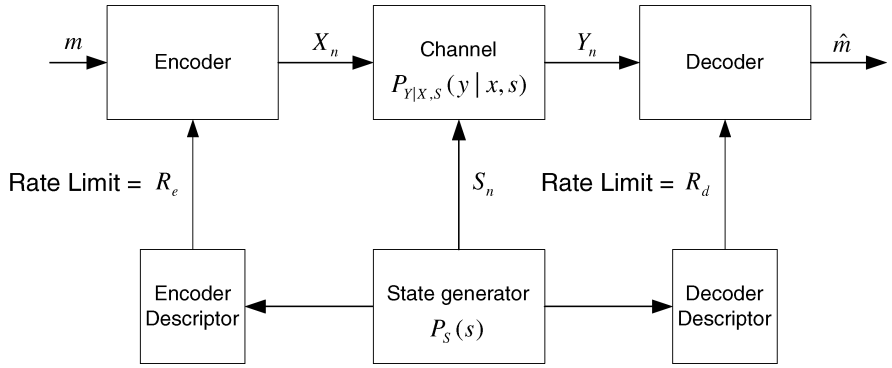

Fig. 3.3 A channel with two-sided non-causal state information with rate constraint model.

$J_d : \mathcal{S}^N \to \{1,\ldots,2^{NR_d}\}$.
$f : \{1,\ldots,2^{NR}\} \times \{1,\ldots,2^{NR_e}\} \to \mathcal{X}^N$.
$g : \mathcal{Y}^N \times \{1,\ldots,2^{NR_d}\} \to \{1,\ldots,2^{NR}\}$.

The mapping J_e generates a description of the state vector S^N to the encoder f, at rate R_e. Similarly, J_d generates a description of S^N to the decoder g, at rate R_d. The mappings J_e and J_d, are implemented by the encoder descriptor (ED) and decoder descriptor (DD) blocks, respectively.

Definition 3.2. For fixed R_e and R_d, a rate triplet (R, R_e, R_d) is achievable if for $\epsilon > 0$ there exists an (N, R, R_e, R_d) code for all sufficiently large N, with error probability $P_e \leq \epsilon$, where P_e is defined in Section 2.

The capacity is defined as the supremum over all R, such that (R, R_e, R_d) is achievable.

The complete characterization of the capacity region is still unknown. An inner bound to the capacity region was provided by Heegard and El Gamal, and is presented in Theorem 3.5.

Theorem 3.5 ([60]). Fix $(\mathcal{S}, P_S(s), \mathcal{X}, P_{Y|X,S}(y|x,s), \mathcal{Y})$ and alphabets $\mathcal{U}, \mathcal{S}_0, \mathcal{S}_e$, and $\mathcal{S}_d$. All rate triples (R, R_e, R_d) in the convex hull of

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

the set

$$\left\{\begin{array}{l}(R,R_e,R_d)|R_e > I(S_0,S_e;S)\\ R_d > I(S_0,S_d;S) - I(S_0,S_d;Y)\\ R_d > I(S_d;S|S_0) - I(S_d;Y|S_0)\\ R_e + R_d > I(S_0,S_e,S_d;S) - I(S_0,S_d;Y) + I(S_e;S_d|S_0)\\ R_e + R_d > I(S_e,S_d;S|S_0) - I(S_d;Y|S_0) + I(S_e;S_d|S_0)\\ R < I(U;Y,S_d|S_0) - I(U;S_e|S_0)\end{array}\right\}$$

for some probability mass function

$$\begin{aligned}&P_{S,S_0,S_e,S_d,U,X}(s,s_0,s_e,s_d,u,x)\\ &\quad = P_S(s)P_{S_0,S_e,S_d|S}(s_0,s_e,s_d|s)P_{U,X|S_0,S_e}(u,x|s_0,s_e)\end{aligned} \tag{3.48}$$

are achievable.

The variables S_0, S_e, S_d denote common side information, encoder side information and decoder side information, respectively. U is an auxiliary random variables that plays the role of the coding RV of Gel'fand and Pinsker. The first condition, is associated with the rate-distortion function for the channel from S to (S_0, S_e). The second condition, is associated with the Wyner–Ziv rate-distortion function [124] for the channel from S to (S_0, S_d) with the channel output Y serving as a decoder side information. The third condition, is associated with the Wyner–Ziv rate-distortion function for the channel from S to S_d with Y serving as a decoder side information, and S_0 as common side information. The fourth and fifth conditions, are combinations of the first condition with the second and third conditions, respectively. We also add to both of these conditions, the information between S_e and S_d, i.e., $I(S_e;S_d|S_0)$. This is due to the fact that if we want to look at the sum of the rates $R_e + R_d$ we have to take into account the influence of S_e on S_d, because now we do not have two separate problems of encoding S_e, and S_d. The last condition is similar to Gel'fand and Pinsker's capacity formula, conditioned on S_0, which is the common part of side information sent to the transmitter and receiver.

