Gaussian Channels: Information, Estimation and Multiuser Detection

Dongning Guo

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Electrical Engineering

November, 2004

Abstract

This thesis represents an addition to the theory of information transmission, signal estimation, nonlinear filtering, and multiuser detection over channels with Gaussian noise. The work consists of two parts based on two problem settings—single-user and multiuser—which draw different techniques in their development.

The first part considers canonical Gaussian channels with an input of arbitrary but fixed distribution. An "incremental channel" is devised to study the mutual information increase due to an infinitesimal increase in the signal-to-noise ratio (SNR) or observation time. It is shown that the derivative of the input-output mutual information (nats) with respect to the SNR is equal to half the minimum mean-square error (MMSE) achieved by optimal estimation of the input given the output. This relationship holds for both scalar and vector signals, as well as for discrete- and continuous-time models. This information-theoretic result has an unexpected consequence in continuous-time estimation: The causal filtering MMSE achieved at SNR is equal to the average value of the noncausal smoothing MMSE achieved with a channel whose signal-to-noise ratio is chosen uniformly distributed between 0 and SNR.

The second part considers Gaussian multiple-access channels, in particular code-division multiple access (CDMA), where the input is the superposition of signals from many users, each modulating independent symbols of an arbitrary distribution onto a random signature waveform. The receiver conducts optimal joint decoding or suboptimal separate decoding that follows a posterior mean estimator front end, which can be particularized to the matched filter, decorrelator, linear MMSE detector, and the optimal detectors. Large-system performance of multiuser detection is analyzed in a unified framework using the replica method developed in statistical physics. It is shown under replica symmetry assumption that the posterior mean estimate, which is generally non-Gaussian in distribution, converges to a deterministic function of a hidden Gaussian statistic. Consequently, the multiuser channel can be decoupled into equivalent single-user Gaussian channels, where the degradation in SNR due to multiple-access interference, called multiuser efficiency, is determined by a fixed-point equation. The multiuser efficiency uniquely characterizes the error performance and input-output mutual information of each user, as well as the overall system spectral efficiency.

Acknowledgements

I am indebted to my advisor, Professor Sergio Verdú, for his support, encouragement, and invaluable advice during the course of my Ph.D. study. Sergio has set an example of a great scholar that would be my goal for many years to come.

I am very grateful to my collaborator, Dr. Shlomo Shamai, whose ideas, suggestions and enormous knowledge of the research literature have benefited me a lot.

I would like to thank Shlomo, Dr. H. Vincent Poor, Dr. Robert Calderbank and Dr. Mung Chiang for serving in my thesis committee.

Thanks to Chih-Chun for many discussions, to Juhua for always sharing her wisdom, and all my friends both at Princeton and elsewhere who made the last five years full of fun.

This thesis is dedicated to my parents and grandparents.

Contents

List of Figures

Chapter 1

Introduction

The exciting revolution in communication technology in recent years would not have been possible without significant advances in detection, estimation, and information theory since the mid-twentieth century. This thesis represents an addition to the theory of information transmission, signal estimation, nonlinear filtering and multiuser detection over channels with Gaussian noise.

"The fundamental problem of communication," described by Shannon in his 1948 landmark paper [86], "is that of reproducing at one point either exactly or approximately a message selected at another point." Typically, communication engineers are given a physical channel, which, upon an input signal representing the message, generates an output signal coupled with the input. The problem is then to determine which message was sent among all possibilities based on the observed output and knowledge about the channel.

A key index of success is how many distinct messages, or in more abstract sense, the amount of information that can be communicated reliably through the channel. This is measured by the notion of *mutual information*. Meanwhile, the difficulty faced by communication engineers is how much error the receiver would make in estimating the input signal given the output. A key measure here is the *minimum mean-square error* (MMSE) due to its tractability and practical effectiveness. This thesis is centered around the input-output mutual information and MMSE of *Gaussian channels*, whose output is equal to the input plus random noise of Gaussian distribution. Two problem settings—single-user and multiuser—which draw different techniques in their development, consist of the two main chapters (Chapters 2 and 3) of the thesis.

1.1 Mutual Information and MMSE

Chapter 2 deals with the transmission of an arbitrarily distributed input signal from a single source (user). Since Wiener's pioneer work in 1940s [114], a rich theory of detection and estimation in Gaussian channels has been developed [59], notably the matched filter [74], the RAKE receiver [79], the likelihood ratio (e.g., [78]), and the estimator-correlator principle (e.g., [92, 84, 56]). Alongside was the development of the theory of stochastic processes and stochastic calculus since 1930s that set the mathematical stage. Shannon, the founder of information theory, obtained the capacity of Gaussian channels under power constraint [86]. Since then, much has been known about the information theory of Gaussian channels, as well as practical codes that achieve rates very close to the channel capacity.

Information theory and the theory of detection and estimation have largely been treated separately, although numerous results and techniques connect the two. In the Gaussian regime, information and estimation have strong ties through known likelihood ratios. For example, by taking the expectation of the "estimator-correlator" type of log-likelihood ratios in continuous-time problems, Duncan found that the input-output mutual information can be expressed as a time-average of the MMSE of causal filtering [21].

Encompassing not only continuous-time, but also discrete-time channels, as well as scalar random transformations with additive Gaussian noise, Chapter 2 finds a fundamental relationship between the input-output mutual information and MMSE, which is unknown until this work. That is, the derivative of the mutual information (nats) with respect to the signal-to-noise ratio (SNR) is equal to half the MMSE, achieved by (noncausal) conditional mean estimation. The relationships holds for vector as well as scalar inputs of arbitrary but fixed distribution. Using these information-theoretic results, a new relationship is found in continuous-time nonlinear filtering: Regardless of the input signal statistics, the causal filtering MMSE achieved at SNR is equal to the expected value of the noncausal smoothing MMSE achieved with a channel whose signal-to-noise ratio is chosen uniformly distributed between 0 and SNR.

The connection between information and estimation is drawn through a key observation that the mutual information of a small SNR Gaussian channel is essentially the input variance times the SNR. This is due to the geometry of the likelihood ratio associated with the Gaussian channel. Major results in Chapter 2 are proved by using the idea of "incremental channels" to investigate the increase of mutual information due to an infinitesimal increase in SNR or observation time.

In a nutshell, Chapter 2 reveals a "hidden" link between information and estimation theory. The new relationship facilitates interactions of the two fields and lead to interesting results.

1.2 Multiuser Channels

Chapter 3 studies multiuser communication systems where individual user's mutual information and estimation error are of interest. Each user modulates coded symbols onto a multidimensional signature waveform which is randomly generated. The input to the Gaussian channel is the superposition of signals from many users, where the users are distinguished by their signature waveforms. This model is often referred to as code-division multiple access (CDMA), but is also a popular paradigm of many multi-input multi-output (MIMO) channels. The most efficient use of such a channel is by optimal joint decoding, the complexity of which is rarely affordable in practice. A common suboptimal strategy is to apply a multiuser detector front end which generates an estimate of the input symbols of each user and then perform independent single-user decoding.

Detection and estimation of multiuser signals are well studied [112]. It is often impossible to maintain orthogonality of the signatures and hence interference among users degrades performance. Various interference suppression techniques provide a wide spectrum of tradeoffs between performance and complexity. Detectors range from the primitive single-user matched filter, to the decorrelator, linear MMSE detector, sophisticated interference cancelers, and the jointly and individually optimal detectors. Performance analysis of various multiuser detection techniques is of great theoretical and practical interest.

Verdú first used the concept of *multiuser efficiency* to refer to the SNR degradation relative to a single-user channel calibrated at the same bit-error-rate (BER) [107]. Exact evaluation of the multiuser efficiency of even the simplest matched filter can be highly complex, since the multiple access interference (MAI), which shows up in detection outputs, often takes an exponential number of different values. More recently, much attention is devoted to the large-system regime, where dependence of performance on signature waveforms diminishes. For most linear front ends of interest, regardless of the input, the MAI in detection output is asymptotically normal, which is tantamount to an enhancement of the background Gaussian noise. Thus the multiuser channel can be decoupled into single-user ones with a degradation in the effective SNR. As far as linear detectors are concerned, the multiuser efficiency directly maps to the SNR of detection output, which completely characterizes large-system performance, and can often be obtained analytically, e.g., [112, 99, 44]. Unfortunately, the traditional wisdom of asymptotic normality fails in case of nonlinear detectors, since the detection output is in general non-Gaussian in the large-system limit. Analytical result for performance is scarce and numerical analysis is costly but often the only resort.

Chapter 3 unravels several outstanding problems in multiuser detection and its information theory. A family of multiuser detectors is analyzed in a unified framework by treating each detector as a posterior mean estimator (PME) informed with a carefully chosen posterior probability distribution. Examples of such detectors include the above-mentioned linear detectors, as well as the optimal nonlinear detectors. One of the key results is that the detection output of all such detectors, although asymptotically non-Gaussian in general, converges to a deterministic function of a "hidden" Gaussian statistic centered at the transmitted symbol. Hence asymptotic normality is still valid subject to an inverse function, and system performance can nonetheless be fully characterized by a scalar parameter, the multiuser efficiency, which is found to satisfy a fixed-point equation along with the MMSE of an adjunct Gaussian channel. Moreover, the spectral efficiencies, i.e., total mutual information per dimension, under both joint and separate decoding are found in simple expressions in the multiuser efficiency.

The methodology applied in Chapter 3 is rather unconventional. The many-user communication system is regarded as a thermodynamic system consisting of a large number of interacting particles, known as a *spin glass*. The system is studied in the large-system limit using the replica method developed in statistical physics. We leave the "self-averaging property", the "replica trick" and replica symmetry assumption unjustified, which are themselves notoriously difficult challenges in mathematical physics. The general results obtained are consistent with known results in special cases, and supported by numerical example.

The use of the replica method in multiuser detection was introduced by Tanaka [96], who also suggested special-case PME with postulated posteriors, although referred to as the marginal-posterior-mode detectors. Chapter 3 puts forth the "decoupling principle" thereby enriching Tanaka's groundbreaking framework. The replica analysis is also developed in this chapter to a full generality in signaling and detection schemes based on [96].

In all, this thesis studies mutual information, posterior mean estimation and their interactions, and presents single-user and multiuser variations on this theme. In the single-user setting (Chapter 2), a fundamental formula that links mutual information and MMSE is discovered, which points to new relationships, applications, and open problems. In the multiuser setting (Chapter 3), the multidimensional channel with a rich structure is essentially decoupled into equivalent single-user Gaussian channels where the SNR degradation is fully quantified.[1] In either case, the results speak of new connections between the fundamental limits of digital communications to those of analog signal processing.

1.3 Notation

Throughout this thesis, random objects and matrices are denoted by upper case letters unless otherwise noted. Vectors and matrices are in bold font. For example, x denotes a number, $\boldsymbol{x}$ a column vector, X a scalar random variable, $\boldsymbol{X}$ a random column vector, $\boldsymbol{S}$ a matrix which is either random or deterministic depending on the context. Gaussian distribution with mean m and variance σ^2 is denoted by $\mathcal{N}\left(m, \sigma^2\right)$ and unit Gaussian random noise is usually denoted by the letter N. In general, P_X denotes the probability measure (or distribution) of the random object X. The probability density function, if exists, is denoted by p_X. Sometimes q_X also denotes a probability density function to distinguish from p_X. An expectation $\mathsf{E}\{\cdot\}$ is taken over the joint distribution of the random variables within the braces. Conditional expectation is denoted by $\mathsf{E}\{\cdot \mid \cdot\}$. In general, notations are introduced at first occurrence.

[1] Chronologically most of the results in Chapter 3 were obtained before those in Chapter 2. In fact, it was the process of proving Theorem 3.1 that prompted Theorem 2.1 and thereby the main theme of Chapter 2. After reading an earlier version of [43] Professor Shlomo Shamai came up independently of us with a proof of Theorem 2.1 using Duncan's Theorem (see Section 2.4.1). Upon finding out that we had obtained similar results independently we joined forces, and the results in Chapter 2 have been obtained in close cooperation with both my thesis advisor, Professor Sergio Verdú and Professor Shlomo Shamai.

Chapter 2

Mutual Information and MMSE

This chapter unveils several new relationships between the input-output mutual information and MMSE of Gaussian channels. The relationships are shown to hold for arbitrarily distributed input signals and the broadest settings of Gaussian channels. Although the signaling can be multidimensional, the input is regarded as a whole from a single source. Multiuser problems where individual user's performance is of interest will be studied in Chapter 3.

2.1 Introduction

Consider an arbitrary pair of jointly distributed random objects (X, Y). The *mutual information*, which stands for the amount of information contained in one of them about the other, is defined as the expectation of the logarithm of the likelihood ratio (Radon-Nikodym derivative) between the joint probability measure and the product of the marginal measures [60, 75]:

$$I(X;Y) = \int \log \frac{\mathrm{d}P_{XY}}{\mathrm{d}P_X \, \mathrm{d}P_Y} \, \mathrm{d}P_{XY}. \tag{2.1}$$

Oftentimes, one would also want to infer the value of X from Y. An *estimate* of X given Y is essentially a function of Y, which is desired to be close to X in some sense. Suppose that X resides in a metric space with L^2-norm defined, then the *mean-square error* of an estimate $f(Y)$ of X is given by

$$\mathsf{mse}_f(X|Y) = \mathsf{E}\left\{|X - f(Y)|^2\right\}, \tag{2.2}$$

where the expectation is taken over the joint distribution P_{XY}. It is well-known that the minimum of (2.2), referred to as the *minimum mean-square error* or MMSE, is achieved by *conditional mean estimation* (CME) (e.g., [76]):

$$\widehat{X}(Y) = \mathsf{E}\left\{ X \mid Y \right\}, \tag{2.3}$$

where the expectation is over the posterior probability distribution $P_{X|Y}$. In Bayesian statistics literature, the CME is also known as the *posterior mean estimation* (PME). Clearly, both mutual information and MMSE are measures of dependence between two random objects.

In general, (X, Y) can be regarded as the input-output pair of a channel characterized by the random transformation $P_{Y|X}$ and input distribution P_X. The mutual information is then a measure of how many distinct input sequences are distinguishable on average by observing the output sequence from repeated and independent use of such a channel. Meanwhile, the MMSE stands for the minimum error in estimating each input X using the observation Y while being informed of the posterior distribution $P_{X|Y}$.

This thesis studies the important case where X and Y denote the input and output of an additive Gaussian noise channel respectively. Take for example the simplest scalar Gaussian channel with an arbitrary input. Fix the input distribution. Let the power ratio of the signal and noise components seen in the channel output, i.e., signal-to-noise ratio, be denoted by snr. Both the input-output mutual information and the MMSE are then monotonic functions of the SNR, denoted by $I(\mathsf{snr})$ and $\mathsf{mmse}(\mathsf{snr})$ respectively. This chapter finds that the mutual information in nats and the MMSE satisfy the following relationship regardless of the input statistics:

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr}). \tag{2.4}$$

Simple as it is, the identity (2.4) was unknown before this work. It is trivial that one can go from one monotonic function to another by simply composing the inverse function of one with the other; what is quite surprising here is that the overall transformation is not only strikingly simple but also independent of the input distribution. In fact the relationship (2.4) and its variations hold under arbitrary input signaling and the broadest settings of Gaussian channels, including discrete-time and continuous-time channels, either in scalar or vector versions.

In a wider context, the mutual information and mean-square error are at the core of information theory and signal processing respectively. Thus not only is the significance of a formula like (2.4) self-evident, but the relationship is intriguing and deserves thorough exposition.

At zero SNR, the right hand side of (2.4) is equal to one half of the input variance. In that special case the identity, and in particular, the fact that at low SNRs the mutual information is insensitive to the input distribution has been remarked before [111, 61, 104]. Relationships between the local behavior of mutual information at vanishing SNR and the MMSE are given in [77].

Formula (2.4) can be proved using a new idea of "incremental channels", which is to analyze the increase in the mutual information due to an infinitesimal increase in SNR, or equivalently, the decrease in mutual information due to an independent extra Gaussian noise which is infinitesimally small. The change in mutual information is found to be equal to the mutual information of a Gaussian channel whose SNR is infinitesimally small, in which region the mutual information is essentially linear in the estimation error, and hence relates the rate of mutual information increase to the MMSE.

A deeper reasoning of the relationship, however, traces to the geometry of Gaussian channels, or, more tangibly, the geometric properties of the likelihood ratio associated with signal detection in Gaussian noise. Basic information-theoretic notions are firmly associated with the likelihood ratio, and foremost is the mutual information. The likelihood ratio also plays a fundamental role in detection and estimation, e.g., in hypothesis testing, it is compared to a threshold to determine which hypothesis to take. Moreover, the likelihood ratio is central in the connection of detection and estimation, in either continuous-time setting [55, 56, 57] or discrete one [48]. In fact, Esposito [27] and Hatsell and Nolte [46] noted simple relationships between conditional mean estimation and the gradient and Laplacian of the log-likelihood ratio respectively, although they did not import mutual information into the picture. Indeed, the likelihood ratio bridges information measures and basic quantities in detection and estimation, and in particular, the estimation errors (e.g., [65]). The relationships between information and estimation have been continuously used to evaluate results in one area taking advantage of known results from the other. This is best exemplified by the classical capacity-rate distortion relations, that have been used to develop lower bounds on estimation errors on one hand [119] and on the other to find achievable bounds for mutual information based on estimation errors associated with linear estimators [31].

The central formula (2.4) holds in case of continuous-time Gaussian channels as well, where the left hand side shall be replaced by the input-output mutual information rate, and the right hand side by the average noncausal smoothing MMSE per unit time. The information-theoretic result has also a surprising consequence in relating the causal and noncausal MMSEs, which becomes clear in Section 2.3.

In fact, the relationship between the mutual information and noncausal estimation error holds in even more general settings of Gaussian channels. Zakai has recently generalized the central formula (2.4) to the abstract Wiener space [120].

The remainder of this chapter is organized as follows. Section 2.2 deals with random variable/vector channels, while continuous-time channels are considered in Section 2.3. Interactions between discrete- and continuous-time models are studied in Section 2.4. Results for general channels and information measures are presented in Section 2.5.

2.2 Scalar and Vector Channels

2.2.1 The Scalar Gaussian-noise Channel

Consider a real-valued scalar Gaussian-noise channel of the canonical form:

$$Y = \sqrt{\mathsf{snr}}\, X + N, \tag{2.5}$$

where snr denotes the signal-to-noise ratio of the observed signal,[1] and the noise $N \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable independent of the input, X. The input-output conditional probability density is described by

$$p_{Y|X;\mathsf{snr}}(y|x;\mathsf{snr}) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(y - \sqrt{\mathsf{snr}}\, x\right)^2\right]. \tag{2.6}$$

Let the distribution of the input be P_X, which does not depend on snr. The marginal probability density function of the output exists:

$$p_{Y;\mathsf{snr}}(y;\mathsf{snr}) = \mathsf{E}\left\{p_{Y|X;\mathsf{snr}}(y|X;\mathsf{snr})\right\}, \quad \forall y. \tag{2.7}$$

Given the channel output, the MMSE in estimating the input is a function of snr:

$$\mathsf{mmse}(\mathsf{snr}) = \mathsf{mmse}\left(X \mid \sqrt{\mathsf{snr}}\, X + N\right). \tag{2.8}$$

The input-output mutual information of the channel (2.5) is also a function of snr. Let it be denoted by

$$I(\mathsf{snr}) = I\left(X; \sqrt{\mathsf{snr}}\, X + N\right). \tag{2.9}$$

[1] If $\mathsf{E}X^2 = 1$ then snr complies with the usual notion of signal-to-noise power ratio E_s/σ^2.

To start with, consider the special case when the distribution P_X of the input X is standard Gaussian. The input-output mutual information is then the well-known channel capacity under constrained input power [86]:

$$I(\mathsf{snr}) = \mathsf{C}(\mathsf{snr}) = \frac{1}{2}\log(1+\mathsf{snr}). \tag{2.10}$$

Meanwhile, the conditional mean estimate of the Gaussian input is merely a scaling of the output:

$$\widehat{X}(Y;\mathsf{snr}) = \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\,Y, \tag{2.11}$$

and hence the MMSE is:

$$\mathsf{mmse}(\mathsf{snr}) = \frac{1}{1+\mathsf{snr}}. \tag{2.12}$$

An immediate observation is

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr})\,\log e. \tag{2.13}$$

Here the base of logarithm is consistent with the unit of mutual information. From this point on throughout this thesis, we assume nats to be the unit of all information measures, and that logarithms have base e, so that $\log e = 1$ disappears from (2.13). It turns out that the above relationship holds not only for Gaussian inputs, but for all inputs of finite power:

Theorem 2.1 *For every input distribution P_X that satisfies $\mathsf{E}X^2 < \infty$,*

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I\left(X;\sqrt{\mathsf{snr}}\,X+N\right) = \frac{1}{2}\mathsf{mmse}\left(X \mid \sqrt{\mathsf{snr}}\,X+N\right). \tag{2.14}$$

Proof: See Section 2.2.3. ∎

The identity (2.14) reveals an intimate and intriguing connection between Shannon's mutual information and optimal estimation in the Gaussian channel (2.5), namely, the rate of the mutual information increase as the SNR increases is equal to half the minimum mean-square error achieved by the optimal (in general nonlinear) estimator.

Theorem 2.1 can also be verified for a simple and important input signaling: ± 1 with equal probability. The conditional mean estimate is given by

$$\widehat{X}(Y;\mathsf{snr}) = \tanh\left(\sqrt{\mathsf{snr}}\,Y\right). \tag{2.15}$$

The MMSE and the mutual information are obtained as:

$$\mathsf{mmse}(\mathsf{snr}) = 1 - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \tanh(\mathsf{snr} - \sqrt{\mathsf{snr}}\,y)\,\mathrm{d}y, \tag{2.16}$$

and (e.g., [6, p. 274] and [30, Problem 4.22])

$$I(\mathsf{snr}) = \mathsf{snr} - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \log\cosh(\mathsf{snr} - \sqrt{\mathsf{snr}}\,y)\,\mathrm{d}y \tag{2.17}$$

respectively. Verifying (2.14) is a matter of algebra [45].

For illustration purposes, the MMSE and the mutual information are plotted against the SNR in Figure 2.1 for Gaussian and binary inputs.

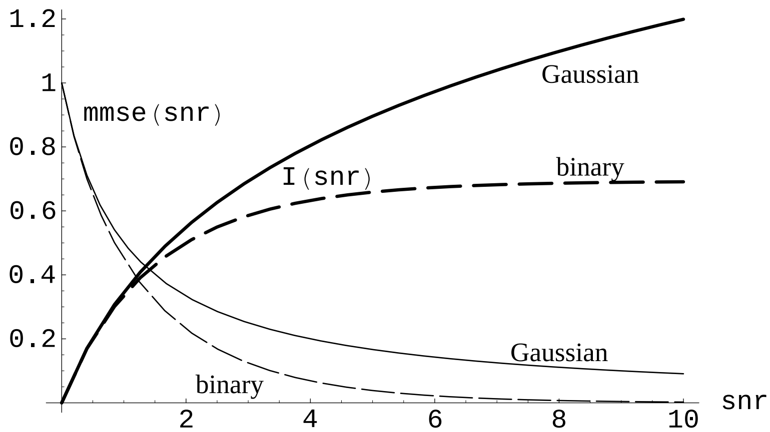

Figure 2.1: The mutual information (in nats) and MMSE of a scalar Gaussian channel with Gaussian and binary inputs, respectively.