Although Theorem 3.5 gives only an inner bound on the achievable region, Heegard and El Gamal found several special cases for which it is the exact achievable region.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

(a) $R_e = R_d = 0$ (no description of defects), the capacity is

$$C = \max_{P_X(x)} I(X;Y). \tag{3.49}$$

(b) $R_e > H(S), R_d > H(S|Y)$ (complete description of defects at encoder and decoder), the capacity is equal to (3.34).
(c) $R_e > H(S), R_d = 0$ (complete description of defects at encoder and no description at decoder, i.e., the Gel'fand–Pinsker model), the capacity is equal to (3.16).
(d) $R_e > 0, R_d = H(S|Y)$ (complete description of defects at decoder).

In case (d), Heegard and El Gamal suggested an expression which is too optimistic. This case was later corrected by Rosenzweig *et. al.* [91], which gave a different expression for the capacity, as we shall see later. We note that $H(S|Y)$ depends on the choice of $P_{X|S}(x|s)$. So $H(S|Y)$ in itself is ill-defined.

The achievability of these capacity formulas follows from Theorem 3.5 by identifying the auxiliary random variables as follows:

(a) $S_0 = S_e = S_d = \emptyset, U = X$.
(b) $S_0 = S, S_e = S_d = \emptyset, U = X$.
(c) $S_0 = S_d = \emptyset, S_e = S$.
(d) $S_e = \emptyset, (S_0, S_d) = S, U = X$.

In [91], it was mentioned that substituting the variables $(S_0, S_d) = S$ in case (d), forces the auxiliary variable S_0 to be a deterministic function of S. Instead, Rosenzweig *et. al.* suggested that the assignment $S_d = S$. In this case, the capacity is given by

$$C = \max_{P_{X|S_0}(x|s_0)} \max_{P_{S_0|S}(s_0|s)} I(X;Y|S,S_0), \tag{3.50}$$

where the second maximum is taken over the distribution $P_{S_0|S}(s_0|s)$ satisfying

$$R_e \leq I(S_0;S). \tag{3.51}$$

Partial CSIT has received considerable attention in the context of multi-input–multi-output (MIMO) systems used over the block fading

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

channel. The presence of CSIT has been shown to yield significant performance gains in various aspects [9, 62, 92]. Many studies consider a MIMO system where the transmitter is given only a quantized finite rate CSI via an error-free feedback channel. Rosenzweig *et. al.* [91], studied these types of channels. They extended Heegard and El Gamal's model for a channel with partial CSIT and perfect CSIR (case (d)), to the case where the CSIT comprises two parts, both subjected to rate constraints. This model was later further extended by Cemal and Steinberg to the case where the transmitter receives multiple partial CSI signals (see [18] and [16]). Cemal and Steinberg [17] extended the model of a single-user channel with partial CSIT and perfect CSIR to multiple access channel (MAC), where the encoders have access to partial rate-limited CSIT and the decoder have access to perfect CSIR. Multiple access channels will be discussed in Section 6.6.1.

Cemal and Steinberg [16] have also extended the model for a channel with partial CSIT and perfect CSIR, to the case where the partial CSI is conveyed through a noisy channel (genie channel) to the encoder. They considered an extended communication model where an information source generates i.i.d random variables $W_n \in \mathcal{W}$, drawn under distribution $P_W(\cdot)$, independent of the state source S_n, is to be transmitted through the channel, where the transmission is subjected to a distortion constraint between the source signal W and its reconstruction $\hat{W}$. This communication model is depicted in Figure 3.4. The Gel'fand–Pinsker channel $P_{Y|X,S}(\cdot|\cdot,\cdot)$ operates at a rate of ρ_c^{-1} channel uses per information symbol and the genie channel $P_{S_e|K}(\cdot|\cdot)$ operates at a rate of ρ_s channel uses per state source (or at ρ_s/ρ_c channel uses per information symbol). The joint source-channel state encoder generates a block code of length $N_s = \rho_s N$, $k^{N_s} \in \mathcal{K}^{N_s}$, from a state source block of length N, $s^N \in \mathcal{S}^N$. The block K^{N_s} is subjected to a transmission cost constraint such that $E[\phi_s(K^{N_s})] \leq N_s \Gamma_s$, where $\phi_s(K^{N_s}) = \sum_{n=1}^{N_s} \phi_s(k_n)$ for some cost function $\phi_s : \mathcal{K} \to \mathbb{R}_+$. The joint source-channel encoder, receives the partial CSIT $S_e^{N_s}$ which is the output of the genie channel, generates the input to the Gel'fand–Pinsker channel $x^N \in \mathcal{X}^N$ based on $S_e^{N_s}$ and an information source block of length $N_c = \rho_c N$, $W^{N_c} \in \mathcal{W}^{N_c}$. The input block x^N is subjected to a transmission cost constraint (2.3). The decoder has access to perfect CSIR s^N and produces an estimate