2.2.2 A Vector Channel

Consider a multiple-input multiple-output system described by the vector Gaussian channel:

$$\boldsymbol{Y} = \sqrt{\mathsf{snr}}\,\boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{N} \tag{2.18}$$

where $\boldsymbol{H}$ is a deterministic $L \times K$ matrix and the noise vector $\boldsymbol{N}$ consists of independent identically distributed (i.i.d.) standard Gaussian entries. The input $\boldsymbol{X}$ (with distribution $P_{\boldsymbol{X}}$) and the output $\boldsymbol{Y}$ are column vectors of appropriate dimensions related by a Gaussian conditional probability density:

$$p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x};\mathsf{snr}) = (2\pi)^{-\frac{L}{2}} \exp\left[-\frac{1}{2}\left\|\boldsymbol{y} - \sqrt{\mathsf{snr}}\,\boldsymbol{H}\boldsymbol{x}\right\|^2\right], \tag{2.19}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector. Let the (weighted) MMSE be defined as the minimum error in estimating $\boldsymbol{H}\boldsymbol{X}$:

$$\mathsf{mmse}(\mathsf{snr}) = \mathsf{E}\left\{\left\|\boldsymbol{H}\,\boldsymbol{X} - \boldsymbol{H}\,\widehat{\boldsymbol{X}}(\boldsymbol{Y};\mathsf{snr})\right\|^2\right\}, \tag{2.20}$$

where $\widehat{\boldsymbol{X}}(\boldsymbol{Y};\mathsf{snr})$ is the conditional mean estimate. A generalization of Theorem 2.1 is the following:

Theorem 2.2 *Consider the vector model (2.18). For every $P_{\boldsymbol{X}}$ satisfying $\mathsf{E}\|\boldsymbol{X}\|^2 < \infty$,*

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I\left(\boldsymbol{X}; \sqrt{\mathsf{snr}}\,\boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{N}\right) = \frac{1}{2}\mathsf{mmse}\left(\boldsymbol{H}\boldsymbol{X} \mid \sqrt{\mathsf{snr}}\,\boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{N}\right). \tag{2.21}$$

Proof: See Section 2.2.3. ■

A verification of Theorem 2.2 in the special case of Gaussian input with positive definite covariance matrix $\boldsymbol{\Sigma}$ is straightforward. The covariance of the conditional mean estimation error is

$$\mathsf{E}\left\{\left(\boldsymbol{X} - \widehat{\boldsymbol{X}}\right)\left(\boldsymbol{X} - \widehat{\boldsymbol{X}}\right)^{\top}\right\} = \left(\boldsymbol{\Sigma}^{-1} + \mathsf{snr}\boldsymbol{H}^{\top}\boldsymbol{H}\right)^{-1}, \tag{2.22}$$

from which one can calculate the MMSE:

$$\mathsf{E}\left\{\left\|\boldsymbol{H}\left(\boldsymbol{X}-\widehat{\boldsymbol{X}}\right)\right\|^2\right\} = \mathsf{tr}\left\{\boldsymbol{H}\left(\boldsymbol{\Sigma}^{-1}+\mathsf{snr}\boldsymbol{H}^\top\boldsymbol{H}\right)^{-1}\boldsymbol{H}^\top\right\}. \tag{2.23}$$

The mutual information is [17, 103]:

$$I(\boldsymbol{X};\boldsymbol{Y}) = \frac{1}{2}\log\det\left(\boldsymbol{I}+\mathsf{snr}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{\Sigma}^{\frac{1}{2}}\right), \tag{2.24}$$

where $\boldsymbol{\Sigma}^{\frac{1}{2}}$ is the unique positive semi-definite symmetric matrix such that $\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^2 = \boldsymbol{\Sigma}$. Taking direct derivative of (2.24) leads to the desired result:[2]

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}}I(\boldsymbol{X};\boldsymbol{Y}) &= \frac{1}{2}\,\mathrm{tr}\left\{\left(\boldsymbol{I}+\mathsf{snr}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{\Sigma}^{\frac{1}{2}}\right\} &(2.25)\\ &= \frac{1}{2}\,\mathsf{E}\left\{\left\|\boldsymbol{H}\left(\boldsymbol{X}-\widehat{\boldsymbol{X}}\right)\right\|^2\right\}. &(2.26)\end{aligned}$$

The versions of Theorems 2.1 and 2.2 for complex-valued channel and signaling hold verbatim if each real/imaginary component of the circularly symmetric Gaussian noise N or $\boldsymbol{N}$ has unit variance, i.e., $\mathsf{E}\left\{\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}}\right\} = 2\boldsymbol{I}$. In particular, the factor of 1/2 in (2.14) and (2.21) remains intact. However, with the more common definition of snr in complex valued channels where the complex noise has real and imaginary components with variance 1/2 each, the factor of 1/2 in (2.14) and (2.21) disappears.

2.2.3 Proof via the SNR-Incremental Channel

The central relationship given by Theorems 2.1 and 2.2 can be proved in various, rather different, ways. In fact, five proofs are given in this thesis, including two direct proofs by taking derivative of the mutual information and a related information divergence respectively, a proof through the de Bruijn identity, and a proof taking advantage of results in the continuous-time domain. However, the most enlightening proof is by considering what we call an "incremental channel" and apply the chain rule for mutual information. A proof of Theorem 2.1 using this technique is given next, while its generalization to the vector version is omitted but straightforward. The alternative proofs are discussed in Section 2.2.6.

The key to the incremental-channel proof is to reduce the proof of the relationship for all SNRs to that for the special case of vanishing SNR, in which domain better is known about the mutual information:

Lemma 2.1 *As $\delta \to 0$, the input-output mutual information of the Gaussian channel:*

$$Y = \sqrt{\delta}\,Z + U, \tag{2.27}$$

where $\mathsf{E}Z^2 < \infty$ and $U \sim \mathcal{N}(0,1)$ is independent of Z, is given by

$$I(Y;Z) = \frac{\delta}{2}\,\mathsf{E}(Z-\mathsf{E}Z)^2 + o(\delta). \tag{2.28}$$

[2] The following identity is useful:

$$\frac{\partial \log\det \boldsymbol{Q}}{\partial x} = \mathsf{tr}\left\{\boldsymbol{Q}^{-1}\frac{\partial \boldsymbol{Q}}{\partial x}\right\}.$$

Essentially, Lemma 2.1 states that the mutual information is half the SNR times the variance of the input at the vicinity of zero SNR, but insensitive to the shape of the input distribution otherwise. Lemma 2.1 has been given in [61] and [104] (also implicitly in [111]). A proof is given here for completeness.

Proof: For any given input distribution P_Z, the mutual information, which is a conditional divergence, allows the following decomposition due to Verdú [111]:

$$I(Y;Z) = \mathsf{D}\left(P_{Y|Z} \| P_Y | P_Z\right) = \mathsf{D}\left(P_{Y|Z} \| P_{Y'} | P_Z\right) - \mathsf{D}\left(P_Y \| P_{Y'}\right), \tag{2.29}$$

where $P_{Y'}$ is an arbitrary distribution as long as the two divergences on the right hand side of (2.29) are well-defined. Choose Y' to be a Gaussian random variable with the same mean and variance as Y. Let the variance of Z be denoted by v. The probability density function associated with Y' is

$$p_{Y'}(y) = \frac{1}{\sqrt{2\pi(\delta v + 1)}} \exp\left[-\frac{y^2}{2(\delta v + 1)}\right]. \tag{2.30}$$

The first term on the right hand side of (2.29) is a divergence between two Gaussian distributions. Using a general formula [111]

$$\mathsf{D}\left(\mathcal{N}\left(m_1, \sigma_1^2\right) \| \mathcal{N}\left(m_0, \sigma_0^2\right)\right) = \frac{1}{2} \log \frac{\sigma_0^2}{\sigma_1^2} + \frac{1}{2}\left(\frac{(m_1 - m_0)^2}{\sigma_0^2} + \frac{\sigma_1^2}{\sigma_0^2} - 1\right) \log e, \tag{2.31}$$

the interested divergence can be easily found as

$$\frac{1}{2}\log(1 + \delta v) = \frac{\delta v}{2} + o(\delta). \tag{2.32}$$

The unconditional output distribution can be expressed as

$$p_Y(y) = \frac{1}{\sqrt{2\pi}} \mathsf{E}\left\{\exp\left[-\frac{1}{2}\left(y - \sqrt{\delta}\, Z\right)^2\right]\right\}. \tag{2.33}$$

By (2.30) and (2.33),

$$\begin{aligned}
\log \frac{p_Y(y)}{p_{Y'}(y)} &= \frac{1}{2}\log(1 + \delta v) + \log \mathsf{E}\left\{\exp\left[\frac{(y - \sqrt{\delta}\,\mathsf{E}Z)^2}{2(\delta v + 1)} - \frac{1}{2}(y - \sqrt{\delta}\, Z)^2\right]\right\} && (2.34)\\
&= \frac{1}{2}\log(1 + \delta v) \\
&\quad + \log \mathsf{E}\left\{\exp\left[\sqrt{\delta}\, y(Z - \mathsf{E}Z) - \frac{\delta}{2}\left(v y^2 + Z^2 - (\mathsf{E}Z)^2\right) + o(\delta)\right]\right\} && (2.35)\\
&= \frac{1}{2}\log(1 + \delta v) + \log \mathsf{E}\left\{1 + \sqrt{\delta}\, y(Z - \mathsf{E}Z)\right. \\
&\quad \left. + \frac{\delta}{2}\left(y^2(Z - \mathsf{E}Z)^2 - v y^2 - Z^2 + (\mathsf{E}Z)^2\right) + o(\delta)\right\} && (2.36)\\
&= \frac{1}{2}\log(1 + \delta v) + \log\left(1 - \frac{\delta v}{2}\right) + o(\delta) && (2.37)\\
&= o(\delta), && (2.38)
\end{aligned}$$

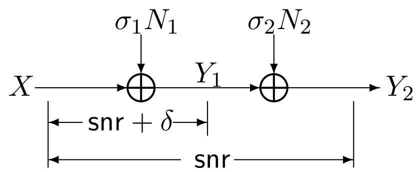

Figure 2.2: An SNR-incremental Gaussian channel.

where the limit $\delta \to 0$ and the expectation can be exchanged in (2.37) as long as $\mathsf{E}Z^2 < \infty$ due to Lebesgue convergence theorem [83]. Therefore, the second divergence on the right hand side of (2.29) is $o(\delta)$. Lemma 2.1 is immediate:

$$I(Y;Z) = \frac{\delta v}{2} + o(\delta). \tag{2.39}$$

■

It is interesting to note that the proof relies on the fact that the divergence between the output distributions of a Gaussian channel under different input distributions is sublinear in the SNR when the noise dominates.

Lemma 2.1 is the special case of Theorem 2.1 at vanishing SNR, which, by means of the incremental-channel method, can be bootstrapped to a proof of Theorem 2.1 for all SNRs.

Proof: [Theorem 2.1] Fix arbitrary $\mathsf{snr} > 0$ and $\delta > 0$. Consider a cascade of two Gaussian channels as depicted in Figure 2.2:

$$\begin{aligned} Y_1 &= X + \sigma_1 N_1, && (2.40a)\\ Y_2 &= Y_1 + \sigma_2 N_2, && (2.40b)\end{aligned}$$

where X is the input, and N_1 and N_2 are independent standard Gaussian random variables. Let σ_1 and σ_2 satisfy:

$$\begin{aligned} \mathsf{snr} + \delta &= \frac{1}{\sigma_1^2}, && (2.41a)\\ \mathsf{snr} &= \frac{1}{\sigma_1^2 + \sigma_2^2}, && (2.41b)\end{aligned}$$

so that the SNR of the first channel (2.40a) is $\mathsf{snr} + \delta$ and that of the composite channel is snr. Such a channel is referred to as an *SNR-incremental Gaussian channel* since the signal-to-noise ratio increases by δ from Y_2 to Y_1. Here we choose to scale the noise for obvious reason.

Since the mutual information vanishes trivially at zero SNR, Theorem 2.1 is equivalent to the following:

$$\begin{aligned} I(X;Y_1) - I(X;Y_2) &= I(\mathsf{snr} + \delta) - I(\mathsf{snr}) && (2.42)\\ &= \frac{\delta}{2}\,\mathsf{mmse}(\mathsf{snr}) + o(\delta). && (2.43)\end{aligned}$$

Noting that X—Y_1—Y_2 is a Markov chain, one has

$$\begin{aligned} I(X;Y_1) - I(X;Y_2) &= I(X;Y_1,Y_2) - I(X;Y_2) && (2.44)\\ &= I(X;Y_1|Y_2), && (2.45)\end{aligned}$$

where (2.45) is by the chain rule for information [17]. Given X, the outputs Y_1 and Y_2 are jointly Gaussian. Hence Y_1 is Gaussian conditioned on X and Y_2. Using (2.40), it is easy to check that

$$(\mathsf{snr}+\delta)\,Y_1-\mathsf{snr}\,Y_2-\delta\,X=\delta\,\sigma_1\,N_1-\mathsf{snr}\,\sigma_2\,N_2. \tag{2.46}$$

Let

$$N=\frac{1}{\sqrt{\delta}}\,(\delta\,\sigma_1\,N_1-\mathsf{snr}\,\sigma_2\,N_2). \tag{2.47}$$

Then N is a standard Gaussian random variable due to (2.41). Given X, N is independent of Y_2 since, by (2.40) and (2.41),

$$\mathsf{E}\left\{N\,Y_2\mid X\right\}=\frac{1}{\sqrt{\delta}}\,\left(\delta\,\sigma_1^2-\mathsf{snr}\,\sigma_2^2\right)=0. \tag{2.48}$$

Therefore, (2.46) is tantamount to

$$(\mathsf{snr}+\delta)\,Y_1=\mathsf{snr}\,Y_2+\delta\,X+\sqrt{\delta}\,N, \tag{2.49}$$

where $N\sim\mathcal{N}(0,1)$ is independent of X and Y_2. Clearly,

$$I(X;Y_1|Y_2)=I\left(X;\delta\,X+\sqrt{\delta}\,N\;\middle|\;Y_2\right). \tag{2.50}$$

Hence given Y_2, (2.49) is equivalent to a Gaussian channel with its SNR equal to δ where the input distribution is $P_{X|Y_2}$. Applying Lemma 2.1 to the Gaussian channel (2.49) conditioned on $Y_2=y_2$, one obtains

$$I(X;Y_1|Y_2=y_2)=\frac{\delta}{2}\,\mathsf{E}\left\{(X-\mathsf{E}\left\{X\mid Y_2\right\})^2\;\middle|\;Y_2=y_2\right\}+o(\delta). \tag{2.51}$$

Taking the expectation over Y_2 on both sides of (2.51), one has

$$I(X;Y_1|Y_2)=\frac{\delta}{2}\,\mathsf{E}\left\{(X-\mathsf{E}\left\{X\mid Y_2\right\})^2\right\}+o(\delta), \tag{2.52}$$

which establishes (2.43) by (2.44) together with the fact that

$$\mathsf{E}\left\{(X-\mathsf{E}\left\{X\mid Y_2\right\})^2\right\}=\mathsf{mmse}(\mathsf{snr}). \tag{2.53}$$

Hence the proof of Theorem 2.1. ∎

2.2.4 Discussions

Mutual Information Chain Rule

Underlying the incremental-channel proof of Theorem 2.1 is the chain rule for information:

$$I(X;Y_1,\dots,Y_n)=\sum_{i=1}^{n}I(X;Y_i\,|\,Y_{i+1},\dots,Y_n). \tag{2.54}$$

In case that X—Y_1—…—Y_n is a Markov chain, (2.54) becomes

$$I(X;Y_1)=\sum_{i=1}^{n}I(X;Y_i\,|\,Y_{i+1}), \tag{2.55}$$

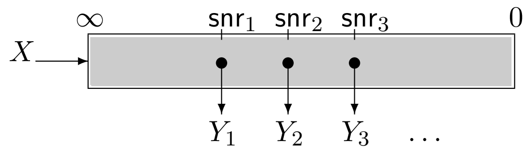

Figure 2.3: A Gaussian pipe where noise is added gradually.

where we let $Y_{n+1} \equiv 0$. This applies to the train of outputs tapped from a Gaussian pipe where noise is added gradually until the SNR vanishes as depicted in Figure 2.3. The sum in (2.55) converges to an integral as Y_i becomes a finer and finer sequence of Gaussian channel outputs by noticing from (2.52) that each conditional mutual information in (2.55) is that of a low-SNR channel and is essentially proportional to the MMSE times the SNR increment. This viewpoint leads us to an equivalent form of Theorem 2.1:

$$I(\mathsf{snr}) = \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma. \tag{2.56}$$

Therefore, the mutual information can be regarded as an accumulation of the MMSE as a function of the SNR, as is illustrated by the curves in Figure 2.1.

The infinite divisibility of Gaussian distributions, namely, the fact that a Gaussian random variable can always be decomposed as the sum of independent Gaussian random variables of smaller variances, is crucial in establishing the incremental channel (or, the Markov chain). This property enables us to study the mutual information increase due to an infinitesimal increase in the SNR, and henceforth obtain the integral equation (2.14) in Theorem 2.1.

Derivative of the Divergence

Consider an input-output pair (X, Y) connected through (2.5). The mutual information $I(X;Y)$ is the average value over the input X of a divergence:

$$\mathsf{D}\left(P_{Y|X=x} \| P_Y\right) = \int \log \frac{\mathrm{d}P_{Y|X=x}(y)}{\mathrm{d}P_Y(y)}\, \mathrm{d}P_{Y|X=x}(y). \tag{2.57}$$

Refining Theorem 2.1, it is possible to directly obtain the derivative of the divergence given any value of the input:

Theorem 2.3 *Consider the channel (2.5). For every input distribution P_X that satisfies $\mathsf{E}X^2 < \infty$,*

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} \mathsf{D}\left(P_{Y|X=x} \| P_Y\right) = \frac{1}{2} \mathsf{E}\left\{ (X - X')^2 \,\middle|\, X = x \right\} - \frac{1}{2\sqrt{\mathsf{snr}}} \mathsf{E}\left\{ X' N \,\middle|\, X = x \right\}, \tag{2.58}$$

where X' is an auxiliary random variable which is i.i.d. with X conditioned on Y.

Proof: See [45]. ∎

The auxiliary random variable X' has an interesting physical meaning. It can be regarded as the output of the so-called "retrochannel" (see also Section 3.2.4), which takes Y as the input and generates a random variable according to the posterior probability distribution $p_{X|Y;\mathsf{snr}}$. Using Theorem 2.3, Theorem 2.1 can be recovered by taking expectation on both sides of (2.58). The left hand side becomes the derivative of the mutual information. The right hand side becomes 1/2 times the following:

$$\frac{1}{\sqrt{\mathsf{snr}}}\mathsf{E}\left\{(X-X')(Y-\sqrt{\mathsf{snr}}X')\right\} = \frac{1}{\sqrt{\mathsf{snr}}}\mathsf{E}\left\{XY-X'Y\right\}+\mathsf{E}\left\{(X')^2-XX'\right\}. \quad (2.59)$$

Since conditioned on Y, X' and X are i.i.d., (2.59) can be further written as

$$\begin{aligned}\mathsf{E}\left\{X^2-XX'\right\} &= \mathsf{E}\left\{X^2-\mathsf{E}\left\{XX' \mid Y;\mathsf{snr}\right\}\right\} && (2.60)\\ &= \mathsf{E}\left\{X^2-\left(\mathsf{E}\left\{X\mid Y;\mathsf{snr}\right\}\right)^2\right\}, && (2.61)\end{aligned}$$

which is the MMSE.

Multiuser Channel

A multiuser system in which users may be received at different SNRs can be better modelled by:

$$\boldsymbol{Y} = \boldsymbol{H}\,\boldsymbol{\Gamma}\,\boldsymbol{X} + \boldsymbol{N} \quad (2.62)$$

where $\boldsymbol{H}$ is a deterministic $L\times K$ matrix, $\boldsymbol{\Gamma} = \mathrm{diag}\{\sqrt{\mathsf{snr}_1},\dots,\sqrt{\mathsf{snr}_K}\}$ consists of the square-root of the SNRs of the K users, and $\boldsymbol{N}$ consists of i.i.d. standard Gaussian entries. The following theorem addresses the derivative of the total mutual information with respect to an individual user's SNR:

Theorem 2.4 *For every input distribution $P_{\boldsymbol{X}}$ that satisfies $\mathsf{E}\|\boldsymbol{X}\|^2<\infty$,*

$$\frac{\partial}{\partial \mathsf{snr}_k} I(\boldsymbol{X};\boldsymbol{Y}) = \frac{1}{2}\sum_{i=1}^{K}\sqrt{\frac{\mathsf{snr}_i}{\mathsf{snr}_k}}\left[\boldsymbol{H}^\top\boldsymbol{H}\right]_{ki}\mathsf{E}\left\{\mathsf{Cov}\left\{X_k,X_i|\boldsymbol{Y};\boldsymbol{\Gamma}\right\}\right\}, \quad (2.63)$$

where $\mathsf{Cov}\{\cdot,\cdot|\cdot\}$ *denotes conditional covariance.*

Proof: The proof follows straightforwardly that of Theorem 2.2 in Appendix A.2 and is omitted. ■

Using Theorem 2.4, Theorem 2.1 can be easily recovered by setting $K=1$ and $\boldsymbol{\Gamma}=\sqrt{\mathsf{snr}}$, since

$$\mathsf{E}\left\{\mathsf{Cov}\left\{X,X|Y;\mathsf{snr}\right\}\right\} = \mathsf{E}\left\{\mathsf{var}\left\{X|Y;\mathsf{snr}\right\}\right\} \quad (2.64)$$

is exactly the MMSE. Theorem 2.2 can also be recovered by letting $\mathsf{snr}_k=\mathsf{snr}$ for all k. Then,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\boldsymbol{X};\boldsymbol{Y}) &= \sum_{k=1}^{K}\frac{\partial}{\partial\mathsf{snr}_k} I(\boldsymbol{X};\boldsymbol{Y}) && (2.65)\\ &= \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{K}\left[\boldsymbol{H}^\top\boldsymbol{H}\right]_{ki}\mathsf{E}\left\{\mathsf{Cov}\left\{X_k,X_i|\boldsymbol{Y};\boldsymbol{\Gamma}\right\}\right\} && (2.66)\\ &= \frac{1}{2}\mathsf{E}\left\{\left\|\boldsymbol{H}\,\boldsymbol{X}-\boldsymbol{H}\,\mathsf{E}\left\{\boldsymbol{X}\mid\boldsymbol{Y};\boldsymbol{\Gamma}\right\}\right\|^2\right\}. && (2.67)\end{aligned}$$

2.2.5 Some Applications of Theorems 2.1 and 2.2

Extremality of Gaussian Inputs

Gaussian inputs are most favorable for Gaussian channels in information-theoretic sense that they maximize mutual information for a given power; on the other hand they are least favorable in estimation-theoretic sense that they maximize MMSE for a given power. These well-known results are seen to be immediately equivalent through Theorem 2.1 (or Theorem 2.2 for the vector case). This also points to a simple proof of the result that Gaussian input is capacity-achieving by showing that the linear estimation upper bound for the MMSE is achieved for Gaussian inputs.

Proof of De Bruijn's Identity

An interesting observation here is that Theorem 2.2 is equivalent to the (multivariate) de Bruijn identity [90, 16]:

$$\frac{\mathrm{d}}{\mathrm{d}t} h\left(\boldsymbol{H}\boldsymbol{X} + \sqrt{t}\,\boldsymbol{N}\right) = \frac{1}{2}\mathsf{tr}\left\{\boldsymbol{J}\left(\boldsymbol{H}\boldsymbol{X} + \sqrt{t}\,\boldsymbol{N}\right)\right\} \tag{2.68}$$

where $h(\cdot)$ stands for the differential entropy and $\boldsymbol{J}(\cdot)$ for Fisher's information matrix [76], which is defined as[3]

$$\boldsymbol{J}(\boldsymbol{y}) = \mathsf{E}\left\{[\nabla \log p_{\boldsymbol{Y}}(\boldsymbol{y})]\,[\nabla \log p_{\boldsymbol{Y}}(\boldsymbol{y})]^{\top}\right\}, \tag{2.69}$$

where the gradient with respect to a vector is defined as $\nabla = \left[\frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_L}\right]^{\top}$. Let $\mathsf{snr} = 1/t$ and $\boldsymbol{Y} = \sqrt{\mathsf{snr}}\,\boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{N}$. Then

$$\begin{aligned} h\left(\boldsymbol{H}\boldsymbol{X} + \sqrt{t}\,\boldsymbol{N}\right) &= h(\boldsymbol{Y}) - \frac{L}{2}\log \mathsf{snr} && (2.70) \\ &= I(\boldsymbol{X};\boldsymbol{Y}) - \frac{L}{2}\log\frac{\mathsf{snr}}{2\pi e}. && (2.71) \end{aligned}$$

In the meantime,

$$\boldsymbol{J}\left(\boldsymbol{H}\boldsymbol{X} + \sqrt{t}\,\boldsymbol{N}\right) = \mathsf{snr}\,\boldsymbol{J}(\boldsymbol{Y}). \tag{2.72}$$

Note that

$$p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y};\mathsf{snr}) = \mathsf{E}\left\{p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X};\mathsf{snr})\right\}, \tag{2.73}$$

where $p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x};\mathsf{snr})$ is a Gaussian density (2.19). It can be shown that

$$\nabla \log p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y};\mathsf{snr}) = \sqrt{\mathsf{snr}}\,\boldsymbol{H}\widehat{\boldsymbol{X}}(\boldsymbol{y};\mathsf{snr}) - \boldsymbol{y}. \tag{2.74}$$

Plugging (2.74) into (2.72) and (2.69) gives

$$\boldsymbol{J}(\boldsymbol{Y}) = \boldsymbol{I} - \mathsf{snr}\,\boldsymbol{H}\,\mathsf{E}\left\{\left(\boldsymbol{X} - \widehat{\boldsymbol{X}}\right)\left(\boldsymbol{X} - \widehat{\boldsymbol{X}}\right)^{\top}\right\}\boldsymbol{H}^{\top}. \tag{2.75}$$

Now de Bruijn's identity (2.68) and Theorem 2.2 prove each other by (2.71) and (2.75). Noting this equivalence, the incremental-channel approach offers an intuitive alternative to the conventional proof of de Bruijn's identity obtained by integrating by parts (e.g., [17]).

[3] The gradient operator can be regarded as $\nabla = \left[\frac{\partial}{\partial y_1}, \cdots, \frac{\partial}{\partial y_L}\right]^{\top}$. For any differentiable function $f : \mathcal{R}^L \to \mathcal{R}$, its gradient at any $\boldsymbol{y}$ is a column vector $\nabla f(\boldsymbol{y}) = \left[\frac{\partial f}{\partial y_1}(\boldsymbol{y}), \cdots, \frac{\partial f}{\partial y_L}(\boldsymbol{y})\right]^{\top}$.

The inverse of Fisher's information is in general a lower bound on estimation accuracy, a result known as the Cramér-Rao lower bound [76]. For Gaussian channels, Fisher's information matrix and the covariance of conditional mean estimation error determine each other in a simple way (2.75). In particular, for a scalar channel,

$$J\left(\sqrt{\mathsf{snr}}\,X+N\right)=1-\mathsf{snr}\cdot\mathsf{mmse}(\mathsf{snr}). \tag{2.76}$$

Joint and Separate Decoding Capacities

Theorem 2.1 is the key to show a relationship between the mutual informations of multiuser channels under joint and separate decoding. This will be relegated to Section 3.2.4 in the chapter on multiuser channels.

2.2.6 Alternative Proofs of Theorems 2.1 and 2.2

The incremental-channel proof of Theorem 2.1 provides much information-theoretic insight into the result. In this subsection, we give an alternative proof of Theorem 2.2, which is a distilled version of the more general result of Zakai [120] (follow-up to this work) that uses the Malliavin calculus and shows that the central relationship between the mutual information and estimation error holds in the abstract Wiener space. This alternative approach of Zakai makes use of relationships between conditional mean estimation and likelihood ratios due to Esposito [27] and Hatsell and Nolte [46].

As mentioned earlier, the central theorems also admit several other alternative proofs. In fact, a third proof using de Bruijn's identity is already evident in Section 2.2.5. A fourth proof taking advantage of results in the continuous-time domain is relegated to Section 2.4. A fifth proof of Theorems 2.1 and 2.2 by taking the derivative of the mutual information is given in Appendices A.1 and A.2 respectively.

It suffices to prove Theorem 2.2 assuming $\boldsymbol{H}$ to be the identity matrix since one can always regard $\boldsymbol{HX}$ as the input. Let $\boldsymbol{Z}=\sqrt{\mathsf{snr}}\,\boldsymbol{X}$. Then the channel (2.18) is represented by the canonical L-dimensional Gaussian channel:

$$\boldsymbol{Y}=\boldsymbol{Z}+\boldsymbol{N}. \tag{2.77}$$

By Verdú's formula (2.29), the mutual information can be expressed in the divergence between the unconditional output distribution and the noise distribution:

$$\begin{aligned} I(\boldsymbol{Y};\boldsymbol{Z}) &= \mathsf{D}\left(P_{\boldsymbol{Y}|\boldsymbol{Z}}\|P_{\boldsymbol{N}}|P_{\boldsymbol{Z}}\right)-\mathsf{D}\left(P_{\boldsymbol{Y}}\|P_{\boldsymbol{N}}\right) &\quad (2.78)\\ &= \frac{1}{2}\mathsf{E}\|\boldsymbol{Z}\|^2-\mathsf{D}\left(P_{\boldsymbol{Y}}\|P_{\boldsymbol{N}}\right). &\quad (2.79)\end{aligned}$$

Hence Theorem 2.2 is equivalent to the following:

Theorem 2.5 *For every $P_{\boldsymbol{X}}$ satisfying* $\mathsf{E}\|\boldsymbol{X}\|^2<\infty$,

$$\frac{\mathrm{d}}{\mathsf{dsnr}}\mathsf{D}\left(P_{\sqrt{\mathsf{snr}}\,\boldsymbol{X}+\boldsymbol{N}}\|P_{\boldsymbol{N}}\right)=\frac{1}{2}\mathsf{E}\left\{\left\|\mathsf{E}\left\{\boldsymbol{X}\mid\sqrt{\mathsf{snr}}\,\boldsymbol{X}+\boldsymbol{N}\right\}\right\|^2\right\}. \tag{2.80}$$

It is clear that, $p_{\boldsymbol{Y}}$, the probability density function for the channel output exists. The likelihood ratio between two hypotheses, one with the input signal $\boldsymbol{Z}$ and the other with zero input, is given by

$$l(\boldsymbol{y})=\frac{p_{\boldsymbol{Y}}(\boldsymbol{y})}{p_{\boldsymbol{N}}(\boldsymbol{y})}. \tag{2.81}$$

Theorem 2.5 can be proved using some geometric properties of the above likelihood ratio. The following lemmas are important steps.

Lemma 2.2 (Esposito [27]) *The gradient of the log-likelihood ratio is equal to the conditional mean estimate:*

$$\nabla \log l(\boldsymbol{y}) = \mathsf{E}\left\{ \boldsymbol{Z} \mid \boldsymbol{Y} = \boldsymbol{y} \right\}. \tag{2.82}$$

Lemma 2.3 (Hatsell and Nolte [46]) *The log-likelihood ratio satisfies Poisson's equation:*[4]

$$\nabla^2 \log l(\boldsymbol{y}) = \mathsf{E}\left\{ \|\boldsymbol{Z}\|^2 \,\middle|\, \boldsymbol{Y} = \boldsymbol{y} \right\} - \left\|\mathsf{E}\left\{ \boldsymbol{Z} \mid \boldsymbol{Y} = \boldsymbol{y} \right\}\right\|^2. \tag{2.83}$$

From Lemmas 2.2 and 2.3,

$$\begin{aligned} \mathsf{E}\left\{ \|\boldsymbol{Z}\|^2 \,\middle|\, \boldsymbol{Y} = \boldsymbol{y} \right\} &= \nabla^2 \log l(\boldsymbol{y}) + \|\nabla \log l(\boldsymbol{y})\|^2 & (2.84)\\ &= \frac{l(y)\nabla^2 \log l(\boldsymbol{y}) - \|\nabla l(\boldsymbol{y})\|^2 + \|\nabla l(\boldsymbol{y})\|^2}{l^2(\boldsymbol{y})}. & (2.85) \end{aligned}$$

Thus we have proved

Lemma 2.4

$$\mathsf{E}\left\{ \|\boldsymbol{Z}\|^2 \,\middle|\, \boldsymbol{Y} = \boldsymbol{y} \right\} = \frac{\nabla^2 l(\boldsymbol{y})}{l(\boldsymbol{y})}. \tag{2.86}$$

A proof of Theorem 2.5 is obtained by taking the derivative directly.

Proof: [Theorem 2.5] Note that the likelihood ratio can be expressed as

$$\begin{aligned} l(\boldsymbol{y}) &= \frac{\mathsf{E}\left\{ p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X}) \right\}}{p_{\boldsymbol{N}}(\boldsymbol{y})} & (2.87)\\ &= \mathsf{E}\left\{ \exp\left[\sqrt{\mathsf{snr}}\, \boldsymbol{y}^\top \boldsymbol{X} - \frac{\mathsf{snr}}{2}\|\boldsymbol{X}\|^2 \right] \right\}. & (2.88) \end{aligned}$$

Hence,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} l(\boldsymbol{y}) &= \frac{1}{2}\mathsf{E}\left\{ \left(\frac{1}{\sqrt{\mathsf{snr}}}\boldsymbol{y}^\top \boldsymbol{X} - \|\boldsymbol{X}\|^2 \right) \exp\left[\sqrt{\mathsf{snr}}\, \boldsymbol{y}^\top \boldsymbol{X} - \frac{\mathsf{snr}}{2}\|\boldsymbol{X}\|^2 \right] \right\} & (2.89)\\ &= \frac{1}{2} l(\boldsymbol{y}) \left[\frac{1}{\sqrt{\mathsf{snr}}}\boldsymbol{y}^\top \mathsf{E}\left\{ \boldsymbol{X} \mid \boldsymbol{Y} = \boldsymbol{y} \right\} - \mathsf{E}\left\{ \|\boldsymbol{X}\|^2 \,\middle|\, \boldsymbol{Y} = \boldsymbol{y} \right\} \right] & (2.90)\\ &= \frac{1}{2\mathsf{snr}} \left[l(\boldsymbol{y})\, \boldsymbol{y}^\top \nabla \log l(\boldsymbol{y}) - \nabla^2 \log l(\boldsymbol{y}) \right]. & (2.91) \end{aligned}$$

Note that the order of expectation with respect to $P_{\boldsymbol{X}}$ and the derivative with respect to the SNR can be exchanged as long as the input has finite power. This is essentially guaranteed by Lemma A.1 in Appendix A.2.

The divergence can be written as

$$\begin{aligned} \mathsf{D}\left(P_{\boldsymbol{Y}} \| P_{\boldsymbol{N}} \right) &= \int p_{\boldsymbol{Y}}(\boldsymbol{y}) \log \frac{p_{\boldsymbol{Y}}(\boldsymbol{y})}{p_{\boldsymbol{N}}(\boldsymbol{y})}\, \mathrm{d}\boldsymbol{y} & (2.92)\\ &= \mathsf{E}\left\{ l(\boldsymbol{N}) \log l(\boldsymbol{N}) \right\}, & (2.93) \end{aligned}$$

[4] For any differentiable $\boldsymbol{f} : \mathcal{R}^L \to \mathcal{R}^L$, $\nabla \cdot \boldsymbol{f} = \sum_{l=1}^{L} \frac{\partial f_l}{\partial y_l}$. Also, if $\boldsymbol{f}$ is doubly differentiable, its Laplacian is defined as $\nabla^2 f = \nabla \cdot (\nabla f) = \sum_{l=1}^{L} \frac{\partial^2 f}{\partial y_l^2}$.

and its derivative

$$\frac{\mathrm{d}}{\mathsf{dsnr}}\mathsf{D}\left(P_{\boldsymbol{Y}}\|P_{\boldsymbol{N}}\right)=\mathsf{E}\left\{\log l(\boldsymbol{N})\frac{\mathrm{d}}{\mathsf{dsnr}}l(\boldsymbol{N})\right\}. \tag{2.94}$$

Again, the derivative and expectation can be exchanged in order by the same argument as in the above. By (2.91), the derivative (2.94) can be evaluated as

$$\begin{aligned}
& \frac{1}{2\mathsf{snr}}\mathsf{E}\left\{l(\boldsymbol{N})\log l(\boldsymbol{N})\,\boldsymbol{N}^\top\nabla\log l(\boldsymbol{N})\right\}-\frac{1}{2\mathsf{snr}}\mathsf{E}\left\{\log l(\boldsymbol{N})\,\nabla^2 l(\boldsymbol{N})\right\} \\
=\ & \frac{1}{2\mathsf{snr}}\mathsf{E}\left\{\nabla\cdot\left[l(\boldsymbol{N})\log l(\boldsymbol{N})\nabla\log l(\boldsymbol{N})\right]-\log l(\boldsymbol{N})\,\nabla^2 l(\boldsymbol{N})\right\} && (2.95)\\
=\ & \frac{1}{2\mathsf{snr}}\mathsf{E}\left\{l(\boldsymbol{N})\,\|\nabla\log l(\boldsymbol{N})\|^2\right\} && (2.96)\\
=\ & \frac{1}{2\mathsf{snr}}\mathsf{E}\,\|\nabla\log l(\boldsymbol{Y})\|^2 && (2.97)\\
=\ & \frac{1}{2}\mathsf{E}\|\mathsf{E}\left\{\boldsymbol{X}\mid\boldsymbol{Y}\right\}\|^2, && (2.98)
\end{aligned}$$

where to write (2.95) one also needs the following result which can be proved easily by integration by parts:

$$\mathsf{E}\left\{\boldsymbol{N}^\top\boldsymbol{f}(\boldsymbol{N})\right\}=\mathsf{E}\left\{\nabla\cdot\boldsymbol{f}(\boldsymbol{N})\right\} \tag{2.99}$$

for all $\boldsymbol{f}:\mathcal{R}^L\to\mathcal{R}^L$ that satisfies $f_i(\boldsymbol{n})e^{-\frac{1}{2}n_i^2}\to 0$ as $n_i\to\infty$. ■

2.2.7 Asymptotics of Mutual Information and MMSE

It can be shown that the mutual information and MMSE are both differentiable functions of the SNR given any finite-power input. In the following, the asymptotics of the mutual information and MMSE at low and high SNRs are studied mainly for the scalar Gaussian channel.

Low-SNR Asymptotics

Using the dominated convergence theorem, one can prove continuity of the MMSE estimate:

$$\lim_{\mathsf{snr}\to 0}\mathsf{E}\left\{X\mid Y;\mathsf{snr}\right\}=\mathsf{E}X, \tag{2.100}$$

and hence

$$\lim_{\mathsf{snr}\to 0}\mathsf{mmse}(\mathsf{snr})=\mathsf{mmse}(0)=\sigma_X^2 \tag{2.101}$$

where σ_X^2 is the input variance. It has been shown in [104] that symmetric (proper-complex in the complex case) signaling is second-order optimal. Indeed, for any real-valued symmetric input with unit variance, the mutual information can be expressed as

$$I(\mathsf{snr})=\frac{1}{2}\mathsf{snr}-\frac{1}{4}\mathsf{snr}^2+o(\mathsf{snr}^2). \tag{2.102}$$

A more refined study of the asymptotics is possible by examining the Taylor expansion of the following:

$$p_i(y;\mathsf{snr})=\mathsf{E}\left\{X^i\,p_{Y|X;\mathsf{snr}}(y\mid X;\mathsf{snr})\right\}, \tag{2.103}$$

which is well-defined at least for $i = 1, 2$, and in case all moments of the input are finite, it is well-defined for all i. Clearly, the unconditional probability density function is a special case:

$$p_{Y;\mathsf{snr}}(y;\mathsf{snr}) = p_0(y;\mathsf{snr}). \tag{2.104}$$

As $\mathsf{snr} \to 0$,

$$\begin{aligned} p_i(y;\mathsf{snr}) =& \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \,\mathsf{E}\left\{ X^i \left[1 + yX\mathsf{snr}^{\frac{1}{2}} + \frac{1}{2}(y^2-1)X^2\mathsf{snr} \right.\right. \\ &\left.\left. + \ \frac{1}{6}(y^2-3)yX^3\mathsf{snr}^{\frac{3}{2}} + \frac{1}{24}(y^4-6y^2+3)X^4\mathsf{snr}^2 + \mathcal{O}\left(\mathsf{snr}^{\frac{5}{2}}\right) \right] \right\}. \end{aligned} \tag{2.105}$$

Without loss of generality, it is assumed that the input has zero mean and unit variance. For convenience, it is also assumed that the input distribution is symmetric, i.e., X and $-X$ are identically distributed. In this case, the odd moments of X vanish and by (2.105),

$$p_{Y;\mathsf{snr}}(y;\mathsf{snr}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \left[1 + \frac{1}{2}(y^2-1)\mathsf{snr} + \frac{1}{24}(y^4-6y^2+3)\mathsf{E}X^4\mathsf{snr}^2 + \mathcal{O}\left(\mathsf{snr}^{\frac{5}{2}}\right) \right], \tag{2.106}$$

and

$$p_1(y;\mathsf{snr}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} y \left[\mathsf{snr}^{\frac{1}{2}} + \frac{1}{6}(y^2-3)\mathsf{E}X^4\mathsf{snr}^{\frac{3}{2}} + \mathcal{O}\left(\mathsf{snr}^{\frac{5}{2}}\right) \right]. \tag{2.107}$$

Thus, the conditional mean estimate is

$$\begin{aligned} \mathsf{E}\{X|Y=y;\mathsf{snr}\} &= \frac{p_1(y;\mathsf{snr})}{p_{Y;\mathsf{snr}}(y;\mathsf{snr})} && (2.108) \\ &= \sqrt{\mathsf{snr}}\, y \left[1 + \left(1 - \mathsf{E}X^4 - y^2 + \frac{1}{3}y^2\mathsf{E}X^4 \right) \frac{\mathsf{snr}}{2} + \mathcal{O}(\mathsf{snr}^2) \right]. && (2.109) \end{aligned}$$

Using (2.109), a finer characterization of the MMSE than (2.101) is obtained by definition (2.8) as

$$\mathsf{mmse}(\mathsf{snr}) = 1 - \mathsf{snr} + \left(3 - \frac{2}{3}\mathsf{E}X^4 \right) \mathsf{snr}^2 + \mathcal{O}\left(\mathsf{snr}^3\right). \tag{2.110}$$

Note that the expression (2.102) for the mutual information can also be refined either by noting that

$$I(\mathsf{snr}) = -\frac{1}{2}\log(2\pi e) - \mathsf{E}\left\{ \log p_{Y;\mathsf{snr}}(Y;\mathsf{snr}) \right\}, \tag{2.111}$$

and using (2.106), or integrating both sides of (2.110) and invoking Theorem 2.1:

$$I(\mathsf{snr}) = \frac{1}{2}\mathsf{snr} - \frac{1}{4}\mathsf{snr}^2 + \left(\frac{1}{2} - \frac{1}{9}\mathsf{E}X^4 \right) \mathsf{snr}^3 + \mathcal{O}\left(\mathsf{snr}^4\right). \tag{2.112}$$

The smoothness of the mutual information and MMSE carries over to the vector channel model (2.18) for finite-power inputs. The asymptotics also have their counterparts. The MMSE of the real-valued vector channel (2.18) is obtained as:

$$\mathsf{mmse}\left(\boldsymbol{H}\boldsymbol{X} \,|\, \sqrt{\mathsf{snr}}\,\boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{N} \right) = \mathsf{tr}\left\{ \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top \right\} - \mathsf{snr}\cdot\mathsf{tr}\left\{ \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top \right\} + \mathcal{O}(\mathsf{snr}^2) \tag{2.113}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the input vector. The input-output mutual information is (see [77]):

$$I\left(\boldsymbol{X}; \sqrt{\mathsf{snr}}\,\boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{N} \right) = \frac{\mathsf{snr}}{2}\mathsf{tr}\left\{ \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top \right\} - \frac{\mathsf{snr}^2}{4}\mathsf{tr}\left\{ \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top \right\} + \mathcal{O}(\mathsf{snr}^3). \tag{2.114}$$

The asymptotics can be refined to any order of the SNR following the above analysis.

High-SNR Asymptotics

At high SNRs, the mutual information does not grow without bound for finite-alphabet inputs such as the binary one (2.17), whereas it can increase at the speed of $\frac{1}{2}\log \mathsf{snr}$ for Gaussian inputs. Using the entropy power inequality [17], the mutual information of the scalar channel given any symmetric input distribution with a density is shown to be bounded:

$$\frac{1}{2}\log(1+\alpha\,\mathsf{snr}) \le I(\mathsf{snr}) \le \frac{1}{2}\log(1+\mathsf{snr}), \tag{2.115}$$

for some $\alpha \in (0,1]$.

The MMSE behavior at high SNR depends on the input distribution. The decay can be as low as $1/\mathsf{snr}$ for Gaussian input, whereas for binary input, the MMSE can also be easily shown to be exponentially small. In fact, for binary equiprobable inputs, the MMSE given by (2.16) allows another representation:

$$\mathsf{mmse}(\mathsf{snr}) = \mathsf{E}\left\{\frac{2}{\exp\left[2(\mathsf{snr}-\sqrt{\mathsf{snr}}\,Y)\right]+1}\right\} \tag{2.116}$$

where $Y \sim \mathcal{N}(0,1)$. The MMSE can then be upper bounded by Jensen's inequality and lower bounded by considering only negative values of Y:

$$\frac{1}{e^{2\mathsf{snr}}+1} < \mathsf{mmse}(\mathsf{snr}) < \frac{2}{e^{2\mathsf{snr}}+1}, \tag{2.117}$$

and hence

$$\lim_{\mathsf{snr}\to\infty} \frac{1}{\mathsf{snr}} \log \mathsf{mmse}(\mathsf{snr}) = -2. \tag{2.118}$$

If the inputs are not equiprobable, then it is possible to have an even faster decay of MMSE as $\mathsf{snr} \to \infty$. For example, using a special input of the type (similar to flash signaling [104])

$$X = \begin{cases} \sqrt{\frac{1-p}{p}} & \text{w.p. } p, \\ -\sqrt{\frac{p}{1-p}} & \text{w.p. } 1-p, \end{cases} \tag{2.119}$$

it can be shown that in this case

$$\mathsf{mmse}(\mathsf{snr}) \le \frac{1}{2p(1-p)} \exp\left[-\frac{\mathsf{snr}}{4p(1-p)}\right]. \tag{2.120}$$

Hence the MMSE can be made to decay faster than any given exponential by choosing a small enough p.

2.3 Continuous-time Channels

The success in the random variable/vector Gaussian channel setting in Section 2.2 can be extended to the more sophisticated continuous-time models. Consider the following continuous-time Gaussian channel:

$$R_t = \sqrt{\mathsf{snr}}\,X_t + N_t, \quad t \in [0,T], \tag{2.121}$$

where $\{X_t\}$ is the input process, $\{N_t\}$ a white Gaussian noise with a flat double-sided spectrum of unit height, and snr denotes the signal-to-noise ratio. Since $\{N_t\}$ is not second-order, it is mathematically more convenient to study an equivalent model obtained by

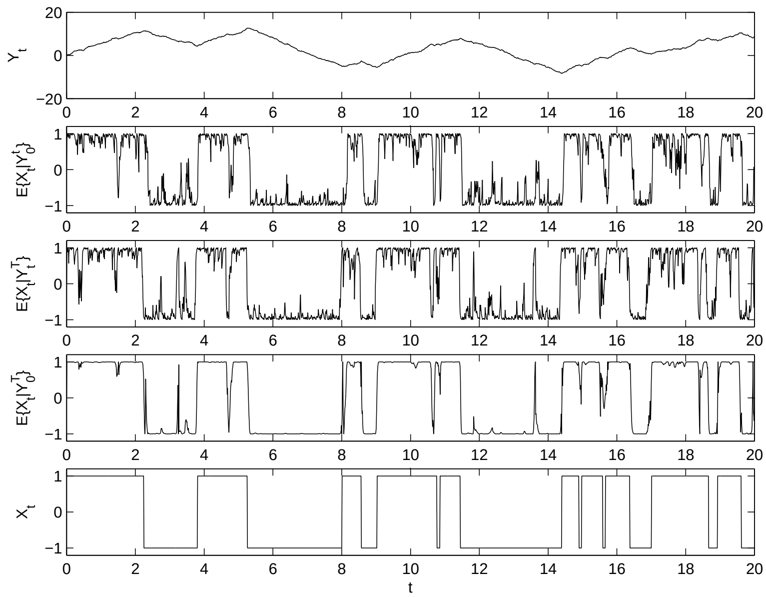

Figure 2.4: Sample paths of the input and output processes of an additive white Gaussian noise channel, the output of the optimal forward and backward filters, as well as the output of the optimal smoother. The input $\{X_t\}$ is a random telegraph waveform with unit transition rate. The SNR is 15 dB.

integrating the observations in (2.121). In a concise form, the input and output processes are related by a standard Wiener process $\{W_t\}$ (also known as the Brownian motion) independent of the input:

$$\mathrm{d}Y_t = \sqrt{\mathsf{snr}}\, X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \in [0, T]. \tag{2.122}$$

An example of the sample paths of the input and output signals is shown in Figure 2.4. Note that instead of scaling the Brownian motion as is ubiquitous in the literature, we choose to scale the input process so as to minimize notation in the analysis and results. The additive Brownian motion model is fundamental in many fields and is central in many textbooks (see e.g. [62]).

In continuous-time signal processing, both the *causal (filtering) MMSE* and *noncausal (smoothing) MMSE* are important performance measures. Suppose for now that the input is a stationary process. Let $\mathsf{cmmse}(\mathsf{snr})$ and $\mathsf{mmse}(\mathsf{snr})$ denote the causal and noncausal MMSEs respectively. Let $I(\mathsf{snr})$ denote now the *mutual information rate*, which measures the average mutual information between the input and output processes per unit time. This section shows that the central formula (2.4) in Section 2.2 also holds literally in this continuous-time setting, i.e., the derivative of the mutual information rate is equal to half the noncausal MMSE. Furthermore, the filtering MMSE is equal to the expected value of the smoothing MMSE:

$$\mathsf{cmmse}(\mathsf{snr}) = \mathsf{E}\{\mathsf{mmse}(\Gamma)\} \tag{2.123}$$

where Γ is chosen uniformly distributed between 0 and snr. In fact, stationarity of the input is not required if the MMSEs are defined as time averages.

Relationships between the causal and noncausal estimation errors have been studied for the particular case of linear estimation (or Gaussian inputs) in [1], where a bound on the loss due to causality constraint is quantified. Duncan [20, 21], Zakai [58, ref. [53]] and Kadota *et al.* [53] pioneered the investigation of relations between the mutual information and conditional mean filtering, which capitalized on earlier research on the "estimator-correlator" principle by Price [78], Kailath [54], and others (see [59]). In particular, Duncan showed that the input-output mutual information can be expressed as a time-integral of the causal MMSE [21].[5] Duncan's relationship is proven to be useful in a wide spectrum of applications in information theory and statistics [53, 52, 4, 10]. There are also a number of other works in this area, most notably those of Liptser [63] and Mayer-Wolf and Zakai [64], where the rate of increase in the mutual information between the sample of the input process at the current time and the entire past of the output process is expressed in the causal estimation error and some Fisher informations. Similar results were also obtained for discrete-time models by Bucy [9]. In [88] Shmelev devised a general, albeit complicated, procedure to obtain the optimal smoother from the optimal filter.

The new relationships as well as Duncan's Theorem are proved in this chapter using incremental channels, which analyze the increase in the input-output mutual information due to an infinitesimal increase in either the SNR or observation time. A counterpart of formula (2.4) in continuous-time setting is first established. The result connecting filtering and smoothing MMSEs admits an information-theoretic proof. So far, no other proof is known.

2.3.1 Mutual Information Rate and MMSEs

We are concerned with three quantities associated with the model (2.122), namely, the causal MMSE achieved by optimal filtering, the noncausal MMSE achieved by optimal smoothing, and the mutual information between the input and output processes. As a convention, let X_τ^t denote the process $\{X_t\}$ in the interval $[\tau, t]$. Also, let μ_X denote the probability measure induced by $\{X_t\}$ in the interval of interest. The input-output mutual information $I\left(X_0^T; Y_0^T\right)$ is defined by (2.1). The causal and noncausal MMSEs at any time $t \in [0, T]$ are defined in the usual way:

$$\mathsf{cmmse}(t, \mathsf{snr}) = \mathsf{E}\left\{\left(X_t - \mathsf{E}\left\{X_t \mid Y_0^t; \mathsf{snr}\right\}\right)^2\right\}, \tag{2.124}$$

and

$$\mathsf{mmse}(t, \mathsf{snr}) = \mathsf{E}\left\{\left(X_t - \mathsf{E}\left\{X_t \mid Y_0^T; \mathsf{snr}\right\}\right)^2\right\}. \tag{2.125}$$

Recall the mutual information rate (mutual information per unit time) defined in the natural way:

$$I(\mathsf{snr}) = \lim_{T\to\infty} \frac{1}{T} I\left(X_0^T; Y_0^T\right). \tag{2.126}$$

Similarly, the average causal and noncausal MMSEs (per unit time) are defined as

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{T}\int_0^T \mathsf{cmmse}(t, \mathsf{snr})\, \mathrm{d}t \tag{2.127}$$

and

$$\mathsf{mmse}(\mathsf{snr}) = \frac{1}{T}\int_0^T \mathsf{mmse}(t, \mathsf{snr})\, \mathrm{d}t \tag{2.128}$$

[5] Duncan's Theorem was independently obtained by Zakai in the more general setting of inputs that may depend causally on the noisy output in a 1969 unpublished Bell Labs Memorandum (see [58]).

respectively.