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

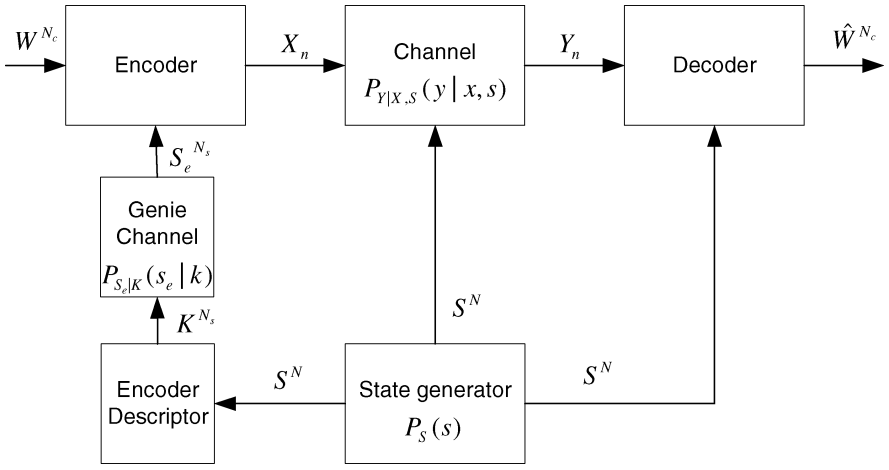

Fig. 3.4 Joint state-channel source–channel coding with partial state information at the transmitter.

of the source $\hat{w}^{N_c} \in \hat{\mathcal{W}}^{N_c}$ from the channel output y^N and the CSIR s^N. The quality of the estimation is measured by the expectation of an additive distortion measure: $d(w^{N_c}, \hat{w}^{N_c}) = \sum_{n=1}^{N_c} d(w_n, \hat{w}_n)$, where $d : \mathcal{W} \times \hat{\mathcal{W}} \to \mathbb{R}_+$ is a single-letter distortion measure.

Definition 3.3. An $(N, N_c, N_s, D, \Gamma, \Gamma_s)$ joint state-source–channel (SSC) code for the information source $P_W(\cdot)$, state source $P_S(\cdot)$, Gel'fand–Pinsker channel $P_{Y|X,S}(\cdot|\cdot,\cdot)$ and genie channel $P_{S_e|K}(\cdot|\cdot)$, consists of three mappings:

$$\begin{aligned} f_s &: S^N \to K^{N_s} \\ f &: W^{N_c} \times S_e^{N_s} \to X^N \\ g &: Y^N \times S^N \to \hat{W}^{N_c} \end{aligned} \tag{3.52}$$

such that $E[d\left(W^{N_c}, g(Y^N, S^N)\right)] \le N_c D$, $E[\phi(f(W^{N_c}, S_e^{N_s}))] \le N\Gamma$, and $E[\phi_s(f_s(S^N))] \le N_s \Gamma_s$.

The distortion cost triple (D, Γ, Γ_s) is achievable with bandwidth expansion factors ρ_c and ρ_s if for any $\epsilon > 0$ and sufficiently large N, there exists an $(N, \rho_c N, \rho_s N, D + \epsilon, \Gamma, \Gamma_s)$ joint SSC code. Given bandwidth expansion factors ρ_c and ρ_s, the distortion-cost region is the

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

closure of the set of achievable distortion-cost triples (D,Γ,Γ_s) and is denoted by $\mathcal{D}$. Cemal and Steinberg [16] have characterized $\mathcal{D}$.