To start with, let $T \to \infty$ and assume that the input to the continuous-time model (2.122) is a stationary[6] Gaussian process with power spectrum $S_X(\omega)$. The mutual information rate was obtained by Shannon [87]:

$$I(\mathsf{snr}) = \frac{1}{2}\int_{-\infty}^{\infty} \log\left(1 + \mathsf{snr}\, S_X(\omega)\right) \frac{\mathrm{d}\omega}{2\pi}. \tag{2.129}$$

In this case optimal filtering and smoothing are both linear. The noncausal MMSE is due to Wiener [114],

$$\mathsf{mmse}(\mathsf{snr}) = \int_{-\infty}^{\infty} \frac{S_X(\omega)}{1 + \mathsf{snr}\, S_X(\omega)} \frac{\mathrm{d}\omega}{2\pi}, \tag{2.130}$$

and the causal MMSE is due to Yovits and Jackson [118, equation (8c)]:

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}}\int_{-\infty}^{\infty} \log\left(1 + \mathsf{snr}\, S_X(\omega)\right) \frac{\mathrm{d}\omega}{2\pi}. \tag{2.131}$$

From (2.129) and (2.130), it is easy to see that the derivative of the mutual information rate (nats per unit time) is equal to half the noncausal MMSE, i.e., the central formula (2.4) for the random variable channel holds literally in case of continuous-time Gaussian input process. Moreover, (2.129) and (2.131) show that the mutual information rate is equal to the causal MMSE scaled by half the SNR, although, interestingly, this connection escaped Yovits and Jackson [118].

In fact, these relationships are true not only for Gaussian inputs. Theorem 2.1 can be generalized to the continuous-time model with an arbitrary input process:

Theorem 2.6 *If the input process $\{X_t\}$ to the Gaussian channel (2.122) has finite average power, i.e.,*

$$\int_0^T \mathsf{E} X_t^2 \,\mathrm{d}t < \infty, \tag{2.132}$$

then

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr}). \tag{2.133}$$

Proof: See Section 2.3.2. ■

What is special for the continuous-time model is the relationship between the mutual information rate and the causal MMSE due to Duncan [21], which is put into a more concise form here:

Theorem 2.7 (Duncan [21]) *For any input process with finite average power,*

$$I(\mathsf{snr}) = \frac{\mathsf{snr}}{2}\mathsf{cmmse}(\mathsf{snr}). \tag{2.134}$$

Together, Theorems 2.6 and 2.7 show that the mutual information, the causal MMSE and the noncausal MMSE satisfy a triangle relationship. In particular, using the mutual information rate as a bridge, the causal MMSE is found to be equal to the noncausal MMSE averaged over SNR:

[6] For stationary input it would be more convenient to shift $[0, T]$ to $[-T/2, T/2]$ and then let $T \to \infty$ so that the causal and noncausal MMSEs at any time $t \in (-\infty, \infty)$ is independent of t. We stick to $[0, T]$ in this chapter for notational simplicity in case of general inputs.

Theorem 2.8 *For any input process with finite average power,*

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma. \tag{2.135}$$

Equality (2.135) is a surprising new relationship between causal and noncausal MMSEs. It is quite remarkable considering the fact that nonlinear filtering is usually a hard problem and few special case analytical expressions are known for the optimal estimation errors in continuous-time problems.

Note that, the equality can be rewritten as

$$\mathsf{cmmse}(\mathsf{snr}) - \mathsf{mmse}(\mathsf{snr}) = -\mathsf{snr} \frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} \mathsf{cmmse}(\mathsf{snr}), \tag{2.136}$$

which quantifies the increase of the minimum estimation error due to the causality constraint. It is interesting to point out that for stationary inputs the anti-causal MMSE is equal to the causal MMSE. The reason is that the noncausal MMSE remains the same in reversed time and white Gaussian noise is reversible. Note that in general the optimal anti-causal filter is different from the optimal causal filter.

It is worth pointing out that Theorems 2.6–2.8 are still valid if the time averages in (2.126)–(2.128) are replaced by their limits as $T \to \infty$. This is particularly relevant to the case of stationary inputs.

Random Telegraph Input

Besides Gaussian inputs, another example of the relation in Theorem 2.8 is an input process called the random telegraph waveform, where $\{X_t\}$ is a stationary Markov process with two equally probable states ($X_t = \pm 1$). See Figure 2.4 for an illustration. Assume that the transition rate of the input Markov process is ν, i.e., for sufficiently small h,

$$\mathsf{P}\{X_{t+h} = X_t\} = 1 - \nu h + o(h), \tag{2.137}$$

the expressions for the MMSEs achieved by optimal filtering and smoothing are obtained as [115, 116]:

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{\int_1^\infty u^{-\frac{1}{2}} (u-1)^{-\frac{1}{2}} e^{-\frac{2\nu u}{\mathsf{snr}}}\, \mathrm{d}u}{\int_1^\infty u^{\frac{1}{2}} (u-1)^{-\frac{1}{2}} e^{-\frac{2\nu u}{\mathsf{snr}}}\, \mathrm{d}u}, \tag{2.138}$$

and

$$\mathsf{mmse}(\mathsf{snr}) = \frac{\int_{-1}^1 \int_{-1}^1 \frac{(1+xy)\exp\left[-\frac{2\nu}{\mathsf{snr}}\left(\frac{1}{1-x^2}+\frac{1}{1-y^2}\right)\right]}{-(1-x)^3(1-y)^3(1+x)(1+y)}\, \mathrm{d}x\, \mathrm{d}y}{\left[\int_1^\infty u^{\frac{1}{2}} (u-1)^{-\frac{1}{2}} e^{-\frac{2\nu u}{\mathsf{snr}}}\, \mathrm{d}u\right]^2} \tag{2.139}$$

respectively. The relationship (2.135) can be verified by algebra [45]. The MMSEs are plotted in Figure 2.5 as functions of the SNR for unit transition rate.

Figure 2.4 shows experimental results of the filtering and smoothing of the random telegraph signal corrupted by additive white Gaussian noise. The forward filter follows Wonham [115]:

$$\mathrm{d}\widehat{X}_t = -\left[2\nu \widehat{X}_t + \mathsf{snr}\, \widehat{X}_t \left(1 - \widehat{X}_t^2\right)\right] \mathrm{d}t + \sqrt{\mathsf{snr}} \left(1 - \widehat{X}_t^2\right) \mathrm{d}Y_t, \tag{2.140}$$

where

$$\widehat{X}_t = \mathsf{E}\left\{ X_t \mid Y_0^t \right\}. \tag{2.141}$$

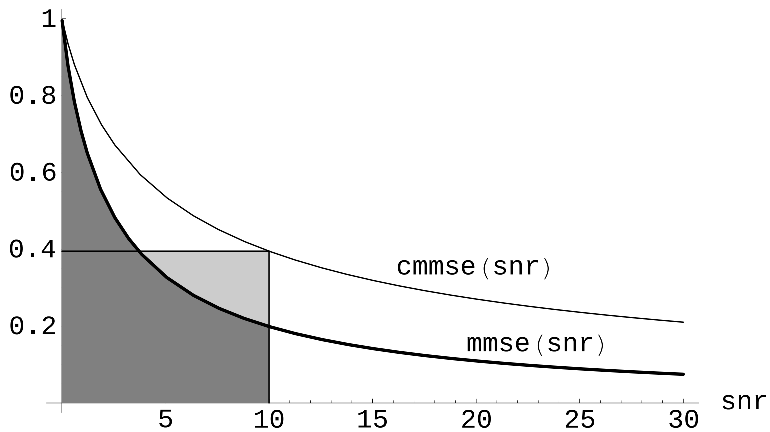

Figure 2.5: The causal and noncausal MMSEs of continuous-time Gaussian channel with a random telegraph waveform input. The transition rate $\nu = 1$. The two shaded regions have the same area due to Theorem 2.8.

This is in fact resulted from a representation theorem of Doob's [18]. The backward filter is merely a time reversal of the filter of the same type. The smoother is due to Yao [116]:

$$\mathsf{E}\left\{ X_t \mid Y_0^T \right\} = \frac{\mathsf{E}\left\{ X_t \mid Y_0^t \right\} + \mathsf{E}\left\{ X_t \mid Y_t^T \right\}}{1 + \mathsf{E}\left\{ X_t \mid Y_0^t \right\} \mathsf{E}\left\{ X_t \mid Y_t^T \right\}}. \tag{2.142}$$

The smoother results in better MMSE of course. Numerical values of the MMSEs in Figure 2.4 are consistent with the curves in Figure 2.5.

Low-SNR and High-SNR Asymptotics

Based on Theorem 2.8, one can study the asymptotics of the mutual information and MMSE in continuous-time setting under low SNR. The relationship (2.135) implies that

$$\lim_{\mathsf{snr} \to 0} \frac{\mathsf{mmse}(0) - \mathsf{mmse}(\mathsf{snr})}{\mathsf{cmmse}(0) - \mathsf{cmmse}(\mathsf{snr})} = 2 \tag{2.143}$$

where

$$\mathsf{cmmse}(0) = \mathsf{mmse}(0) = \frac{1}{T} \int_0^T \mathsf{var}\left\{ X_t \right\} \, \mathrm{d}t. \tag{2.144}$$

Hence the rate of decrease (with snr) of the noncausal MMSE is twice that of the causal MMSE at low SNRs.

In the high SNR regime, there exist inputs that make the MMSE exponentially small. However, in case of Gauss-Markov input processes, Steinberg *et al.* [91] observed that the causal MMSE is asymptotically twice the noncausal MMSE, as long as the input-output relationship is described by

$$\mathrm{d}Y_t = \sqrt{\mathsf{snr}}\, h(X_t) \, \mathrm{d}t + \mathrm{d}W_t \tag{2.145}$$

where $h(\cdot)$ is a differentiable and increasing function. In the special case where $h(X_t) = X_t$, Steinberg *et al.*'s observation can be justified by noting that in the Gauss-Markov case, the

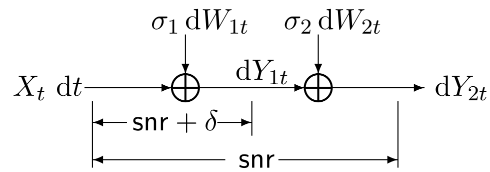

Figure 2.6: A continuous-time SNR-incremental Gaussian channel.

smoothing MMSE satisfies [8]:

$$\mathsf{mmse}(\mathsf{snr}) = \frac{c}{\sqrt{\mathsf{snr}}} + o\left(\frac{1}{\mathsf{snr}}\right), \tag{2.146}$$

which implies according to (2.135) that

$$\lim_{\mathsf{snr}\to\infty} \frac{\mathsf{cmmse}(\mathsf{snr})}{\mathsf{mmse}(\mathsf{snr})} = 2. \tag{2.147}$$

Unlike the universal factor of 2 result in (2.143) for the low SNR regime, the factor of 2 result in (2.147) for the high SNR regime fails to hold in general. For example, for the random telegraph waveform input, the causality penalty increases in the order of $\log \mathsf{snr}$ [116].

2.3.2 The SNR-Incremental Channel

Theorem 2.6 can be proved using the SNR-incremental channel approach developed in Section 2.2. Consider a cascade of two Gaussian channels with independent noise processes as depicted in Figure 2.6:

$$\mathrm{d}Y_{1t} = X_t\ \mathrm{d}t + \sigma_1\ \mathrm{d}W_{1t}, \tag{2.148a}$$

$$\mathrm{d}Y_{2t} = \mathrm{d}Y_{1t} + \sigma_2\ \mathrm{d}W_{2t}, \tag{2.148b}$$

where $\{W_{1t}\}$ and $\{W_{2t}\}$ are independent standard Wiener processes also independent of $\{X_t\}$, and σ_1 and σ_2 satisfy (2.41) so that the signal-to-noise ratio of the first channel and the composite channel is $\mathsf{snr}+\delta$ and snr respectively. Given $\{X_t\}$, $\{Y_{1t}\}$ and $\{Y_{2t}\}$ are jointly Gaussian processes. Following steps similar to those that lead to (2.49), it can be shown that

$$(\mathsf{snr}+\delta)\ \mathrm{d}Y_{1t} = \mathsf{snr}\ \mathrm{d}Y_{2t} + \delta\, X_t\,\mathrm{d}t + \sqrt{\delta}\ \mathrm{d}W_t, \tag{2.149}$$

where $\{W_t\}$ is a standard Wiener process independent of $\{X_t\}$ and $\{Y_{2t}\}$. Hence conditioned on the process $\{Y_{2t}\}$ in $[0,T]$, (2.149) can be regarded as a Gaussian channel with an SNR of δ. Similar to Lemma 2.1, the following result holds.

Lemma 2.5 *As $\delta\to 0$, the input-output mutual information of the following Gaussian channel:*

$$\mathrm{d}Y_t = \sqrt{\delta}\, Z_t\,\mathrm{d}t + \mathrm{d}W_t, \quad t\in[0,T], \tag{2.150}$$

where $\{W_t\}$ is standard Wiener process independent of the input $\{Z_t\}$, which satisfies

$$\int_0^T \mathsf{E}Z_t^2\ \mathrm{d}t < \infty, \tag{2.151}$$

is given by the following:

$$\lim_{\delta\to 0}\frac{1}{\delta} I\left(Z_0^T; Y_0^T\right) = \frac{1}{2}\int_0^T \mathsf{E}\left(Z_t - \mathsf{E}Z_t\right)^2 \, \mathrm{d}t. \tag{2.152}$$

Proof: See Appendix A.3. ∎

Applying Lemma 2.5 to the Gaussian channel (2.149) conditioned on $\{Y_{2t}\}$ in $[0,T]$, one has

$$I\left(X_0^T; Y_{1,0}^T | Y_{2,0}^T\right) = \frac{\delta}{2}\int_0^T \mathsf{E}\left\{\left(X_t - \mathsf{E}\left\{X_t \mid Y_{2,0}^T\right\}\right)^2\right\} \mathrm{d}t + o(\delta). \tag{2.153}$$

Since $\{X_t\}$—$\{Y_{1t}\}$—$\{Y_{2t}\}$ is a Markov chain, the left hand side of (2.153) is recognized as the mutual information increase:

$$\begin{aligned} I\left(X_0^T; Y_{1,0}^T \,|\, Y_{2,0}^T\right) &= I\left(X_0^T; Y_{1,0}^T\right) - I\left(X_0^T; Y_{2,0}^T\right) &(2.154)\\ &= T\left[I(\mathsf{snr}+\delta) - I(\mathsf{snr})\right]. &(2.155) \end{aligned}$$

By (2.155) and the definition of the noncausal MMSE (2.125), (2.153) can be rewritten as

$$I(\mathsf{snr}+\delta) - I(\mathsf{snr}) = \frac{\delta}{2T}\int_0^T \mathsf{mmse}(t,\mathsf{snr}) \, \mathrm{d}t + o(\delta). \tag{2.156}$$

Hence the proof of Theorem 2.6.

The property that independent Wiener processes sum up to a Wiener process is essential in the above proof. The incremental channel device is very useful in proving integral equations such as in Theorem 2.6. Indeed, by the SNR-incremental channel it has been shown that the mutual information at a given SNR is an accumulation of the MMSEs of degraded channels due to the fact that an infinitesimal increase in the SNR adds to the total mutual information an increase proportional to the MMSE.

2.3.3 The Time-Incremental Channel

Note Duncan's Theorem (Theorem 2.7) that links the mutual information and the causal MMSE is yet another integral equation, although inexplicit, where the integral is with respect to time on the right hand side of (2.134). Analogous to the SNR-incremental channel, one can investigate the mutual information increase due to an infinitesimal extra time duration of observation of the channel output. This leads to a new proof of Theorem 2.7 in the following, which is more intuitive than Duncan's original one [21].

Theorem 2.7 is equivalent to

$$I\left(X_0^{t+\delta}; Y_0^{t+\delta}\right) - I\left(X_0^t; Y_0^t\right) = \delta\,\frac{\mathsf{snr}}{2}\,\mathsf{E}\left\{\left(X_t - \mathsf{E}\left\{X_t \mid Y_0^t; \mathsf{snr}\right\}\right)^2\right\} + o(\delta), \tag{2.157}$$

which is to say that the mutual information increase due to the extra observation time is proportional to the causal MMSE. The left hand side of (2.157) can be written as

$$\begin{aligned} & I\left(X_0^{t+\delta}; Y_0^{t+\delta}\right) - I\left(X_0^t; Y_0^t\right) \\ =\; & I\left(X_0^t, X_t^{t+\delta}; Y_0^t, Y_t^{t+\delta}\right) - I\left(X_0^t; Y_0^t\right) &(2.158)\\ =\; & I\left(X_0^t, X_t^{t+\delta}; Y_0^t\right) + I\left(X_t^{t+\delta}; Y_t^{t+\delta} \,|\, Y_0^t\right) \\ & + I\left(X_0^t; Y_t^{t+\delta} \,|\, X_t^{t+\delta}, Y_0^t\right) - I\left(X_0^t; Y_0^t\right) &(2.159)\\ =\; & I\left(X_t^{t+\delta}; Y_t^{t+\delta} \,|\, Y_0^t\right) + I\left(X_0^t; Y_t^{t+\delta} \,|\, X_t^{t+\delta}, Y_0^t\right) + I\left(X_t^{t+\delta}; Y_0^t \,|\, X_0^t\right). &(2.160) \end{aligned}$$

Since Y_0^t—X_0^t—$X_t^{t+\delta}$—$Y_t^{t+\delta}$ is a Markov chain, the last two mutual informations in (2.160) vanish due to conditional independence. Therefore,

$$I\left(X_0^{t+\delta};Y_0^{t+\delta}\right) - I\left(X_0^t;Y_0^t\right) = I\left(X_t^{t+\delta};Y_t^{t+\delta}\,|\,Y_0^t\right), \tag{2.161}$$

i.e., the increase in the mutual information is the conditional mutual information between the input and output during the extra time interval given the past observation. This can be understood easily by considering a conceptual "time-incremental channel". Note that conditioned on Y_0^t, the channel in $(t, t+\delta)$ remains the same but with a different input distribution due to conditioning on Y_0^t. Let us denote this new channel by

$$\mathrm{d}\tilde{Y}_t = \sqrt{\mathsf{snr}}\,\tilde{X}_t\,\mathrm{d}t + \mathrm{d}W_t, \quad t \in [0,\delta], \tag{2.162}$$

where the time duration is shifted to $[0,\delta]$, and the input process $\tilde{X}_0^\delta$ has the same law as $X_t^{t+\delta}$ conditioned on Y_0^t. Instead of looking at this new problem of an infinitesimal time interval $[0,\delta]$, we can convert the problem to a familiar one by an expansion in the time axis. Since

$$\sqrt{\delta}\,W_{t/\delta} \tag{2.163}$$

is also a standard Wiener process, the channel (2.162) in $[0,\delta]$ is equivalent to a new channel described by

$$\mathrm{d}\tilde{\tilde{Y}}_\tau = \sqrt{\delta\,\mathsf{snr}}\,\tilde{\tilde{X}}_\tau\,\mathrm{d}\tau + \mathrm{d}W'_\tau, \quad \tau \in [0,1], \tag{2.164}$$

where $\tilde{\tilde{X}}_\tau = \tilde{X}_{\tau\delta}$, and $\{W'_t\}$ is a standard Wiener process. The channel (2.164) is of (fixed) unit duration but a diminishing signal-to-noise ratio of $\delta\,\mathsf{snr}$. It is interesting to note here that the trick here performs a "time-SNR" transform (see also Section 2.4.1). By Lemma 2.5, the mutual information is

$$\begin{aligned}
I\left(X_t^{t+\delta};Y_t^{t+\delta}|Y_0^t\right) &= I\left(\tilde{\tilde{X}}_0^1;\tilde{\tilde{Y}}_0^1\right) &(2.165)\\
&= \frac{\delta\,\mathsf{snr}}{2}\int_0^1 \mathsf{E}(\tilde{\tilde{X}}_\tau - \mathsf{E}\tilde{\tilde{X}}_\tau)^2\,\mathrm{d}\tau + o(\delta) &(2.166)\\
&= \frac{\delta\,\mathsf{snr}}{2}\int_0^1 \mathsf{E}\left\{\left(X_{t+\tau\delta} - \mathsf{E}\left\{X_{t+\tau\delta}\,|\,Y_0^t;\mathsf{snr}\right\}\right)^2\right\}\mathrm{d}\tau + o(\delta) &(2.167)\\
&= \frac{\delta\,\mathsf{snr}}{2}\mathsf{E}\left\{\left(X_t - \mathsf{E}\left\{X_t\,|\,Y_0^t;\mathsf{snr}\right\}\right)^2\right\} + o(\delta), &(2.168)
\end{aligned}$$

where (2.168) is justified by the continuity of the MMSE. The relation (2.157) is then established due to (2.161) and (2.168), and hence the proof of Theorem 2.7.

Similar to the discussion in Section 2.2.4, the integral equations in Theorems 2.6 and 2.7 proved by using the SNR- and time-incremental channels are also consequences of the mutual information chain rule applied to a Markov chain of the channel input and degraded versions of channel outputs. The independent-increment property both SNR-wise and time-wise is quintessential in establishing the results.

2.4 Discrete-time vs. Continuous-time

In Sections 2.2 and 2.3, the mutual information and the estimation errors have been shown to satisfy similar relations in the random variable/vector and continuous-time random process models. This section bridges these results for different models under certain circumstances. Moreover, discrete-time models can be analyzed by considering piecewise constant input to the continuous-time channel.

2.4.1 A Fourth Proof of Theorem 2.1

Besides the direct and incremental-channel approaches, a fourth proof of the mutual information and MMSE relationship in the random variable/vector model can be obtained using continuous-time results in Section 2.3. For simplicity we prove Theorem 2.1 using Theorem 2.7. The proof can be easily modified to show Theorem 2.2, using the vector version of Duncan's Theorem [21].

A continuous-time counterpart of the model (2.5) can be constructed by letting $X_t \equiv X$ for $t \in [0,1]$ where X is a random variable not dependent on t:

$$\mathrm{d}Y_t = \sqrt{\mathsf{snr}}\, X \;\mathrm{d}t + \mathrm{d}W_t. \tag{2.169}$$

For every $u \in [0,1]$, Y_u is a sufficient statistic of the observation Y_0^u for X and hence also for X_0^u. Therefore, the input-output mutual information of the scalar channel (2.5) is equal to that of the continuous-time channel (2.169):

$$I(\mathsf{snr}) = I(X; Y_1) = I\left(X_0^1; Y_0^1\right). \tag{2.170}$$

Integrating both sides of (2.169), one has

$$Y_u = \sqrt{\mathsf{snr}}\, u\, X + W_u, \quad u \in [0,1], \tag{2.171}$$

where $W_u \sim \mathcal{N}(0,u)$. Note that (2.171) is exactly a scalar Gaussian channel with a signal-to-noise ratio of $u\,\mathsf{snr}$. Clearly, the MMSE of the continuous-time model given the observation Y_0^u, i.e., the causal MMSE at time u with a signal-to-noise ratio of snr, is equal to the MMSE of a scalar Gaussian channel with a signal-to-noise ratio of $u\,\mathsf{snr}$:

$$\mathsf{cmmse}(u, \mathsf{snr}) = \mathsf{mmse}(u\,\mathsf{snr}). \tag{2.172}$$

By Theorem 2.7, the mutual information can be written as

$$\begin{aligned} I(X_0^1; Y_0^1) &= \frac{\mathsf{snr}}{2} \int_0^1 \mathsf{cmmse}(u, \mathsf{snr}) \;\mathrm{d}u & (2.173)\\ &= \frac{\mathsf{snr}}{2} \int_0^1 \mathsf{mmse}(u\,\mathsf{snr}) \;\mathrm{d}u & (2.174)\\ &= \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) \;\mathrm{d}\gamma. & (2.175) \end{aligned}$$

Thus Theorem 2.1 follows by noticing (2.170).

Note also that in this setting, the MMSE at any time t of a continuous-time Gaussian channel with a signal-to-noise ratio of $u\,\mathsf{snr}$ is equal to the MMSE of a scalar Gaussian channel at the same SNR:

$$\mathsf{mmse}(t, u\,\mathsf{snr}) = \mathsf{mmse}(u\,\mathsf{snr}), \quad \forall t \in [0,T]. \tag{2.176}$$

Together, (2.172) and (2.176) yield (2.135) for this special input by taking average over time u.

Indeed, for an observation time duration $[0,u]$ of the continuous-time channel output, the corresponding signal-to-noise ratio is $u\,\mathsf{snr}$ in the equivalent scalar channel model; or in other words, the useful signal energy is accumulated over time. The integral over time in (2.134) and the integral over SNR are interchangeable in this case. This is clearly another example of the "time-SNR" transform which is also used in Section 2.3.3.

In retrospect of the above proof, the time-invariant input can be replaced by a general form of $X\,h(t)$, where $h(t)$ is any deterministic continuous signal.

2.4.2 Discrete-time Channels

Consider the case where the input is a discrete-time process and the channel is

$$Y_i = \sqrt{\mathsf{snr}}\, X_i + N_i, \quad i = 1, 2, \ldots, \tag{2.177}$$

where the noise N_i is a sequence of i.i.d. standard Gaussian random variables. Given that we have already studied the case of a finite-dimensional vector channel, an advantageous analysis of (2.177) consists of treating the finite-horizon case $i = 1, \ldots, n$ and then taking the limit as $n \to \infty$.

Let $\boldsymbol{X}^n$ denote a column vector formed by the sequence $X_1, \ldots, X_n$. Putting the finite-horizon version of (2.177) in a vector form results in a MIMO channel of the form (2.18) with $\boldsymbol{H}$ being the identity matrix. Therefore the relation (2.21) between the mutual information and the MMSE holds also in this case:

Theorem 2.9 *If* $\sum_{i=1}^n \mathsf{E} X_i^2 < \infty$*, then*

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I\left(\boldsymbol{X}^n; \sqrt{\mathsf{snr}}\, \boldsymbol{X}^n + \boldsymbol{N}^n\right) = \frac{1}{2} \sum_{i=1}^n \mathsf{mmse}(i, \mathsf{snr}), \tag{2.178}$$

where

$$\mathsf{mmse}(i, \mathsf{snr}) = \mathsf{E}\left\{ (X_i - \mathsf{E}\left\{ X_i \mid \boldsymbol{Y}^n; \mathsf{snr} \right\})^2 \right\} \tag{2.179}$$

is the noncausal MMSE at time i given the entire observation $\boldsymbol{Y}^n$.

It is important to note that the MMSE in this case is noncausal since the estimate is obtained through optimal smoothing. It is also interesting to consider optimal filtering and prediction in this setting. Let the MMSE of optimal filtering be defined as

$$\mathsf{cmmse}(i, \mathsf{snr}) = \mathsf{E}\left\{ \left(X_i - \mathsf{E}\left\{ X_i \mid \boldsymbol{Y}^i; \mathsf{snr} \right\}\right)^2 \right\}, \tag{2.180}$$

and the MMSE of optimal one-step prediction as

$$\mathsf{pmmse}(i, \mathsf{snr}) = \mathsf{E}\left\{ \left(X_i - \mathsf{E}\left\{ X_i \mid \boldsymbol{Y}^{i-1}; \mathsf{snr} \right\}\right)^2 \right\}. \tag{2.181}$$

In discrete-time setting, the identity (2.4) still holds, while the relationship between the mutual information and the causal MMSEs (Duncan's Theorem) does not: Instead, the mutual information is lower bounded by the filtering error but upper bounded by the prediction error.

Theorem 2.10 *The input-output mutual information satisfies:*

$$\frac{\mathsf{snr}}{2} \sum_{i=1}^n \mathsf{cmmse}(i, \mathsf{snr}) \le I\left(\boldsymbol{X}^n; \boldsymbol{Y}^n\right) \le \frac{\mathsf{snr}}{2} \sum_{i=1}^n \mathsf{pmmse}(i, \mathsf{snr}). \tag{2.182}$$

Proof: Consider the discrete-time model (2.177) and its piecewise constant continuous-time counterpart:

$$\mathrm{d}Y_t = \sqrt{\mathsf{snr}}\, X_{\lceil t \rceil}\, \mathrm{d}t + \mathrm{d}W_t, \quad t \in [0, \infty). \tag{2.183}$$

It is clear that in the time interval $(i-1, i]$ the input to the continuous-time model is equal to the random variable X_i. Note the delicacy in notation. Y_0^n stands for a sample path of the continuous-time random process $\{Y_t, t \in [0,n]\}$, $\boldsymbol{Y}^n$ stands for a discrete-time process $\{Y_1, \dots, Y_n\}$, or the vector consisting of samples of $\{Y_t\}$ at integer times, whereas Y_i is either the i-th point of $\boldsymbol{Y}^n$ or the sample of $\{Y_t\}$ at $t = i$ depending on the context. It is easy to see that the samples of $\{Y_t\}$ at natural numbers are sufficient statistics for the input process $\boldsymbol{X}^n$. Hence

$$I\left(\boldsymbol{X}^n; \boldsymbol{Y}^n\right) = I\left(\boldsymbol{X}^n; Y_0^n\right), \quad n = 1, 2, \dots. \tag{2.184}$$

Note that the causal MMSE of the continuous-time model takes the same value as the causal MMSE of the discrete-time model at integer values i. Thus it suffices to use $\mathsf{cmmse}(\cdot, \mathsf{snr})$ to denote the causal MMSE under both discrete- and continuous-time models. Here, $\mathsf{cmmse}(i, \mathsf{snr})$ is the MMSE of the estimation of X_i given the observation $\boldsymbol{Y}^i$ which is a sufficient statistic of Y_0^i, while $\mathsf{pmmse}(i, \mathsf{snr})$ is the MMSE of the estimation of X_i given the observation $\boldsymbol{Y}^{i-1}$ which is a sufficient statistic of Y_0^{i-1}. Suppose that $t \in (i-1, i]$. Since the filtration generated by Y_0^i (or $\boldsymbol{Y}^i$) contains more information about X_i than the filtration generated by Y_0^t, which in turn contains more information about X_i than Y_0^{i-1}, one has

$$\mathsf{cmmse}(\lceil t \rceil, \mathsf{snr}) \le \mathsf{cmmse}(t, \mathsf{snr}) \le \mathsf{pmmse}(\lceil t \rceil, \mathsf{snr}). \tag{2.185}$$

Integrating (2.185) over t establishes Theorem 2.10 by noting also that

$$I\left(\boldsymbol{X}^n; \boldsymbol{Y}^n\right) = \frac{\mathsf{snr}}{2} \int_0^n \mathsf{cmmse}(t, \mathsf{snr}) \, \mathrm{d}t \tag{2.186}$$

due to Theorem 2.7. ∎

The above analysis can also be reversed to prove the continuous-time results (Theorems 2.6 and 2.7) starting from the discrete-time ones (Theorems 2.9 and 2.10) through piecewise constant process approximations at least for continuous input processes. In particular, let $\boldsymbol{X}^n$ be the samples of X_t equally spaced in $[0, T]$. Letting $n \to \infty$ allows Theorem 2.7 to be recovered from Theorem 2.10, since the sum on both sides of (2.182) (divided by n) converge to integrals and the prediction MMSE converges to the causal MMSE due to continuity.

2.5 Generalizations and Observations

2.5.1 General Additive-noise Channel

Theorems 2.1 and 2.2 show the relationship between the mutual information and the MMSE as long as the mutual information is between a stochastic signal and an observation of it in Gaussian noise. Let us now consider the more general setting where the input is preprocessed arbitrarily before contamination by additive Gaussian noise as depicted in Figure 2.7. Let X be a random object jointly distributed with a real-valued random variable Z. The channel output is expressed as

$$Y = \sqrt{\mathsf{snr}}\, Z + N, \tag{2.187}$$

where the noise $N \sim \mathcal{N}(0, 1)$ is independent of X and Z. The preprocessor can be regarded as a channel with arbitrary conditional probability distribution $P_{Z|X}$. Since X—Z—Y is a Markov chain,

$$I(X; Y) = I(Z; Y) - I(Z; Y \mid X). \tag{2.188}$$

Note that given (X, Z), the channel output Y is Gaussian. Two applications of Theorem 2.1 to the right hand side of (2.188) give the following:

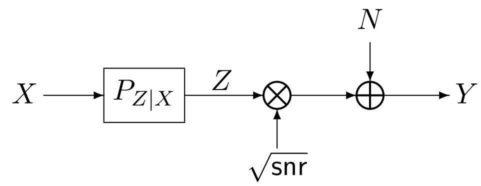

Figure 2.7: A general additive-noise channel.

Theorem 2.11 *Let X—Z—Y be a Markov chain and Z and Y be connected through (2.187). If $\mathsf{E}Z^2 < \infty$, then*

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I\left(X; \sqrt{\mathsf{snr}}\, Z + N\right) = \frac{1}{2}\mathsf{mmse}\left(Z|\sqrt{\mathsf{snr}}\, Z + N\right) - \frac{1}{2}\mathsf{mmse}\left(Z|X, \sqrt{\mathsf{snr}}\, Z + N\right). \quad (2.189)$$

The special case of this result for zero SNR is given by Theorem 1 of [77]. As a simple illustration of Theorem 2.11, consider a scalar channel where $X \sim \mathcal{N}\left(0, \sigma_X^2\right)$ and $P_{Z|X}$ is a Gaussian channel with noise variance σ^2. Then straightforward calculations yield

$$I(X;Y) = \frac{1}{2}\log\left(1 + \frac{\mathsf{snr}\,\sigma_X^2}{1 + \mathsf{snr}\,\sigma^2}\right), \quad (2.190)$$

and

$$\mathsf{mmse}\left(Z|\sqrt{\mathsf{snr}}\, Z + N\right) = \frac{\sigma_X^2 + \sigma^2}{1 + \mathsf{snr}\left(\sigma_X^2 + \sigma^2\right)}, \quad (2.191)$$

$$\mathsf{mmse}\left(Z|X, \sqrt{\mathsf{snr}}\, Z + N\right) = \frac{\sigma^2}{1 + \mathsf{snr}\,\sigma^2}. \quad (2.192)$$

The relationship (2.189) is easy to check.

In the special case where the preprocessor is a deterministic function of the input, e.g., $Z = g(X)$ where $g(\cdot)$ is an arbitrary deterministic mapping, the second term on the right hand side of (2.189) vanishes. Note also that since $I(X;Y) = I(g(X);Y)$ in this case, one has

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(X; \sqrt{\mathsf{snr}}\, g(X) + N) = \frac{1}{2}\mathsf{mmse}\left(g(X)\,|\,\sqrt{\mathsf{snr}}\, g(X) + N\right). \quad (2.193)$$

Hence (2.14) holds verbatim where the MMSE in this case is defined as the minimum error in estimating $g(X)$. Indeed, the vector channel in Theorem 2.2 is merely a special case of the vector version of this general result.

One of the many scenarios in which the general result can be useful is the intersymbol interference channel. The input (Z_i) to the Gaussian channel is the desired symbol (X_i) corrupted by a function of the previous symbols $(X_{i-1}, X_{i-2}, \dots)$. Theorem 2.11 can possibly be used to calculate (or bound) the mutual information given a certain input distribution. Another domain of applications of Theorem 2.11 is the case of fading channels known or unknown at the receiver.

Using similar arguments as in the above, nothing prevents us from generalizing the continuous-time results in Section 2.3 to a much broader family of models:

$$\mathrm{d}Y_t = \sqrt{\mathsf{snr}}\, Z_t\ \mathrm{d}t + \mathrm{d}W_t, \quad (2.194)$$

where $\{Z_t\}$ is a random process jointly distributed with the random message X, and $\{W_t\}$ is a Wiener process independent of X and $\{Z_t\}$. The following is straightforward in view of Theorem 2.11.

Theorem 2.12 *As long as the input $\{Z_t\}$ to the channel (2.194) has finite average power,*

$$\frac{\mathrm{d}}{\mathrm{dsnr}} I\left(X; Y_0^T\right) = \frac{1}{2T} \int_0^T \mathsf{mmse}\left(Z_t \,|\, Y_0^T\right) - \mathsf{mmse}\left(Z_t \,|\, X, Y_0^T\right) \,\mathrm{d}t. \tag{2.195}$$

In case $Z_t = g_t(X)$, where $g_t(\cdot)$ is an arbitrary time-varying mapping, Theorems 2.6-2.8 hold verbatim except that the finite-power requirement now applies to $g_t(X)$, and the MMSEs in this case refer to the minimum errors in estimating $g_t(X)$.

Extension of the results to the case of colored Gaussian noise is straightforward by filtering the observation to whiten the noise and recover the canonical model of the form (2.187).

2.5.2 New Representation of Information Measures

Consider a discrete random variable X. The mutual information between X and its observation through a Gaussian channel converges to the entropy of X as the SNR of the channel goes to infinity.

Lemma 2.6 *For any discrete real-valued random variable X,*

$$H(X) = \lim_{\mathsf{snr}\to\infty} I\left(X; \sqrt{\mathsf{snr}}\, X + N\right). \tag{2.196}$$

Proof: See Appendix A.7. ■

Note that if $H(X)$ is infinity then the mutual information in (2.196) also increases without bound as $\mathsf{snr} \to \infty$. Moreover, the result holds if X is subject to an arbitrary one-to-one mapping $g(\cdot)$ before going through the channel. In view of (2.193), the following theorem is immediate.

Theorem 2.13 *For any discrete random variable X and one-to-one mapping $g(\cdot)$ that maps X to real numbers, the entropy in nats can be obtained as*

$$H(X) = \frac{1}{2} \int_0^\infty \mathsf{E}\left\{\left(g(X) - \mathsf{E}\left\{ g(X) \,|\, \sqrt{\mathsf{snr}}\, g(X) + N\right\}\right)^2\right\} \,\mathsf{dsnr}. \tag{2.197}$$

It is interesting to note that the integral on the right hand side of (2.197) is not dependent on the choice of $g(\cdot)$, which is not evident from estimation-theoretic properties alone. It is possible, however, to check this in special cases.

Other than for discrete random variables, the entropy is not defined and the input-output mutual information is in general unbounded as SNR increases. One may consider the divergence between the input distribution and a Gaussian distribution with the same mean and variance:

Lemma 2.7 *For any real-valued random variable X. Let X' be Gaussian with the same mean and variance as X, i.e., $X' \sim \mathcal{N}\left(\mathsf{E}X, \sigma_X^2\right)$. Let Y and Y' be the output of the channel (2.5) with X and X' as the input respectively. Then*

$$\mathsf{D}\left(P_X \| P_{X'}\right) = \lim_{\mathsf{snr}\to\infty} \mathsf{D}\left(P_Y \| P_{Y'}\right). \tag{2.198}$$

The lemma can be proved using monotone convergence and the fact that data processing reduces divergence. Note that in case the divergence between P_X and $P_{X'}$ is infinity, the divergence between P_Y and $P_{Y'}$ also increases without bound. Since

$$\mathsf{D}\left(P_Y \| P_{Y'}\right) = I(X';Y') - I(X;Y), \tag{2.199}$$

the following theorem is straightforward by applying Theorem 2.1.

Theorem 2.14 *For any random variable X with $\sigma_X^2 < \infty$,*

$$\mathsf{D}\left(P_X \| \mathcal{N}(\mathsf{E}X, \sigma_X^2)\right) = \frac{1}{2}\int_0^\infty \frac{\sigma_X^2}{1+\mathsf{snr}\,\sigma_X^2} - \mathsf{mmse}\left(X|\sqrt{\mathsf{snr}}\,X+N\right)\,\mathrm{d}\mathsf{snr}. \tag{2.200}$$

Note that the integrand in (2.200) is always positive since Gaussian inputs maximizes the MMSE. Also, Theorem 2.14 holds even if the divergence is infinity, for example in the case that X is not a continuous random variable.

In view of Theorem 2.14, the differential entropy of X can also be expressed as a function of the MMSE:

$$\begin{aligned} h(X) &= \frac{1}{2}\log\left(2\pi e\,\sigma_X^2\right) - \mathsf{D}\left(P_X \| P_{X'}\right) && (2.201)\\ &= \frac{1}{2}\log\left(2\pi e\,\sigma_X^2\right) - \frac{1}{2}\int_0^\infty \frac{\sigma_X^2}{1+\mathsf{snr}\,\sigma_X^2} - \mathsf{mmse}\left(X|\sqrt{\mathsf{snr}}\,X+N\right)\,\mathrm{d}\mathsf{snr}. && (2.202)\end{aligned}$$

Theorem 2.11 provides an apparently new means of representing the mutual information between an arbitrary random variable X and a real-valued random variable Z:

$$I(X;Z) = \frac{1}{2}\int_0^\infty \mathsf{E}\left\{\left(\mathsf{E}\left\{Z \mid \sqrt{\mathsf{snr}}\,Z+N, X\right\}\right)^2 - \left(\mathsf{E}\left\{Z \mid \sqrt{\mathsf{snr}}\,Z+N\right\}\right)^2\right\}\,\mathrm{d}\mathsf{snr}, \tag{2.203}$$

where N is standard Gaussian.

It is remarkable that the entropy, differential entropy, divergence and mutual information in fairly general settings admit expressions in pure estimation-theoretic quantities. It remains to be seen whether such representations find any application.

2.5.3 Generalization to Vector Models

Just as that Theorem 2.1 obtained under a scalar model has its counterpart (Theorem 2.2) under a vector model, all the results in Sections 2.3 and 2.4 are generalizable to vector models, under both discrete-time and continuous-time settings. For example, the vector continuous-time model takes the form of

$$\mathrm{d}\boldsymbol{Y}_t = \sqrt{\mathsf{snr}}\,\boldsymbol{X}_t\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t, \tag{2.204}$$

where $\{\boldsymbol{W}_t\}$ is an m-dimensional Wiener process, and $\{\boldsymbol{X}_t\}$ and $\{\boldsymbol{Y}_t\}$ are m-dimensional random processes. Theorem 2.6 holds literally, while the mutual information rate, estimation errors, and power are now defined with respect to the vector signals and their Euclidean norms. Note also that Duncan's Theorem was originally given in vector form [21]. It should be noted that the incremental-channel devices are directly applicable to the vector models.

In view of the above generalizations, the discrete- and continuous-time results in Sections 2.5.1 and 2.5.2 also extend straightforwardly to vector models.

2.6 Summary

This chapter reveals for the first time that the input-output mutual information and the (noncausal) MMSE in estimating the input given the output determine each other by a simple differential formula under both discrete- and continuous-time, scalar and vector Gaussian channel models (Theorems 2.1, 2.2, 2.4, 2.6 and 2.9). A consequence of this relationship is the coupling of the MMSEs achievable by smoothing and filtering with arbitrary signals corrupted by Gaussian noise (Theorems 2.8 and 2.10). Moreover, new expressions in terms of MMSE are found for information measures such as entropy and input-output mutual information of a general channel with real/complex-valued output (Theorems 2.3, 2.5, 2.11, 2.12, 2.13 and 2.14). Asymptotics of the mutual information and MMSE are studied in both the low- and high-SNR domains.

The idea of incremental channels is the underlying basis for the most streamlined proof of the main results and for their interpretation. The white Gaussian nature of the noise is key to this approach since 1) the sum of independent Gaussian variates is Gaussian; and 2) the Wiener process (time-integral of white Gaussian noise) has independent increments.

Besides those given in Section 2.2.5, applications of the relationships revealed in this chapter are abundant. The fact that the mutual information and the (noncausal) MMSE determine each other by a simple formula also provides a new means to calculate or bound one quantity using the other. An upper (resp. lower) bound for the mutual information is immediate by bounding the MMSE using a suboptimal (resp. genie aided) estimator. Lower bounds on the MMSE, e.g., [7], may also lead to new lower bounds on the mutual information.

Results in this chapter have been published in part in [37] and are included in a submitted paper [36]. Extensions of this work are found in [34, 35]. In a follow-up to this work, Zakai has recently extended the central formula (2.4) to the abstract Wiener space [120], which generalizes the classical m-dimensional Wiener process.

Chapter 3

Multiuser Channels

In contrast to the canonical additive Gaussian channels studied in Chapter 2, this chapter assumes a specific structure of the communication channel where independent inputs come from multiple users, and study individual user's error performance and reliable information rate, as well as the overall efficiency of the multiuser system.

3.1 Introduction

Consider a multidimensional Euclidean space in which each user (or source) randomly picks a signature vector and modulates its own symbol onto it. The channel output is then a superposition of all users' signals corrupted by Gaussian noise. Such a model, best described as a matrix channel, is very versatile and is widely used in applications that include code-division multiple access, multi-input multi-output systems, etc. With knowledge of all signature waveforms, the task of an estimator is to recover the transmitted symbols from some or all users.

This chapter focuses on a paradigm of multiuser channels, randomly spread code-division multiple access, in which a number of users share a common media to communicate to a single receiver simultaneously over the same bandwidth. Each user in a CDMA system employs a "signature waveform" with a large time-bandwidth product that results in many advantages particularly in wireless communications: frequency diversity, robustness to channel impairment, ease of resource allocation, etc. The price to pay is multiple-access interference due to non-orthogonal spreading sequences from all users. Numerous multiuser detection techniques have been proposed to mitigate the MAI to various degrees. This work concerns the efficiency of the multiuser system in two aspects: One is *multiuser efficiency*, which in general measures the quality of multiuser detection outputs under uncoded transmission; the other is *spectral efficiency*, which is the total information rate normalized by the dimensionality of the CDMA channel achievable by coded transmission. As one shall see, the multiuser efficiency and spectral efficiency are tightly related to the mean-square error of multiuser detection and the mutual information between the input and detection output respectively.

3.1.1 Gaussian or Non-Gaussian?

The most efficient use of a multiuser channel is through jointly optimal decoding, which is an NP-complete problem [110]. A common suboptimal strategy is to apply a multiuser detector front end with polynomial complexity and then perform independent single-user decoding. The quality of the detection output fed to decoders is of great interest.

In [106, 107, 108], Verdú first used the concept of multiuser efficiency in binary uncoded transmission to refer to the degradation of the output signal-to-noise ratio relative to a single-user channel calibrated at the same BER. The multiuser efficiencies of matched filter, decorrelator, and linear MMSE detector, were found as functions of the correlation matrix of the spreading sequences. Particular attention has been given to the asymptotic multiuser efficiency in the more tractable region of high SNR. Expressions for the optimum uncoded asymptotic multiuser efficiency were found in [108, 109].

In the large-system limit, where the number of users and the spreading factor both tend to infinity with a fixed ratio, the dependence of system performance on the spreading sequences vanishes, and random matrix theory proves to be an excellent tool for analyzing linear detectors. The large-system multiuser efficiency of the matched filter is trivial [112]. The multiuser efficiency of the MMSE detector is obtained explicitly in [112] for the equal-power case and in [99] as the solution to the so-called Tse-Hanly fixed-point equation in the case with flat fading. The decorrelator is also analyzed [23, 39]. The success with a wide class of linear detectors hinges on the facts that 1) The detection output (e.g., for user k) is a sum of independent components: the desired signal, the MAI and Gaussian background noise:

$$\tilde{X}_k = X_k + \mathsf{mai}_k + N_k; \tag{3.1}$$

and 2) The multiple-access interference mai_k is asymptotically Gaussian (e.g., [44]). Clearly, as far as linear multiuser detectors are concerned, the performance is fully characterized by the SNR degradation due to MAI regardless of the input distribution, i.e., the multiuser efficiency. Therefore, by incorporating the linear detector into the channel, an individual user experiences essentially a single-user Gaussian channel with noise enhancement associated with the MAI variance.

The error performance of nonlinear detectors such as the optimal ones are hard problems. The difficulty here is inherent to nonlinear operations in estimation. That is, the detection output cannot be decomposed as a sum of independent components associated with the desired signal and interferences respectively. Moreover, the detection output is in general asymptotically non-Gaussian conditioned on the input. An extreme case is the maximum-likelihood multiuser detector for binary transmission, the hard decision output of which takes only two values. The difficulty remains if one looks at the soft detection output, which is the mean value of the posterior probability distribution. Hence, unlike for a Gaussian output, the conditional variance of a general detection output does not lead to simple characterization of multiuser efficiency and error performance. For illustration, Figure 3.1 plots the probability density function obtained from the histogram of the soft output statistic of the individually optimal detector given that +1 was transmitted. The simulated system has 8 users, a spreading factor of 12, and an SNR of 2 dB. Note that negative decision values correspond to decision error; hence the area under the curve on the left half plane gives the BER. The distribution shown in Figure 3.1 is far from Gaussian. Thus the usual notion of output SNR fails to capture the essence of system performance. In fact, much literature is devoted to evaluating the error performance by Monte Carlo simulation.

This chapter makes a contribution to the understanding of the multiuser detection in the large-system regime. It is found under certain assumptions that the output decision statistic of a nonlinear detector, such as the one whose distribution is depicted by Figure 3.1, converges in fact to a very simple monotonic function of a "hidden" Gaussian random

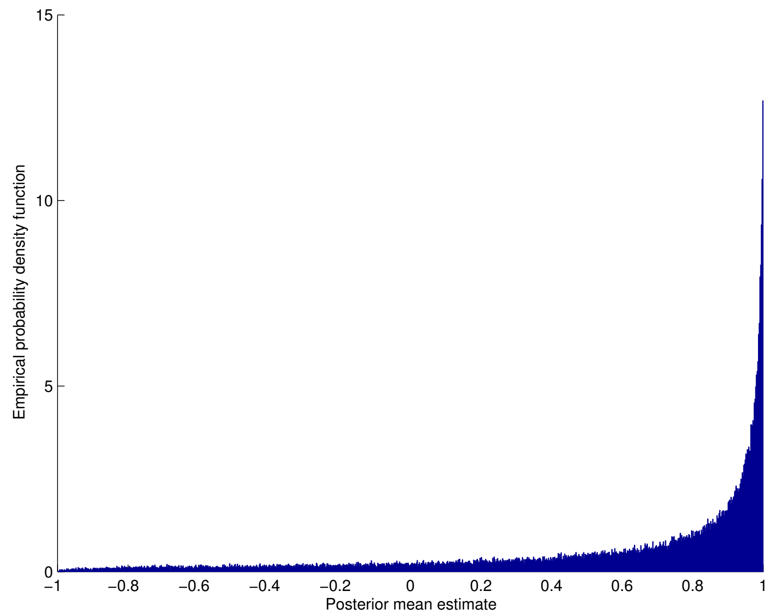

Figure 3.1: The probability density function obtained from the histogram of an individually optimal soft detection output conditioned on +1 being transmitted. The system has 8 users, the spreading factor is 12, and the SNR 2 dB.