Theorem 3.6 ([16]). For any information source $P_W(\cdot)$, state source $P_S(\cdot)$, Gel'fand–Pinsker channel $P_{Y|X,S}(\cdot|\cdot,\cdot)$ and genie channel $P_{S_e|K}(\cdot|\cdot)$ with bandwidth expansion factors ρ_c and ρ_s, $\mathcal{D}$ is the set of all distortion-cost triples (D,Γ,Γ_s) for which there exist a random variable S_0 taking values in $\mathcal{S}_0$ such that the following conditions are satisfied simultaneously: (i) The Markov relations $S \to S_0 \to X$, $S_0 \to (X,S) \to Y$ hold. (ii) The auxiliary alphabet $\mathcal{S}_0$ satisfies $|\mathcal{S}_0| \le |S| + 1$. (iii) The distortion-cost triple (D,Γ,Γ_s) satisfies

$$\begin{aligned} \rho_c R_W(D) &\le I(X;Y|S,S_0) \\ I(S,S_0) &\le \rho_s C_g(\Gamma_s) \\ E[\phi(X)] &\le \Gamma, \end{aligned} \tag{3.53}$$

where $R_W(D) \triangleq \min_{P_{\hat{W}|W}(\hat{w}|w):E[d(w,\hat{w})]\le D} I(W;\hat{W})$ is the rate-distortion function and $C_g(\Gamma_s) \triangleq \max_{P_K(k):E[\phi_s(k)]\le\Gamma_s} I(K;S_e)$ is the capacity–cost function.

If we compare the achievable rate region given in (3.50), (3.51) to the distortion-cost region characterized in Theorem 3.6, we see that the partial description of R_e in (3.51) is replaced by $\rho_s C_g(\Gamma_s)$ and the Gel'fand–Pinsker channel capacity C in (3.50) is replaced by $\rho_c R_W(D)$. This connection implies a separation principle in our extended scenario. A separation principle holds for state coding independent of the genie channel (but depending on the Gel'fand–Pinsker channel), and a separation principle holds for the main source coding, independent of the Gel'fand–Pinsker channel statistics.

In [102, 103] a dual problem was investigated, where the transmitter have access to full non-causal CSIT and the receiver have access to rate-limited CSIR. The capacity of this model is given by

$$C = \max\left[I(U;Y|S_d) - I(U;S|S_d)\right], \tag{3.54}$$

where the maximization is over all $P_{X,S_d,U|S}(\cdot,\cdot,\cdot|\cdot)$ satisfying

$$R_d \ge I(S;S_d) - I(Y;S_d). \tag{3.55}$$

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

Coding of side information intended for the channel decoder is a Wyner–Ziv like problem, since the channel output depends statistically on the state, thus serving as side information in the decoding of the encoded state. Steinberg [104] suggested as an application for this model the problem of reversible information embedding with compressed host at the decoder. Information embedding, and reversible information embedding is discussed in Section 5.2.

The problem of coding for a channel with rate-limited CSIR and no CSIT was investigated by Ahlswede and Han in [2]. They have showed that the channel can be viewed as a special case of a relay channel.[3] The receiver in this problem, has access to a rate-limited side information S_d via a separate communication channel at a fixed rate R_d. The capacity of this channel is given by

$$C = \max I(X;Y|S_d), \tag{3.56}$$

where the maximum is over all joint distribution of the form $P_X(X)P_S(s)P_{Y|X,S}(y|x,s)P_{S_d|S}(s_d|s)$ such that

$$I(S;S_d|Y) \leq R_d, \tag{3.57}$$

where the cardinality of S_d is bounded by $|\mathcal{S}_d| \leq |\mathcal{S}| + 1$. In [33], a special case of the problem was considered where the channel state S is a deterministic function of (X,Y), e.g., $S = X + Y$ as in the binary channel. For this case, the capacity is given by

$$C = \max_{P_X(x)} \min\{I(X;Y) + R_d, I(X;Y,S)\}. \tag{3.58}$$

Another special case of the problem in which the channel is a modulo sum relay channel was considered in [3] (see [3] for more details).

[3] A relay channel is a channel in which there is one sender and one receiver with a number of intermediate nodes which act as relays to help the communication from the sender to the receiver.

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information

4

Specific Channel Models

In this section, we describe two specific models of channels with side information. We use the results presented in Section 3 to find the capacities of these channels and we present coding techniques. The specific channels are the dirty-paper channel, the AWGN channel with fading and the modulo additive noise channel.

4.1 The Dirty-Paper Channel

In this subsection, we consider the problem of a power constrained additive noise channel, where part of the noise is known at the transmitter as side information. This part of the noise may be the result of an additive interference. This model is suitable to describe, for example, a scenario in which a user, which is located close to the main transmitter, is interfering, but cooperating by revealing his messages in advance. It is also suitable to describe a scenario in which one transmitting antenna, is interfering another antenna, serving the same user in a MIMO scenario.

The channel output, in this model, is given by

$$Y^N = X^N + S^N + Z^N, \tag{4.1}$$

Documents
Document Sharing Hub
Full document details: IEEE Channel Coding in the Presence of Side Information