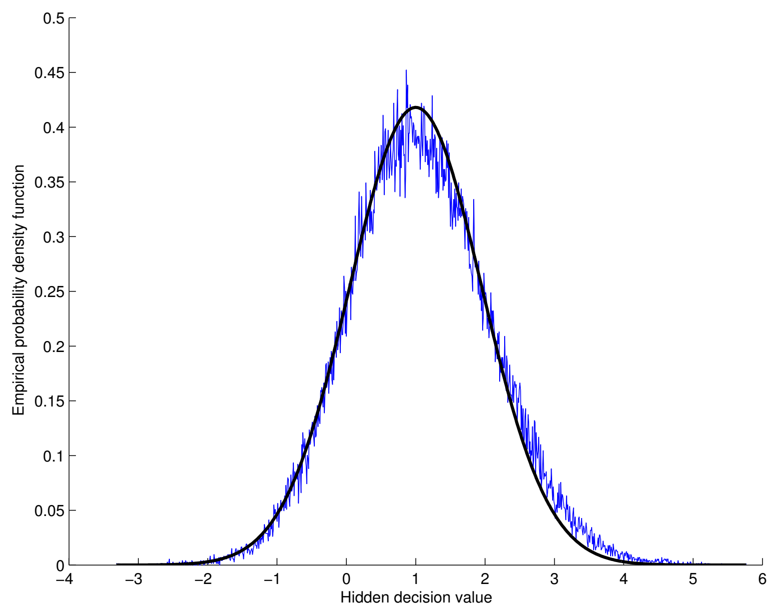

Figure 3.2: The probability density function obtained from the histogram of the hidden equivalent Gaussian statistic conditioned on +1 being transmitted. The system has 8 users, the spreading factor is 12, and the SNR 2 dB. The asymptotic Gaussian distribution is also plotted for comparison.

variable, i.e.,

$$\tilde{X}_k = f(X_k + \mathsf{mai}_k + N_k). \tag{3.2}$$

One may contend that it is always possible to monotonically map a non-Gaussian distribution to a Gaussian one. What is surprisingly simple and powerful here is that the mapping f does not depend on the transmitted symbols which we wish to estimate in the first place; neither does it depend on the instantaneous spreading sequences in the large-system regime. Indeed, the function f is determined by merely a few parameters of the system. By applying an inverse of this function, an equivalent Gaussian statistic is recovered, so that we are back to the familiar ground where the output SNR (defined for the Gaussian statistic) completely characterizes system performance. In other words, the multiuser efficiency can still be obtained as the ratio of the output and input SNRs. Since each user enjoys now an equivalent single-user Gaussian channel with an enhanced Gaussian noise in lieu of the MAI, we will refer to this result as the "decoupling principle". In this chapter, under certain assumption, the decoupling principle will be shown to hold for not only optimal detection, but also a generic multiuser detector front end, called the *generalized posterior mean estimator*, which can be particularized to the matched filter, decorrelator, linear MMSE detector, as well as the jointly and individually optimal detectors. Moreover, the principle holds for arbitrary input distributions. Although results on performance analysis of multiuser detections are abundant, we believe that this work is the first to point out the decoupling principle, and henceforth introduces a completely new angle to multiuser detection problems.

For illustration, Figure 3.2 plots the probability density function of the Gaussian statistic (obtained by applying the inverse function f^{-1}) corresponding to the non-Gaussian one in Figure 3.1. The theoretically predicted density function is also shown for comparison. The "fit" is good considering that a relatively small system of 8 users with a processing gain of 12 is considered. Note that in case the multiuser detector is linear, the mapping f is also linear, and (3.2) reduces to (3.1).

By merit of the decoupling principle, the mutual information between the input and the output of the generic detector for each user is exactly the input-output mutual information of the equivalent scalar Gaussian channel under the same input, which now admits a simple analytical expression. Hence the large-system spectral efficiency of several well-known linear detectors, first found in [105] and [85] with and without fading respectively, can be recovered straightforwardly using the multiuser efficiency and the decoupling principle. New results for spectral efficiency of nonlinear detector and arbitrary inputs under both joint and separate decoding are obtained. The additive decomposition of optimal spectral efficiency as a sum of single-user efficiencies and a joint decoding gain [85] applies under more general conditions than originally thought.

It should be pointed out here that the large-system results are close representatives of practical system of moderate size. As seen numerically, a system of as few as 8 users can often be well approximated as a large system.

3.1.2 Random Matrix vs. Spin Glass

Much of the early success in the large-system analysis of linear detectors relies on the fact that the multiuser efficiency of a finite-size system can be written as an explicit function of the singular values of the random correlation matrix, the empirical distributions of which converge to a known function in the large-system limit [100, 3]. As a result, the limit of the multiuser efficiency can be obtained as an integral with respect to the limiting singular-value distribution. Indeed, this random matrix technique is applicable to analyzing any performance measure that can be expressed as a function of the singular values. Based on an explicit expression for CDMA channel capacity in [103], Verdú and Shamai quantified the optimal spectral efficiency in the large-system limit [105, 85] (see also [32, 81]). The expression found in [105] also solved the capacity of single-user narrowband multiantenna channels as the number of antennas grows—a problem that was open since the pioneering work of Foschini [29] and Teletar [97]. Unfortunately, few explicit expressions of the efficiencies in terms of singular values are available beyond the above cases. Much less success has been reported in the application of random matrix theory in other problems, for example, the multiuser efficiency of nonlinear detectors and the spectral efficiency achieved by practical signaling constellations such as m-PSK.

A major consequence of random matrix theory is that the dependence of the performance measures on the spreading sequences vanishes as the system size increases without bound. In other words, the performance measures are "self-averaging." This property is nothing but a manifestation of a fundamental law that the fluctuation of macroscopic properties of certain many-body systems vanishes in thermodynamic limit, i.e., when the number of interacting bodies becomes large. This falls under the general scope of statistical physics, whose principal goal is to study the macroscopic properties of physical systems containing a large number of particles starting from knowledge of microscopic interactions. Indeed, the asymptotic eigenvalue distribution of large correlation matrices can be derived via statistical physics [82]. Tanaka pioneered statistical physics concepts and methodologies in multiuser detection and obtained the large-system uncoded minimum BER (hence the multiuser efficiency) and spectral efficiency with equal-power antipodal inputs [93, 94, 95, 96]. We further elucidated the relationship between CDMA and statistical physics and generalized Tanaka's results to the case of non-equal powers [39]. Inspired by [96], Müller and Gerstacker [72] studied the channel capacity under separate decoding and noticed that the additive decomposition of the optimum spectral efficiency in [85] holds also for binary inputs. Müller thus further conjectured the same formula to be valid regardless of the input distribution [71].

In this chapter, we continue along the line of [96, 39] and present a unified treatment of Gaussian CDMA channels and multiuser detection assuming an arbitrary input distribution and fading characteristic. A wide class of multiuser detectors, optimal as well as suboptimal, are treated uniformly under the framework of posterior mean estimators. The central result is the above-mentioned decoupling principle, and analytical solutions are also reported on the multiuser efficiency and spectral efficiency under both general and specific settings.

The key technique in this chapter, the replica method, has its origin in spin glass theory in statistical physics [22]. Analogies between statistical physics and neural networks, coding, image processing, and communications have long been noted (e.g., [73, 89]). There have been many recent activities to apply statistical physics wisdom in error-correcting codes [49, 69, 50]. The rather unconventional method was first used by Tanaka to analyze several well-known CDMA multiuser detectors [96]. In this chapter, we extend Tanaka's original ideas and draw a parallel between the general statistical inference problem in multiuser communications and the problem of determining the configuration of random spins subject to quenched randomness. A mathematical framework is then developed in the CDMA context based on the replica recipe. For the purpose of analytical tractability, we will assume that, 1) the self-averaging property applies, 2) the "replica trick" is valid, and 3) replica symmetry holds. These assumptions have been used successfully in many problems in statistical physics as well as neural networks and coding theory, to name a few, while a complete justification of the replica method is an ongoing effort in mathematics and

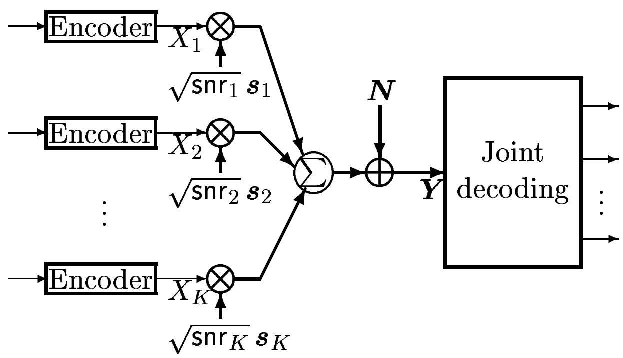

Figure 3.3: The CDMA channel with joint decoding.

physics communities. The results in this chapter are based on these assumptions and therefore a rigorous justification is pending on breakthroughs in those problems. In case the assumptions fail to hold, results obtained using the replica method may still capture many of the qualitative features of the system performance. Indeed, such results are often found as good approximations in many cases where some of the assumptions are not true [68, 19]. Furthermore, the decoupling principle carries great practicality and may find convenient uses as long as the analytical predictions are close to the reality even if not exact.

The remainder of this chapter is organized as follows. In Section 3.2, we give the model and summarize major results. Relevant concepts and methodologies in statistical physics are introduced in Section 3.3. Detailed calculation based on a real-valued channel is presented in Section 3.4. Complex-valued channels are discussed in Section 3.5 followed by some numerical examples in Section 3.6.

3.2 Model and Summary of Results

3.2.1 System Model

Consider the K-user CDMA system with spreading factor L depicted in Figure 3.3. Each encoder maps its message into a sequence of channel symbols. All users employ the same type of signaling so that at each interval the K symbols are i.i.d. random variables with distribution P_X, which has zero mean and unit variance. Let $\boldsymbol{X} = [X_1, \dots, X_K]^\top$ denote the vector of input symbols from the K users. For notational convenience in explaining some of the ideas, it is assumed that either a probability density function or a probability mass function of P_X exists, and is denoted by p_X. Let also $p_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{k=1}^{K} p_X(x_k)$ denote the joint (product) distribution. The results in this chapter, however, hold in full generality and do not depend on the existence of a probability density or mass function.

Let user k's instantaneous SNR be denoted by snr_k and $\boldsymbol{\Gamma} = \operatorname{diag}\{\sqrt{\mathsf{snr}_1}, \dots, \sqrt{\mathsf{snr}_K}\}$. Denote the spreading sequence of user k by $\boldsymbol{s}_k = \frac{1}{\sqrt{L}}[S_{1k}, S_{2k}, \dots, S_{Lk}]^\top$, where S_{nk} are i.i.d. random variables with zero mean and finite moments. Let the symbols and spreading sequences be randomly chosen for each user and not dependent on the SNRs. The $L \times K$ channel state matrix is denoted by $\boldsymbol{S} = [\sqrt{\mathsf{snr}_1}\,\boldsymbol{s}_1, \dots, \sqrt{\mathsf{snr}_K}\,\boldsymbol{s}_K]$. Assuming symbol-synchronous

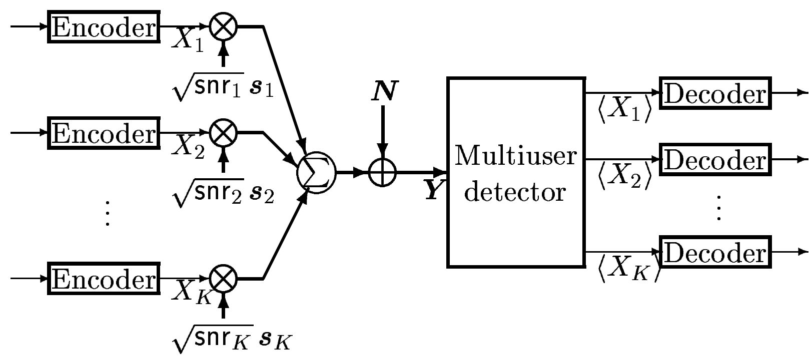

Figure 3.4: The CDMA channel with separate decoding.

transmission, a memoryless Gaussian CDMA channel with flat fading is described by:

$$\begin{aligned} \boldsymbol{Y} &= \sum_{k=1}^{K} \sqrt{\mathsf{snr}_k}\, \boldsymbol{s}_k X_k + \boldsymbol{N} & (3.3)\\ &= \boldsymbol{S}\boldsymbol{X} + \boldsymbol{N} & (3.4) \end{aligned}$$

where $\boldsymbol{N}$ is a vector consisting of i.i.d. zero-mean Gaussian random variables. Depending on the domain that the inputs and the spreading chips take values, the input-output relationship (3.4) describes either a real-valued or a complex-valued fading channel.

The linear system (3.4) is quite versatile. In particular, with $\mathsf{snr}_k = \mathsf{snr}$ for all k and snr deterministic, it models the canonical MIMO channel in which all propagation coefficients are i.i.d. (see (2.18)). An example is single-user communication with K transmit antennas and L received antennas, where the channel coefficients are not known to the transmitter.

3.2.2 Posterior Mean Estimation

The most efficient use of channel (3.4) in terms of information capacity is achieved by optimal joint decoding as depicted in Figure 3.3. The input-output mutual information of the CDMA channel given the channel state $\boldsymbol{S}$ is $I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S})$. Due to the NP-complete complexity of joint decoding, one often breaks the process into multiuser detection followed by separate decoding as shown in Figure 3.4. A multiuser detector front end with no knowledge of the error-control codes used by the encoder outputs an estimate of the transmitted symbols given the received signal and the channel state. Each decoder only takes the decision statistic of a single user of interest to decoding without awareness of the existence of any other users (in particular, without knowledge of the spreading sequences). By adopting this separate decoding approach, the channel together with the multiuser detector front end is viewed as a single-user channel for each user. The detection output sequence for an individual user is in general not a sufficient statistic for decoding even this user's own information. Hence the sum of the single-user mutual informations is less than the input-output mutual information of the multiple-access channel.

To capture the intended suboptimal structure, one has to restrict the capability of the multiuser detector here; otherwise the detector could in principle encode the channel state and the received signal $(\boldsymbol{S},\boldsymbol{Y})$ into a single real number as its output to each user, which is a sufficient statistic for all users! A plausible choice is the *posterior mean estimator*, which computes the mean of the posterior probability distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$, hereafter denoted by angle brackets $\langle\cdot\rangle$:

$$\langle \boldsymbol{X} \rangle = \mathsf{E}\left\{ \boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{S} \right\}. \tag{3.5}$$

In view of Chapter 2, (3.5) represents exactly the conditional mean estimator, or CME, which achieves the MMSE. It is however termed PME here in accordance to the Bayesian statistics literature for reasons to be clear shortly.

The PME can be understood as an "informed" optimal estimator which is supplied with the posterior probability distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$. Of course we always assume that the channel state $\boldsymbol{S}$ is revealed to the estimator, so that upon receipt of the channel output $\boldsymbol{Y}$, the informed estimator computes the posterior mean. A generalization of the PME is conceivable: Instead of informing the estimator with the actual posterior distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$, we can supply at will any other well-defined conditional distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$. The estimator can nonetheless perform "optimal" estimation based on the postulated measure q. We call this the *generalized posterior mean estimation*, which is conveniently denoted as

$$\langle \boldsymbol{X} \rangle_q = \mathsf{E}_q \left\{ \boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{S} \right\} \tag{3.6}$$

where $\mathsf{E}_q\{\cdot\}$ stands for the expectation with respect to the postulated measure q. For consistency, the subscripts in (3.6) can be dropped if the postulated measure q coincides with the true one p.

A postulated measure q different from p in general causes degradation in detection output. Such a strategy may be either due to lack of knowledge of the true statistics or a particular choice that corresponds to a certain estimator of interest. In principle, any deterministic estimation can be regarded as a generalized PME since we can always choose to put a unit mass at the desired estimation output given $(\boldsymbol{Y}, \boldsymbol{S})$. We will see in Section 3.2.3 that by choosing an appropriate measure q, it is easy to particularize the generalized PME to many important multiuser detectors. As will be shown in this chapter, the generic representation (3.6) allows a uniform treatment of a large family of multiuser detectors which results in a simple performance characterization. In general, information about the channel state $\boldsymbol{S}$ is also subject to manipulation, but this is out of the scope of this thesis.

Clearly, $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is induced from the input distribution p_X and the conditional Gaussian density function $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ of the channel (3.4) using the Bayes formula:[1]

$$p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = \frac{p_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S})}{\int p_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S})\, \mathrm{d}\boldsymbol{x}}. \tag{3.7}$$

In this work, the knowledge supplied to the generalized PME is assumed to be the posterior probability distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ corresponding to a postulated CDMA system, where the input distribution is an arbitrary q_X, and the input-output relationship of the postulated channel differs from the actual channel (3.4) by only the noise variance. That is, the postulated channel is characterized by

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{X} + \sigma \boldsymbol{N}' \tag{3.8}$$

where the channel state matrix $\boldsymbol{S}$ is the same as that of the actual channel, and $\boldsymbol{N}'$ is statistically the same as the Gaussian noise $\boldsymbol{N}$ in (3.4). Easily, $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is determined by

[1] Existence of a probability density function p_X is assumed for notational convenience, although the Bayes hold for an arbitrary measure P_X.

$q_{\boldsymbol{X}}$ and $q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ according to the Bayes formula akin to (3.7). Here, σ serves as a control parameter of the postulated channel. Also, the postulated input distribution q_X has zero-mean and finite moments. In short, we study a family of multiuser detectors parameterized by the postulated input and noise level (q_X, σ).

It should be noted that posterior mean estimation under postulated probability distributions has been used in Bayes statistics literature. This technique is introduced to multiuser detection by Tanaka in the special case of equal-power users with binary or Gaussian inputs [93, 94, 96]. This work, however, is the first to treat multiuser detection in a general scope that includes arbitrary input and arbitrary SNR distribution.

3.2.3 Specific Detectors

We identify specific choices of the postulated input distribution q_X and noise level σ under which the generalized PME is particularized to well-known multiuser detectors.

Linear Detectors

Let the postulated input distribution be standard Gaussian

$$q_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{3.9}$$

Then the posterior probability distribution is

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = [Z(\boldsymbol{y},\boldsymbol{S})]^{-1} \exp\left[-\frac{1}{2}\|\boldsymbol{x}\|^2 - \frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{3.10}$$

where $Z(\boldsymbol{y},\boldsymbol{S})$ is a normalization factor such that (3.10) is a probability density. Since (3.10) is a Gaussian density, it is easy to see that its mean is a linear filtering of the received signal $\boldsymbol{Y}$:

$$\langle \boldsymbol{X} \rangle_q = \left[\boldsymbol{S}^\top \boldsymbol{S} + \sigma^2 \boldsymbol{I}\right]^{-1} \boldsymbol{S}^\top \boldsymbol{Y}. \tag{3.11}$$

If $\sigma \to \infty$, (3.11) gives

$$\sigma^2 \langle X_k \rangle_q \longrightarrow \boldsymbol{s}_k^\top \boldsymbol{Y}, \tag{3.12}$$

and hence the generalized PME estimate is consistent with the matched filter output. If $\sigma = 1$, (3.11) is exactly the soft output of the linear MMSE detector. If $\sigma \to 0$, (3.11) converges to the soft output of the decorrelator. In general, the generalized PME reduces to a linear detector by postulating Gaussian inputs regardless of the postulated noise level. The control parameter σ can then be tuned to chose from the matched filter, decorrelator, MMSE detector, etc.

Optimal Detectors

Let the postulated input distribution q_X be identical to the true one, p_X. The posterior probability distribution is

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = [Z(\boldsymbol{y},\boldsymbol{S})]^{-1} p_{\boldsymbol{X}}(\boldsymbol{x}) \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{3.13}$$

where $Z(\boldsymbol{y},\boldsymbol{S})$ is a normalization factor.

Suppose that the postulated noise level $\sigma \to 0$, then most of the probability mass of the distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is concentrated on a vector that achieves the minimum of $\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|$, which also maximizes the likelihood function $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S})$. The generalized PME output $\lim_{\sigma\to 0} \langle \boldsymbol{X} \rangle_q$ is thus equivalent to that of jointly optimal (or maximum-likelihood) detection [112].

Alternatively, if $\sigma = 1$, then the postulated measure q coincides with the true measure p, i.e., $q \equiv p$. The PME outputs $\langle \boldsymbol{X} \rangle$ and is referred to as the soft version of the individually optimal multiuser detector [112]. Indeed, in case of m-PSK (resp. binary) inputs, hard decision of its soft output gives the most likely value of the transmitted symbol (resp. bit).

Also worth mentioning here is that, if $\sigma \to \infty$, the generalized PME reduces to the matched filter. This can be easily verified by noticing that

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = p_{\boldsymbol{X}}(\boldsymbol{x}) \left[1 - \frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2 + \mathcal{O}\left(\sigma^{-4}\right)\right], \tag{3.14}$$

and hence

$$\sigma^2 \langle \boldsymbol{X} \rangle_q \to \boldsymbol{S}^\top \boldsymbol{Y} \quad \text{in } L^2 \text{ as } \sigma \to \infty. \tag{3.15}$$

3.2.4 Main Results

This subsection summarizes the main results of this chapter. The replica analysis we carry out to obtain these results is relegated to Section 3.3 and 3.4.

Consider the generalized posterior mean estimator parameterized by (q_X, σ). The goal here is to quantify for each user k the distribution of the detection output $\langle X_k \rangle_q$ conditioned on the input X_k, the mutual information $I(X_k; \langle X_k \rangle_q | \boldsymbol{S})$ between the input and the output of the front end, as well as the overall spectral efficiency of optimal joint decoding $\frac{1}{L} I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{S})$. Although these quantities are all dependent on the realization of the channel state, in the large-system asymptote such dependence vanishes. By a *large system* we refer to the limit that both the number of users K and the spreading factor L tend to infinity but with K/L, known as the *system load*, converging to a positive number β, which may be greater than 1. It is also assumed that $\{\mathsf{snr}_k\}_{k=1}^K$ are i.i.d. with distribution P_{snr}, hereafter referred to as the *SNR distribution*. Clearly, the empirical distributions of the SNR of all users converge to the same distribution P_{snr} as $K \to \infty$. Note that this SNR distribution captures the fading characteristics of the channel. All moments of the SNR distribution are assumed to be finite.

The main results are stated in the following assuming a real-valued model. The complex-valued model will be discussed in Section 3.5. In the real-valued system, the inputs X_k, the spreading chips S_{nk}, and all entries of the noise $\boldsymbol{N}$ take real values and have unit variance. The characteristics of the actual channel and the postulated channel are

$$p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = (2\pi)^{-\frac{L}{2}} \exp\left[-\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right], \tag{3.16}$$

and

$$q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = \left(2\pi\sigma^2\right)^{-\frac{L}{2}} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{3.17}$$

respectively.

Given the system load β, the input distribution p_X to the actual CDMA channel, the SNR distribution P_{snr}, and the input distribution q_X and variance σ^2 of the postulated CDMA system, we express in these parameters the large-system limit of the multiuser efficiency and spectral efficiency under both separate and joint decoding.

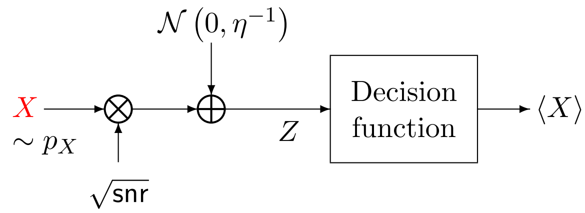

Figure 3.5: The equivalent scalar Gaussian channel followed by a decision function.

The Equivalent Single-user Channel

Consider the scalar Gaussian channel depicted in Figure 3.5:

$$Z = \sqrt{\mathsf{snr}}\, X + \frac{1}{\sqrt{\eta}}\, N \tag{3.18}$$

where $\mathsf{snr} > 0$ is the *input SNR*, $\eta > 0$ the *inverse noise variance* and $N \sim \mathcal{N}(0, 1)$ the noise independent of the input X. The conditional distribution associated with the channel is Gaussian:

$$p_{Z|X,\mathsf{snr};\eta}(z|x, \mathsf{snr}; \eta) = \sqrt{\frac{\eta}{2\pi}} \exp\left[-\frac{\eta}{2}\left(z - \sqrt{\mathsf{snr}}\, x\right)^2\right]. \tag{3.19}$$

Let the input distribution be p_X. The posterior mean estimate of X given the output Z is

$$\langle X \rangle = \mathsf{E}\left\{X | Z, \mathsf{snr}; \eta\right\} \tag{3.20}$$

where the expectation is over the posterior probability distribution $p_{X|Z,\mathsf{snr};\eta}$, which can be obtained through Bayes formula from the input distribution p_X and $p_{Z|X,\mathsf{snr};\eta}$. It is important to note that the single-user PME $\langle X \rangle$ is in general a nonlinear function of Z parameterized by snr and η. Clearly, $\langle X \rangle$ is also the (nonlinear) MMSE estimate, since it achieves the minimum mean-square error:

$$\mathsf{mmse}(\eta\, \mathsf{snr}) = \mathsf{E}\left\{ (X - \langle X \rangle)^2 \,\middle|\, \mathsf{snr}; \eta \right\}. \tag{3.21}$$

Note that this definition of $\mathsf{mmse}(\cdot)$ is consistent with the one (2.8) in Chapter 2 under the same input. The following is claimed for the multiuser posterior mean estimator (3.5).[2]

Claim 3.1 *In the large-system limit, the distribution of the posterior mean estimate $\langle X_k \rangle$ of the multiple-access channel (3.4) conditioned on $X_k = x$ being transmitted with signal-to-noise ratio snr_k is identical to the distribution of the posterior mean estimate $\langle X \rangle$ of the scalar Gaussian channel (3.18) conditioned on $X = x$ being transmitted with $\mathsf{snr} = \mathsf{snr}_k$, where the PME multiuser efficiency η (inverse noise variance of channel (3.18)) satisfies a fixed-point equation:*

$$\eta^{-1} = 1 + \beta\, \mathsf{E}\left\{\mathsf{snr} \cdot \mathsf{mmse}(\eta\, \mathsf{snr})\right\} \tag{3.22}$$

where the expectation is taken over the SNR distribution P_{snr}.

[2] Since the key assumptions made in Section 3.1 (essentially the replica method) are not rigorously justified, some of the results in this Chapter are referred to as claims. Nonetheless, proofs are provided in Section 3.4 based on those assumptions.

The large-system limit result here is understood as the following. Let $P^K_{\langle X_k\rangle|X_k=x}$ denote the input-output conditional distribution defined on a system of size K, then as $K \to \infty$, $P^K_{\langle X_k\rangle|X_k=x}$ converges weakly to the distribution $P_{\langle X\rangle|X=x}$ associated with the equivalent single-user channel for almost every x.

Claim 3.1 reveals that each single-user channel seen at the output of the PME for a CDMA channel is equivalent to a scalar Gaussian channel with enhanced noise followed by a (posterior mean) decision function as depicted in Figure 3.5. There exists a number $\eta \in [0,1]$ associated with the multiuser system, called the *multiuser efficiency*, which is a solution to the fixed-point equation (3.22). The effective SNR in each equivalent Gaussian channel is the input SNR times the multiuser efficiency. In other words, the multiple-access channel can be decoupled under nonlinear MMSE detection, where the effect of the MAI is summarized as a single parameter η^{-1} as the noise enhancement. As stated in Section 3.1, although the multiuser PME output $\langle X_k\rangle$ is in general non-Gaussian, it is in fact asymptotically a function (the decision function (3.20)) of Z, a conditionally Gaussian random variable centered at the actual input X_k scaled by $\sqrt{\mathsf{snr}_k}$.

It is straightforward to determine the multiuser efficiency η by (3.22). The following functions play an important role in our development (cf. (2.103)):

$$p_i(z, \mathsf{snr}; \eta) = \mathsf{E}\left\{ X^i \, p_{Z|X,\mathsf{snr};\eta}(z \,|\, X, \mathsf{snr}; \eta) \,\Big|\, \mathsf{snr} \right\}, \quad i = 0, 1, \dots \tag{3.23}$$

where the expectation is taken over the input distribution p_X, and $p_{Z|X,\mathsf{snr};\eta}$ is the Gaussian density associated with the scalar channel (3.18). The decision function (3.20) can be written as

$$\mathsf{E}\left\{X|Z, \mathsf{snr}; \eta\right\} = \frac{p_1(Z, \mathsf{snr}; \eta)}{p_0(Z, \mathsf{snr}; \eta)}, \tag{3.24}$$

and the MMSE

$$\mathsf{mmse}(\eta\, \mathsf{snr}) = 1 - \int \frac{p_1^2(z, \mathsf{snr}; \eta)}{p_0(z, \mathsf{snr}; \eta)} \, \mathrm{d}z. \tag{3.25}$$

The MMSE often allows simpler expressions than (3.25) for practical inputs (see examples in Section 3.2.3). Otherwise numerical integrals can be applied to evaluate (3.23) and henceforth (3.25). Thus, solutions to the fixed-point equation (3.22) can be found without much difficulty. There are cases that (3.22) has more than one solution. The ambiguity is resolved shortly in Claim 3.2.

Separate and Joint Decoding Spectral Efficiencies

The posterior mean decision function (3.24) is strictly monotone increasing (to be shown in Section 3.4.2) and thus inconsequential in both detection- and information-theoretic viewpoints. Hence the following corollary:

Corollary 3.1 *In the large-system limit, the mutual information between the input X_k and the output of the multiuser posterior mean estimator $\langle X_k\rangle$ is equal to the input-output mutual information of the equivalent scalar Gaussian channel (3.18) with the same input and SNR, and an inverse noise variance η as the PME multiuser efficiency.*

According to Corollary 3.1, the large-system mutual information for a user with signal-to-noise ratio snr is

$$I\left(X_k; \langle X_k\rangle \,|\boldsymbol{S}\right) \to I(\eta\, \mathsf{snr}) \quad \text{as } K \to \infty \tag{3.26}$$

where the mutual information $I(\cdot)$ as a function of the SNR is consistent with the definition (2.9) in Chapter 2, i.e.,

$$I(\eta\,\mathsf{snr}) = \mathsf{D}(p_{Z|X,\mathsf{snr};\eta} \,||\, p_{Z|\mathsf{snr};\eta} \,|\, p_X), \tag{3.27}$$

where $p_{Z|\mathsf{snr};\eta}$ is the marginal probability distribution of the output of channel (3.18). The overall spectral efficiency under suboptimal separate decoding is the sum of the single-user mutual informations divided by the dimensionality of the CDMA channel, which is simply

$$\mathsf{C}_{\text{sep}}(\beta) = \beta\,\mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\}. \tag{3.28}$$

The optimal spectral efficiency under joint decoding is greater than (3.28), where the difference is given by the following:

Claim 3.2 *The gain of optimal joint decoding over multiuser posterior mean estimation followed by separate decoding in the large-system spectral efficiency of the multiple-access channel (3.4) is determined by the PME multiuser efficiency η as*[3]

$$\mathsf{C}_{\text{joint}}(\beta) - \mathsf{C}_{\text{sep}}(\beta) = \frac{1}{2}(\eta - 1 - \log\eta) = \mathsf{D}\left(\mathcal{N}(0,\eta) \,||\, \mathcal{N}(0,1)\right). \tag{3.29}$$

In other words, the spectral efficiency under joint decoding is

$$\mathsf{C}_{\text{joint}}(\beta) = \beta\,\mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\} + \frac{1}{2}(\eta - 1 - \log\eta). \tag{3.30}$$

In case of multiple solutions to (3.22), the PME multiuser efficiency η is the one that gives the smallest spectral efficiency under optimal joint decoding.

Indeed, Müller's conjecture on the mutual information loss [71] is true for arbitrary inputs and SNRs. Incidentally, the loss is identified as a (Kullback-Leibler) divergence between two Gaussian distributions.

The fixed-point equation (3.22), which is obtained using the replica method, may have multiple solutions. This is known as phase coexistence in statistical physics. Among those solutions, the PME multiuser efficiency is the thermodynamically dominant one that gives the smallest value of the joint decoding spectral efficiency (3.30). It is the solution that carries relevant operational meaning in the communication problem. In general, as the system parameters (such as the load) change, the dominant solution may switch from one of the coexisting solutions to another. This is known as *phase transition* (refer to Section 3.6 for numerical examples).

Equal-power Gaussian input is the first known case that admits a closed form solution for the multiuser efficiency [112] and henceforth also the spectral efficiencies. The spectral efficiencies under joint and separate decoding were found for Gaussian inputs with fading in [85], and then found implicitly in [96] and later explicitly [72] for equal-power users with binary inputs. Formulas (3.28) and (3.30) give general formulas for arbitrary input distributions and received powers.

Interestingly, the spectral efficiencies under joint and separate decoding are also related by an integral equation. Even more so is the proof based on the central formula that links mutual information and MMSE in Chapter 2.

---

[3] Note that natural logarithm is assumed throughout.

Theorem 3.1 *For every input distribution P_X,*

$$\mathsf{C}_{\mathrm{joint}}(\beta) = \int_0^\beta \frac{1}{\beta'} \mathsf{C}_{\mathrm{sep}}(\beta')\,\mathrm{d}\beta'. \tag{3.31}$$

Proof: Since $\mathsf{C}_{\mathrm{joint}}(0) = 0$, it suffices to show

$$\beta \frac{\mathrm{d}}{\mathrm{d}\beta} \mathsf{C}_{\mathrm{joint}}(\beta) = \mathsf{C}_{\mathrm{sep}}(\beta). \tag{3.32}$$

By (3.28) and (3.30), it is enough to show

$$\beta \frac{\mathrm{d}}{\mathrm{d}\beta} \mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\} + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\beta}[\eta - 1 - \log\eta] = 0. \tag{3.33}$$

Noticing that η is a function of β, (3.33) is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\eta} \mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\} + \frac{1}{2\beta}\left(1 - \eta^{-1}\right) = 0. \tag{3.34}$$

By the central formula that links the mutual information and MMSE in Gaussian channels (Theorem 2.1),

$$\frac{\mathrm{d}}{\mathrm{d}\eta} I(\eta\mathsf{snr}) = \frac{\mathsf{snr}}{2}\,\mathsf{mmse}(\eta\,\mathsf{snr}). \tag{3.35}$$

Thus (3.34) holds as η satisfies the fixed-point equation (3.22). ■

An interpretation of Theorem 3.1 through mutual information chain rule is given in Section 3.2.6.

Generalized Posterior Mean Estimation

Given $\xi > 0$, consider a postulated scalar Gaussian channel similar to (3.18) with input signal-to-noise ratio snr and inverse noise variance ξ:

$$Z = \sqrt{\mathsf{snr}}\,X + \frac{1}{\sqrt{\xi}}U \tag{3.36}$$

where $U \sim \mathcal{N}(0,1)$ is independent of X. Let the input distribution to this postulated channel be q_X. Denote the underlying measure of the postulated system by q. A *retrochannel* is characterized by the posterior probability distribution $q_{X|Z,\mathsf{snr};\xi}$, namely, it takes in an input Z and outputs a random variable X according to $q_{X|Z,\mathsf{snr};\xi}$. Note that the retrochannel is nothing but a materialization of the Bayes posterior distribution. The PME estimate of X given Z is therefore the posterior mean under the measure q:

$$\langle X\rangle_q = \mathsf{E}_q\left\{X \mid Z, \mathsf{snr}; \xi\right\}. \tag{3.37}$$

Consider now a concatenation of the scalar Gaussian channel (3.18) and the retrochannel as depicted in Figure 3.6. Let the input to the Gaussian channel (3.18) be denoted by X_0 to distinguish it from the output X of the retrochannel. The probability law of the composite channel is determined by snr and two parameters η and ξ. Let the generalized PME estimate be defined as in (3.37). We define the mean-square error of the estimate as

$$\mathsf{mse}(\mathsf{snr};\eta,\xi) = \mathsf{E}\left\{\left(X_0 - \langle X\rangle_q\right)^2 \,\middle|\, \mathsf{snr};\eta,\xi\right\}. \tag{3.38}$$

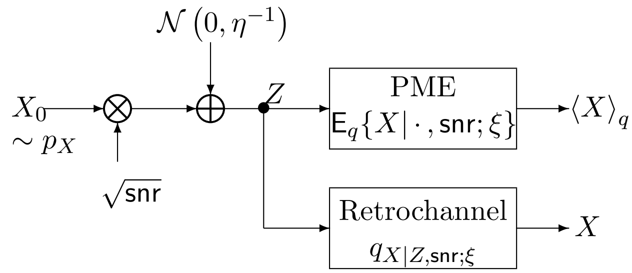

Figure 3.6: The equivalent single-user Gaussian channel, posterior mean estimator and retrochannel.

We also define the variance of the retrochannel as

$$\mathsf{var}(\mathsf{snr};\eta,\xi) = \mathsf{E}\left\{ \left(X - \langle X\rangle_q\right)^2 \,\middle|\, \mathsf{snr};\eta,\xi \right\}. \tag{3.39}$$

Note that in the special case where $q \equiv p$ and $\sigma = 1$, the postulated channel is identical to the actual channel, and consequently

$$\mathsf{mse}(\mathsf{snr};x,x) = \mathsf{var}(\mathsf{snr};x,x) = \mathsf{mmse}(x\,\mathsf{snr}), \quad \forall x. \tag{3.40}$$

We claim the following for the generalized multiuser PME.

Claim 3.3 *Let the generalized posterior mean estimator of the multiuser channel (3.4) be defined by (3.6) parameterized by the postulated input distribution q_X and noise level σ. Then, in the large-system limit, the distribution of the multiuser detection output $\langle X_k\rangle_q$ conditioned on $X_k = x$ being transmitted with signal-to-noise ratio snr_k is identical to the distribution of the generalized PME $\langle X\rangle_q$ of the equivalent scalar Gaussian channel (3.18) conditioned on $X = x$ being transmitted with input signal-to-noise ratio $\mathsf{snr} = \mathsf{snr}_k$, where the multiuser efficiency η and the inverse noise variance ξ of the postulated scalar channel (3.36) satisfy the coupled equations:*

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{mse}(\mathsf{snr};\eta,\xi)\right\}, \tag{3.41a}$$

$$\xi^{-1} = \sigma^2 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{var}(\mathsf{snr};\eta,\xi)\right\}, \tag{3.41b}$$

where the expectations are taken over P_{snr}. In case of multiple solutions to (3.41), (η,ξ) is chosen to minimize the free energy expressed as

$$\begin{aligned}\mathcal{F} = &-\mathsf{E}\left\{\int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta)\,\log q_{Z|\mathsf{snr};\xi}(z|\mathsf{snr};\xi)\,\mathrm{d}\,z\right\}\\ &-\frac{1}{2}\log\frac{2\pi}{\xi} - \frac{\xi}{2\eta} + \frac{\sigma^2\xi(\eta-\xi)}{2\beta\eta} + \frac{1}{2\beta}(\xi - 1 - \log\xi) + \frac{1}{2\beta}\log(2\pi) + \frac{\xi}{2\beta\eta}.\end{aligned} \tag{3.42}$$

Claim 3.3 is a generalization of Claim 3.1 to the case of generalized multiuser PME. The multiuser detector in this case is parameterized by (q_X,σ). Nonetheless, the single-user channel seen at the output of the generalized PME is equivalent to a degraded Gaussian channel followed by a decision function. However, the multiuser efficiency, same for all users, is now a solution to the coupled fixed-point equations (3.41), which is obtained using the replica method to be shown later. In case of multiple solutions, only the parameters (η, ξ) that minimize the free energy (3.42) conform with the operational meaning of the communication system.

The decision function (3.37) is akin to (3.7):

$$\mathsf{E}_q \{ X \mid Z, \mathsf{snr}; \xi \} = \frac{q_1(Z, \mathsf{snr}; \xi)}{q_0(Z, \mathsf{snr}; \xi)} \tag{3.43}$$

where

$$q_i(z, \mathsf{snr}; \xi) = \mathsf{E}_q \left\{ X^i \, q_{Z|X,\mathsf{snr};\xi}(z|X, \mathsf{snr}; \xi) \,\middle|\, \mathsf{snr} \right\}, \quad i = 0, 1, \dots. \tag{3.44}$$

Some algebra leads to

$$\mathsf{mse}(\mathsf{snr}; \eta, \xi) = 1 + \int p_0(z, \mathsf{snr}; \eta) \frac{q_1^2(z, \mathsf{snr}; \xi)}{q_0^2(z, \mathsf{snr}; \xi)} - 2p_1(z, \mathsf{snr}; \eta) \frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)} \, \mathrm{d}z \tag{3.45}$$

and

$$\mathsf{var}(\mathsf{snr}; \eta, \xi) = \int p_0(z, \mathsf{snr}; \eta) \frac{q_0(z, \mathsf{snr}; \xi) q_2(z, \mathsf{snr}; \xi) - q_1^2(z, \mathsf{snr}; \xi)}{q_0^2(z, \mathsf{snr}; \xi)} \, \mathrm{d}z. \tag{3.46}$$

Using (3.45) and (3.46), it is in general viable to find solutions to (3.41) numerically. Since the decision function (3.43) is strictly monotone, the following result is straightforward:

Corollary 3.2 *In the large-system limit, the mutual information between one user's input symbol and the output of the generalized multiuser posterior mean estimator for this user is equal to the input-output mutual information of the equivalent scalar Gaussian channel (3.18) with the same input distribution and SNR, and an inverse noise variance η as the multiuser efficiency given by Claim 3.3.*

3.2.5 Recovering Known Results

As shown in 3.2.3, several well-known multiuser detectors can be regarded as the generalized PME with appropriate parameters. Thus many previously known results can be recovered as special case of the new findings in Section 3.2.4.

Linear Detectors

For linear multiuser detectors, standard Gaussian prior is postulated for the generalized multiuser PME as well as the postulated scalar channel (3.36). Since the input Z and output X of the retrochannel are jointly Gaussian (refer to Figure 3.6), the PME is simply a linear attenuator (cf. (2.11)):

$$\langle X \rangle_q = \frac{\xi \sqrt{\mathsf{snr}}}{1 + \xi \mathsf{snr}} Z. \tag{3.47}$$

The variance of X conditioned on Z is independent of Z. Hence the variance of the retrochannel output is independent of η (cf. (2.12)):

$$\mathsf{var}(\mathsf{snr}; \eta, \xi) = \frac{1}{1 + \xi \mathsf{snr}}. \tag{3.48}$$

From Claim 3.3, one finds that ξ is the solution to

$$\xi^{-1} = \sigma^2 + \beta\,\mathsf{E}\left\{\frac{\mathsf{snr}}{1+\xi\mathsf{snr}}\right\}. \tag{3.49}$$

Meanwhile, the mean-square error is

$$\begin{aligned}\mathsf{E}\left\{\left(X_0 - \langle X\rangle_q\right)^2\right\} &= \mathsf{E}\left\{\left[X_0 - \frac{\xi\sqrt{\mathsf{snr}}}{1+\xi\mathsf{snr}}\left(\sqrt{\mathsf{snr}}X_0 + \frac{1}{\sqrt{\eta}}N\right)\right]^2\right\} &(3.50)\\ &= \frac{\eta+\xi^2\mathsf{snr}}{\eta(1+\xi\mathsf{snr})^2}. &(3.51)\end{aligned}$$

After some algebra, the multiuser efficiency is determined as

$$\eta = \xi + \xi\left(\sigma^2 - 1\right)\left[1 + \beta\,\mathsf{E}\left\{\frac{\mathsf{snr}}{(1+\xi\mathsf{snr})^2}\right\}\right]^{-1}. \tag{3.52}$$

Clearly, the large-system multiuser efficiency of such a linear detector is independent of the input distribution.

For the matched filter, we let the postulated noise level $\sigma \to \infty$. One finds $\xi\sigma^2 \to 1$ by (3.49) and consequently, the multiuser efficiency of matched filter is [112]

$$\eta^{(\mathrm{mf})} = \frac{1}{1+\beta\,\mathsf{E}\left\{\mathsf{snr}\right\}}. \tag{3.53}$$

In case of MMSE detector, the control parameter $\sigma = 1$. By (3.52), one finds that $\eta = \xi$ and by (3.49), the multiuser efficiency η satisfies the Tse-Hanly equation [99, 105]:

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\frac{\mathsf{snr}}{1+\eta\mathsf{snr}}\right\}, \tag{3.54}$$

which has a unique positive solution $\eta^{(\mathrm{mmse})}$.

In case of decorrelator, the control parameter $\sigma \to 0$. If $\beta < 1$, then (3.49) gives $\xi \to \infty$ and $\xi\sigma^2 \to 1-\beta$, and the multiuser efficiency is found as $\eta = 1-\beta$ by (3.52) regardless of the SNR distribution. If $\beta > 1$, and assuming the generalized form of the decorrelator as the Moore-Penrose inverse of the correlation matrix [112], then ξ is the unique solution to

$$\xi^{-1} = \beta\,\mathsf{E}\left\{\frac{\mathsf{snr}}{1+\xi\mathsf{snr}}\right\} \tag{3.55}$$

and the multiuser efficiency is found by (3.52) with $\sigma = 0$. In the special case of equal SNR from all users, an explicit expression can be found [23, 39]

$$\eta^{(\mathrm{dec})} = \frac{\beta-1}{\beta+\mathsf{snr}(\beta-1)^2}, \qquad \beta > 1. \tag{3.56}$$

By Corollary 3.2, the mutual information with input distribution p_X for a user with signal-to-noise ratio snr under linear multiuser detection is the input-output mutual information of the scalar Gaussian channel (3.18) with the same input:

$$I(X;\langle X\rangle_q|\mathsf{snr}) = I(\eta\,\mathsf{snr}), \tag{3.57}$$

where η depends on which type of linear detector is in use. Gaussian priors are known to achieve the capacity:

$$\mathsf{C}(\mathsf{snr}) = \frac{1}{2}\log(1+\eta\,\mathsf{snr}). \tag{3.58}$$

By Corollary 3.2, the total spectral efficiency under Gaussian inputs is expressed in terms of the linear MMSE multiuser efficiency:

$$\mathsf{C}_{\text{joint}}^{(\text{Gaussian})} = \frac{\beta}{2}\,\mathsf{E}\left\{\log\left(1+\eta^{(\text{mmse})}\mathsf{snr}\right)\right\} + \frac{1}{2}\left(\eta^{(\text{mmse})} - 1 - \log\eta^{(\text{mmse})}\right). \tag{3.59}$$

This is exactly Shamai and Verdú's result for fading channels [85].

Optimal Detectors

Using the actual input distribution p_X as the postulated priors of the generalized PME results in optimum multiuser detectors. In case of the jointly optimum detector, the postulated noise level $\sigma = 0$, and (3.41) becomes

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{mse}(\mathsf{snr};\eta,\xi)\right\}, \tag{3.60a}$$

$$\xi^{-1} = \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{var}(\mathsf{snr};\eta,\xi)\right\}, \tag{3.60b}$$

where $\mathsf{mse}(\cdot)$ and $\mathsf{var}(\cdot)$ are given by (3.45) and (3.46) with $q_i(z,\mathsf{snr};x) = p_i(z,\mathsf{snr};x)$, $\forall x$. The parameters have to be solved numerically.

In case of the individually optimal detector, $\sigma = 1$. It is clear that the postulated measure $q_{\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{S}}$ is identical to the true measure $p_{\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{S}}$. In the equivalent scalar channel and its retrochannel, the parameters satisfy

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{mse}(\mathsf{snr};\eta,\xi)\right\}, \tag{3.61a}$$

$$\xi^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{var}(\mathsf{snr};\eta,\xi)\right\}. \tag{3.61b}$$

We take the symmetric solution $\eta = \xi$ due to (3.40). The multiuser efficiency η is thus the solution to the fixed-point equation (3.22) given in Claim 3.1. It should be cautioned that (3.61) may have other solutions with $\eta \neq \xi$ in the unlikely case that replica symmetry does not hold.

It is of practical interest to find the spectral efficiency under the constraint that the input symbols are antipodally modulated as in the popular BPSK. In this case, the distribution $p_X(x) = 1/2$, $x = \pm 1$, maximizes the mutual information. The MMSE is given by (2.16). By Claim 3.1, The multiuser efficiency, $\eta^{(\text{b})}$, where the superscript (b) stands for binary inputs, is a solution to the fixed-point equation:

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\left[1 - \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh\left(\eta\mathsf{snr} - z\sqrt{\eta\mathsf{snr}}\right)\,\mathrm{d}z\right]\right\}, \tag{3.62}$$

which is a generalization to unequal-power distribution [39] of an earlier result on equal SNRs due to Tanaka [96]. The single-user channel capacity for a user with signal-to-noise ratio snr is the same as that obtained by Müller and Gerstacker [72] and is given by (2.17) with snr replaced by $\eta^{(\text{b})}\,\mathsf{snr}$. The total spectral efficiency of the CDMA channel subject to binary inputs is thus

$$\begin{aligned}\mathsf{C}_{\text{joint}}^{(\text{b})} =& \beta\,\mathsf{E}\left\{\eta^{(\text{b})}\,\mathsf{snr} - \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \log\cosh\left(\eta^{(\text{b})}\mathsf{snr} - z\sqrt{\eta^{(\text{b})}\mathsf{snr}}\right)\,\mathrm{d}z\right\} \\ &+ \frac{1}{2}\left[\eta^{(\text{b})} - 1 - \log\eta^{(\text{b})}\right],\end{aligned} \tag{3.63}$$

which is also a generalization in [39] of Tanaka's implicit result [96].

3.2.6 Discussions

Successive Interference Cancellation

Theorem 3.1 is an outcome of the chain rule of mutual information, which holds for all inputs and arbitrary number of users:

$$I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) = \sum_{k=1}^{K} I(X_k;\boldsymbol{Y}|\boldsymbol{S}, X_{k+1},\ldots,X_K). \tag{3.64}$$

The left hand side of (3.64) is the total capacity of the multiuser channel. Each summand on the right hand side of (3.64) is a single-user capacity over the multiuser channel conditioned on the symbols of previously decoded users. As argued in the following, the limit of (3.64) as $K \to \infty$ becomes the integral equation (3.31).

We conceive an interference canceler that decodes the users successively in which reliably decoded symbols as well as the generalized PME estimates of the yet undecoded users are used to reconstruct the interference for cancellation. Since the error probability of intermediate decisions vanishes with code block-length, the MAI from decoded users are asymptotically completely removed. Without loss of generality assume that the users are decoded in reverse order, then the generalized PME for user k sees only $k-1$ interfering users. Hence the performance for user k under successive decoding is identical to that under parallel separate decoding in a CDMA system with k instead of K users. Nonetheless, the equivalent single-user channel for each user is Gaussian by Claim 3.3. The multiuser efficiency experienced by user k is $\eta\left(\frac{k}{L}\right)$ where we use the fact that it is a function of the load $\frac{k}{L}$ seen by the generalized PME for user k. Following (3.26), the single-user capacity for user k is therefore

$$I\left(\eta\left(\frac{k}{L}\right)\mathsf{snr}_k\right). \tag{3.65}$$

Assuming that the i.i.d. snr_k are not dependent on the indexes k, the overall spectral efficiency under successive decoding converges almost surely:

$$\frac{1}{L}\sum_{k=1}^{K} I\left(\eta\left(\frac{k}{L}\right)\mathsf{snr}_k\right) \to \mathsf{E}\left\{\int_0^{\beta} I(\beta'\,\mathsf{snr})\,\mathrm{d}\beta'\right\}. \tag{3.66}$$

Note that the above result on successive decoding is true for arbitrary input distribution p_X and generalized PME detectors. In the special case of the PME, for which the postulated system is identical to the actual one, the right hand side of (3.66) is equal to $\mathsf{C}_{\text{joint}}(\beta)$ by Theorem 3.1. We can summarize this principle as:

Proposition 3.1 *In the large-system limit, successive decoding with a PME front end against yet undecoded users achieves the optimal CDMA channel capacity under arbitrary input distributions.*

Proposition 3.1 is a generalization of the previous result that a successive canceler with a linear MMSE front end against undecoded users achieves the capacity of the CDMA channel under Gaussian inputs [102, 80, 105, 2, 70, 33]. In the special case of Gaussian inputs, however, the optimality is known to hold for any finite number of users [102, 105].

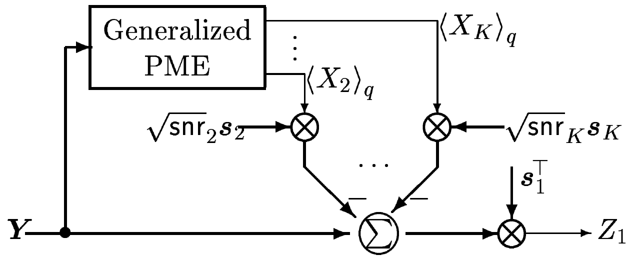

Figure 3.7: A canonical interference canceler equivalent to the single-user Gaussian channel.

Parallel Interference Cancellation Puzzle

We give a quick and specious interpretation of the large-system results. The assumptions that will be made are erroneous, hence the interpretation is fundamentally rootless. However, the argument leads surprisingly to the correct result. We suspect that this interpretation can be fixed to obtain a correct justification of the results. The author has failed but nonetheless would like to include the argument in case the reader may turn it to a success.

We construct without loss of generality an estimator for user 1 using interference cancellation as depicted in Figure 3.7. Let $\langle X_2 \rangle_q, \dots, \langle X_K \rangle_q$ be the generalized PME estimates for user 2 through user K. A decision statistic for user 1 can be generated by first subtracting the reconstructed interferences using those estimates and then matched filtering with respect to user 1's spreading sequence:

$$Z_1 = \sqrt{\mathsf{snr}_1} X_1 + \sum_{k=2}^{K} \boldsymbol{s}_1^\top \boldsymbol{s}_k \sqrt{\mathsf{snr}_k} \left(X_k - \langle X_k \rangle_q \right) + N_1 \tag{3.67}$$

where N_1 is a standard Gaussian random variable. If the desired symbol X_1, the Gaussian noise N_1, and the residual errors $\left(X_k - \langle X_k \rangle_q \right)$ were independent, by virtue of the central limit theorem, the sum of the residual MAI and Gaussian noise converges to a Gaussian random variable as $K \to \infty$. The variance of $X_k - \langle X_k \rangle_q$ is $\mathsf{mse}(\mathsf{snr}_k; \eta, \xi)$ by Claim 3.3. The variance of the total interference in (3.67) is therefore

$$1 + \beta \, \mathsf{E} \left\{ \mathsf{snr} \cdot \mathsf{mse}(\mathsf{snr}; \eta, \xi) \right\}, \tag{3.68}$$

which, by the fixed-point equation (3.41a) in Claim 3.3, is equal to η^{-1}. Thus if the independence assumption were valid, we would have found a degraded Gaussian channel for user 1 equivalent to the single-user channel as shown in Figure 3.6. That is also to say that, given the generalized PME estimates of all interfering users, the interference canceler produces for the desired user an estimate that is as good as the generalized PME output. We can also argue that every user enjoys the same efficiency since otherwise users with worse efficiency may benefit from users with better efficiency until an equilibrium is reached. Roughly speaking, the generalized PME output is a "fixed-point" of a parallel interference canceler. The multiuser efficiency, in a sense, is the outcome of such an equilibrium.

It should be emphasized that the above interpretation does not hold due to the erroneous independence assumption. In particular, $\boldsymbol{s}_1^\top \boldsymbol{s}_k$ are not independent, albeit uncorrelated, for all k. Also, $\langle X_k \rangle_q$ is dependent on the desired signal X_1 and the noise N_1. This is evident in the special case of linear MMSE detection. One is tempted to fix the argument by first decorrelating the residual errors, the desired symbol and the noise to satisfy the requirement of central limit theorem. It seems to be possible because correlation between the MAI and the desired signal may neutralize the correlations among the components of the MAI as well as with the noise. The author has failed to show this.

3.3 Multiuser Communications and Statistical Physics

In this section, we prepare the reader with concepts and methodologies that will be needed to prove the results given in Section 3.2.4. The replica method, which we will make use of, was originally developed in spin glass theory in statistical physics [22]. Although one can work with the mathematical framework only and avoid foreign concepts, we believe it is more enlightening to draw an equivalence in between multiuser communications and many-body problems in statistical physics. Such an analogy is first seen in a primitive form in [96] and will be developed to a full generality here.

3.3.1 A Note on Statistical Physics

Let the microscopic state of a system be described by the configuration of some K variables as a vector $\boldsymbol{x}$. The *Hamiltonian* is a function of the configuration, denoted by $H(\boldsymbol{x})$. The state of the system evolves over time according to some physical laws, and after long enough time it reaches thermal equilibrium. The time average of an observable quantity can be obtained by averaging over the ensemble of the states. In particular, the *energy* of the system is

$$\mathcal{E} = \sum_{\boldsymbol{x}} p(\boldsymbol{x}) H(\boldsymbol{x}) \tag{3.69}$$

where $p(\boldsymbol{x})$ is the probability of the system being found in configuration $\boldsymbol{x}$. In other words, as far as the macroscopic properties are concerned, it suffices to describe the system statistically instead of solving the exact dynamic trajectories. Another fundamental quantity is the *entropy*, defined as

$$\mathcal{S} = -\sum_{\boldsymbol{x}} p(\boldsymbol{x}) \log p(\boldsymbol{x}). \tag{3.70}$$

It is assumed that the system is not isolated and may interact with the surroundings. As a result, at thermal equilibrium, the energy of the system remains a constant and the entropy is the maximum possible.

Given the energy $\mathcal{E}$, one can use the Lagrange multiplier method to show that the equilibrium probability distribution that maximizes the entropy is the Boltzmann distribution

$$p(\boldsymbol{x}) = Z^{-1} \exp\left[-\frac{1}{T} H(\boldsymbol{x})\right] \tag{3.71}$$

where

$$Z = \sum_{\boldsymbol{x}} \exp\left[-\frac{1}{T} H(\boldsymbol{x})\right] \tag{3.72}$$

is the *partition function*, and the parameter T is the *temperature*, which is determined by the energy constraint (3.69). The system is found in each configuration with a probability that is negative exponential in the Hamiltonian associated with the configuration. The most probable configuration is the ground state which has the minimum Hamiltonian.

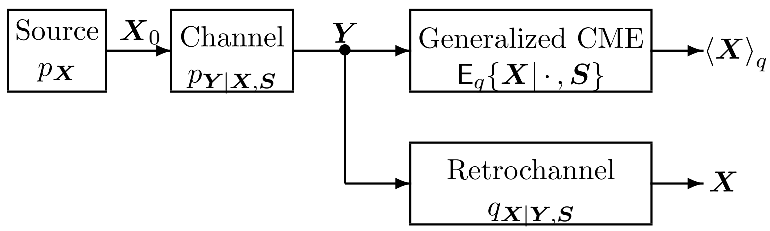

Figure 3.8: A canonical channel, its corresponding retrochannel, and the generalized PME.

One particularly useful macroscopic quantity of the thermodynamic system is the *free energy*, defined as

$$\mathcal{E} - T\,\mathcal{S}. \tag{3.73}$$

Since the entropy is the maximum at equilibrium, the free energy is at its minimum. Using (3.69)–(3.72), one finds that the free energy at equilibrium can also be expressed as $-T\,\log Z$. The free energy is often the starting point for calculating macroscopic properties of a thermodynamic system.

3.3.2 Communications and Spin Glasses

The statistical inference problem faced by an estimator can be described as follows. A (vector) source symbol $\boldsymbol{X}_0$ is drawn according to a prior distribution $p_{\boldsymbol{X}}$. The channel response to the input $\boldsymbol{X}_0$ is an output $\boldsymbol{Y}$ generated according to a conditional probability distribution $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ where $\boldsymbol{S}$ is the channel state. The estimator would like to draw some conclusion about the original symbol $\boldsymbol{X}_0$ upon receiving $\boldsymbol{Y}$ using knowledge about the state $\boldsymbol{S}$. Naturally, the posterior distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is central in the statistical inference problem.

If $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is revealed to the estimator, we have the posterior mean estimator, which is optimal in mean-square sense. One may choose, however, to supply any posterior distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ in lieu of $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$, henceforth resulting a generalized PME: $\langle\boldsymbol{X}\rangle_q = \mathsf{E}_q\{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}\}$. As shown in Section 3.2.3, the freedom in choosing the postulated measure allows treatment of various detection techniques in a uniform framework.

It is conceptually helpful here to also introduce the retrochannel induced by the postulated system. The multiuser channel, the generalized multiuser PME, and the associated retrochannel are depicted in Figure 3.8 (to be compared with its single-user counterpart in Figure 3.6). Upon receiving a signal $\boldsymbol{Y}$ with a channel state $\boldsymbol{S}$, the retrochannel outputs a random vector according to the probability distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$, which is induced by the postulated prior distribution $q_{\boldsymbol{X}}$ and the postulated conditional distribution $q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$. Clearly, given $(\boldsymbol{Y},\boldsymbol{S})$, the generalized PME output $\langle\boldsymbol{X}\rangle_q$ is the expected value of the retrochannel output $\boldsymbol{X}$.

In the multiple-access channel (3.4), the channel state consists of the spreading sequences and the SNRs, collectively represented by matrix $\boldsymbol{S}$. The conditional distribution $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ is a Gaussian density (3.16). In this work, the postulated channel (3.17) differs from the actual one only in the noise variance. Assuming an input distribution of q_X, the posterior distribution of the postulated channel can be obtained by using the Bayes formula (cf. (3.7)) as

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = \left[q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\right]^{-1} q_{\boldsymbol{X}}(\boldsymbol{x}) \left(2\pi\sigma^2\right)^{-\frac{L}{2}} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{3.74}$$

where

$$q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S}) = \left(2\pi\sigma^2\right)^{-\frac{L}{2}} \mathsf{E}_q\left\{\exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{X}\|^2\right] \middle| \boldsymbol{S}\right\} \tag{3.75}$$

and where the expectation in (3.75) is taken conditioned on $\boldsymbol{S}$ over $\boldsymbol{X}$ with distribution $q_{\boldsymbol{X}}$.

Interestingly, we can associate the posterior probability distribution (3.74) with the characteristics of a thermodynamic system called spin glass. We believe this work is the first to draw this analogy in the general setting, although in certain special cases the argument is found in Tanaka's important paper [96]. A *spin glass* is a system consisting of many directional spins, in which the interaction of the spins is determined by the so-called *quenched random variables* whose values are determined by the realization of the spin glass.[4] Let the microscopic state of a spin glass be described by a K-dimensional vector $\boldsymbol{x}$. Let the quenched random variables be denoted by $(\boldsymbol{y}, \boldsymbol{S})$. The system can be understood as K random spins sitting in quenched randomness $(\boldsymbol{y}, \boldsymbol{S})$. The basic quantity characterizing a microscopic state is the Hamiltonian, which is a function of the configuration dependent on the quenched randomness, denoted by $H_{\boldsymbol{y},\boldsymbol{S}}(\boldsymbol{x})$. At thermal equilibrium, the spin glass is found in each configuration with the Boltzmann distribution:

$$q(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = Z^{-1}(\boldsymbol{y},\boldsymbol{S}) \exp\left[-\frac{1}{T}H_{\boldsymbol{y},\boldsymbol{S}}(\boldsymbol{x})\right] \tag{3.76}$$

where

$$Z(\boldsymbol{y},\boldsymbol{S}) = \sum_{\boldsymbol{x}} \exp\left[-\frac{1}{T}H_{\boldsymbol{y},\boldsymbol{S}}(\boldsymbol{x})\right] \tag{3.77}$$

is the partition function. This equilibrium distribution maximizes the entropy subject to energy constraint.

The reader may have noticed the similarity between (3.74)–(3.75) and (3.76)–(3.77). Indeed, if the temperature $T = 1$ in (3.76)–(3.77) and that the Hamiltonian is defined as

$$H_{\boldsymbol{y},\boldsymbol{S}}(\boldsymbol{x}) = \frac{L}{2}\log\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{x}\|^2 - \log q_{\boldsymbol{X}}(\boldsymbol{x}), \tag{3.78}$$

then

$$q(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}), \tag{3.79}$$

$$Z(\boldsymbol{y},\boldsymbol{S}) = q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S}). \tag{3.80}$$

In other words, by defining an appropriate Hamiltonian, the configuration distribution of the spin glass at equilibrium is identical to the posterior probability distribution associated with a multiuser communication system. Precisely, the probability that the transmitted symbol is $\boldsymbol{x}$ under the postulated model, given the observation $\boldsymbol{y}$ and the channel state $\boldsymbol{S}$, is equal to the probability that the spin glass is at configuration $\boldsymbol{x}$, given the values of the quenched

[4] An example is a system consisting molecules with magnetic spins that evolve over time, while the positions of the molecules that determine the amount of interactions are random (disordered) but remain fixed for each concrete instance as in a piece of glass (hence the name of spin glass).

random variables $(\boldsymbol{Y}, \boldsymbol{S}) = (\boldsymbol{y}, \boldsymbol{S})$. It is interesting to note that Gaussian distribution is a natural Boltzmann distribution with squared Euclidean norm as the Hamiltonian.

The quenched randomness $(\boldsymbol{Y}, \boldsymbol{S})$ takes a specific distribution in our problem, i.e., $(\boldsymbol{Y}, \boldsymbol{S})$ is a realization of the received signal and channel state matrix according to the prior and conditional distributions that underlie the "original" spins. Indeed, the communication system depicted in Figure 3.8 can be also understood as a spin glass $\boldsymbol{X}$ under physical law q sitting in the quenched randomness caused by another spin glass $\boldsymbol{X}_0$ under physical law p. The channel corresponds to the random mapping from a given spin glass configuration to an induced quenched randomness. Conversely, the retrochannel corresponds to the random mechanism that maps some quenched randomness into an induced spin glass configuration.

The free energy of the thermodynamic (or communication) system is obtained as:

$$-T \, \log Z(\boldsymbol{Y}, \boldsymbol{S}). \tag{3.81}$$

In the following, we show some useful properties of the free energy and associate it with the spectral efficiency of communication systems.

3.3.3 Free Energy and Self-averaging Property

The free energy (3.81) with $T = 1$ and normalized by the number of users is (via (3.80))

$$-\frac{1}{K} \log q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S}). \tag{3.82}$$

As mentioned in Section 3.1, the randomness in (3.82) vanishes as $K \to \infty$ due to the self-averaging assumption. As a result, the free energy normalized by the number of users converges in probability to its expected value over the distribution of the quenched random variables $(\boldsymbol{Y}, \boldsymbol{S})$ in the large-system limit, which is denoted by $\mathcal{F}$,

$$\mathcal{F} = - \lim_{K\to\infty} \mathsf{E}\left\{ \frac{1}{K} \log Z(\boldsymbol{Y}, \boldsymbol{S}) \right\}. \tag{3.83}$$

Hereafter, by the free energy we refer to the large-system limit $\mathcal{F}$, which we will calculate in Section 3.4.

The reader should be cautioned that for disordered systems, thermodynamic quantities may or may not be self-averaging [14]. The self-averaging property remains to be proved or disproved in the CDMA context. This is a challenging problem on its own. In this work we take the self-averaging property for granted with a good faith that it be correct. Again, we believe that even if the self-averaging property breaks down, the main results in this chapter are good approximations and still useful to some extent in practice.

The self-averaging property resembles the asymptotic equipartition property (AEP) in information theory [17]. An important consequence is that a macroscopic quantity of a thermodynamic system, which is a function of a large number of random variables, may become increasingly predictable from merely a few parameters independent of the realization of the random variables as the system size grows without bound. Indeed, the macroscopic quantity converges in probability to its ensemble average in the thermodynamic limit.

In the CDMA context, the self-averaging property leads to a strong consequence that for almost all realizations of the received signal and the spreading sequences, macroscopic quantities such as the BER, the output SNR and the spectral efficiency, averaged over data, converge to deterministic quantities in the large-system limit. Previous work (e.g.

[105, 99, 44]) has shown convergence of performance measures for almost all spreading sequences. The self-averaging property results in convergence of empirical measures of error performance and information rate, which holds for almost all realizations of the data and noise.

3.3.4 Spectral Efficiency of Jointly Optimal Decoding

This section associates the optimal spectral efficiency of a multiuser communication system with the free energy of a corresponding spin glass. This is again a full generalization of an observation by Tanaka [96] in special cases. In fact, the analogy between free energy and information-theoretic quantities has been noticed in belief propagation [117], coding [101] and optimization problems [11] as well.

For a fixed input distribution p_X, the total mutual information under joint decoding is

$$I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) = \mathsf{E}\left\{ \log \frac{p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S})}{p_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S})} \,\middle|\, \boldsymbol{S} \right\} \tag{3.84}$$

where the expectation is taken over the conditional joint distribution $p_{\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{S}}$. Since the channel characteristic given by (3.16) is a standard L-dimensional Gaussian density, one has

$$\mathsf{E}\left\{ \log p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}) \,\middle|\, \boldsymbol{S} \right\} = -\frac{L}{2}\log(2\pi e). \tag{3.85}$$

Suppose that the postulated measure q is the same as the actual measure p, then by (3.80), (3.84) and (3.85), the spectral efficiency in nats per degree of freedom achieved by optimal joint decoding is

$$\begin{aligned} \mathsf{C}(\boldsymbol{S}) &= \frac{1}{L} I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) && (3.86)\\ &= -\beta\,\mathsf{E}\left\{ \frac{1}{K}\log Z(\boldsymbol{Y},\boldsymbol{S}) \,\middle|\, \boldsymbol{S} \right\} - \frac{1}{2}\log(2\pi e). && (3.87) \end{aligned}$$

To calculate (3.87) is formidable for an arbitrary realization of $\boldsymbol{S}$ but due to the self-averaging property, the spectral efficiency converges in probability as $K, N \to \infty$ to

$$\mathsf{C} = \beta \mathcal{F}|_{q=p} - \frac{1}{2}\log(2\pi e) \tag{3.88}$$

where $\mathcal{F}$ is defined in (3.83). Note that the constant term in (3.88) can be removed by redefining the partition function up to a constant coefficient. In either way, the spectral efficiency under optimal joint decoding is affine in the free energy under a postulated measure q identical to the true measure p.

3.3.5 Separate Decoding

In case of a multiuser detector front end, one is interested in the distribution of the detection output $\langle X_k \rangle_q$ conditioned on the input X_{0k}. Here, X_{0k} is used to denote the input to the CDMA channel to distinguish it from the retrochannel output X_k. Our approach is to calculate joint moments

$$\mathsf{E}\left\{ X_{0k}^j \, \langle X_k \rangle_q^i \,\middle|\, \boldsymbol{S} \right\}, \quad i, j = 0, 1, \ldots \tag{3.89}$$

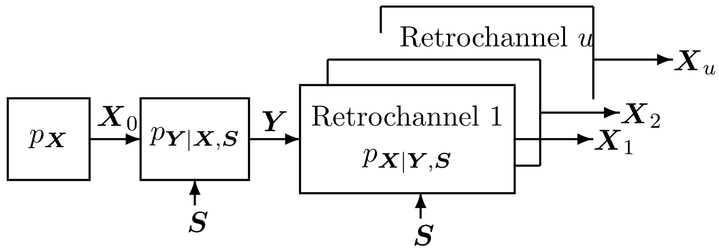

Figure 3.9: The replicas of the retrochannel.

and then infer the distribution of $(\langle X_k \rangle_q - X_{0k})$. In the spin glass context, one may interpret (3.89) as the joint moments of the spins $\boldsymbol{X}_0$ that induced the quenched randomness, and the conditional expectation of the induced spins $\boldsymbol{X}$. The calculation of (3.89) is again formidable due to its dependence on the channel state, but in the large-system limit, the result is surprisingly simple. Due to the self-averaging property, the moments also converge in probability, and it suffices to calculate the moments with respect to the joint distribution of the spins and the quenched randomness as

$$\lim_{K\to\infty} \mathsf{E}\left\{ X_{0k}^j \, \langle X_k \rangle_q^i \right\}. \tag{3.90}$$

Note that $\boldsymbol{X}_0 \to (\boldsymbol{Y}, \boldsymbol{S}) \to \boldsymbol{X}$ is a Markov chain. It can be shown that (3.90) is equivalent to

$$\lim_{K\to\infty} \mathsf{E}\left\{ X_{0k}^j \, X_k^i \right\}, \tag{3.91}$$

which turns out to be easier to calculate by studying the free energy associated with a modified version of the partition function (3.75). More on this later.

The mutual information between the input and the output of a multiuser detector front end for an arbitrary user k is given by

$$I(X_{0k}; \langle X_k \rangle_q \mid \boldsymbol{S}), \tag{3.92}$$

which can be derived once the input-output relationship is known. It will be shown that conditioning on the channel state $\boldsymbol{S}$ is asymptotically inconsequential.

We have distilled our problems under both joint and separate decoding to finding some ensemble averages, namely, the free energy (3.83) and the moments (3.91). In order to calculate these quantities, we resort to a powerful technique developed in the theory of spin glass, the heart of which is sketched in the following subsection.

3.3.6 Replica Method

The replica method was introduced to the field of multiuser detection by Tanaka to analyze optimal detectors under equal power Gaussian or binary input. In the following we outline the method in a more general setting following Tanaka's pioneering work [96].

The expected value of the logarithm in (3.83) can be reformulated as

$$\mathcal{F} = -\lim_{K\to\infty} \frac{1}{K} \lim_{u\to 0} \frac{\partial}{\partial u} \log \mathsf{E}\left\{ Z^u(\boldsymbol{Y}, \boldsymbol{S}) \right\}. \tag{3.93}$$

The equivalence of (3.83) and (3.93) can be easily verified by noticing that

$$\lim_{u\to 0}\frac{\partial}{\partial u}\log \mathsf{E}\left\{\Theta^u\right\} = \lim_{u\to 0}\frac{\mathsf{E}\left\{\Theta^u\log\Theta\right\}}{\mathsf{E}\left\{\Theta^u\right\}} = \mathsf{E}\left\{\log\Theta\right\}, \quad \forall\Theta. \tag{3.94}$$

For an arbitrary integer replica number u, we introduce u independent replicas of the retrochannel (or the spin glass) with the same received signal $\boldsymbol{Y}$ and channel state matrix $\boldsymbol{S}$ as depicted in Figure 3.9. Conditioned on $(\boldsymbol{Y},\boldsymbol{S})$, $\boldsymbol{X}_a$ are independent. By (3.80), the partition function of the replicated system is

$$Z^u(\boldsymbol{y},\boldsymbol{S}) = \mathsf{E}_q\left\{\prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_a,\boldsymbol{S}) \,\middle|\, \boldsymbol{S}\right\} \tag{3.95}$$

where the expectation is taken over the i.i.d. symbols $\{X_{ak}|a=1,\dots,u,\, k=1,\dots,K\}$, with distribution q_X. Note that X_{ak} are i.i.d. since $\boldsymbol{Y}=\boldsymbol{y}$ is given. We can henceforth evaluate

$$-\lim_{K\to\infty}\frac{1}{K}\log\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} \tag{3.96}$$

as a function of the integer u. The *replica trick* assumes that the resulting expression is also valid for an arbitrary real number u and finds the derivative at $u=0$ as the free energy. Besides validity of continuing to non-integer values of the replica number u, it is also necessary to assume that the two limits in (3.93) can be exchanged in order.

It remains to calculate (3.96). Note that $(\boldsymbol{Y},\boldsymbol{S})$ is induced by the transmitted symbols $\boldsymbol{X}_0$. By taking expectation over $\boldsymbol{Y}$ first and then averaging over the spreading sequences, one finds that

$$\frac{1}{K}\log\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} = \frac{1}{K}\log\mathsf{E}\left\{\exp\left[\beta^{-1}K\,G_K^{(u)}(\boldsymbol{\Gamma},\underline{\boldsymbol{X}})\right]\right\} \tag{3.97}$$

where $G_K^{(u)}$ is some function of the SNRs and the transmitted symbols and their replicas, collectively denoted by a $K\times(u+1)$ matrix $\underline{\boldsymbol{X}} = [\boldsymbol{X}_0,\dots,\boldsymbol{X}_u]$. By first conditioning on the correlation matrix $\boldsymbol{Q}$ of $\boldsymbol{\Gamma}\,\underline{\boldsymbol{X}}$, the central limit theorem helps to reduce (3.97) to

$$\frac{1}{K}\log\int\exp\left[\beta^{-1}K\,G^{(u)}(\boldsymbol{Q})\right]\mu_K^{(u)}(\,\mathrm{d}\boldsymbol{Q}) \tag{3.98}$$

where $G^{(u)}$ is some function of the $(u+1)\times(u+1)$ correlation matrix $\boldsymbol{Q}$, and $\mu_K^{(u)}$ is the probability measure of the random matrix $\boldsymbol{Q}$. Large deviations can be invoked to show that there exists a rate function $I^{(u)}$ such that the measure $\mu_K^{(u)}$ satisfies

$$-\lim_{K\to\infty}\frac{1}{K}\log\mu_K^{(u)}(\mathcal{A}) = \inf_{\boldsymbol{Q}\in\mathcal{A}} I^{(u)}(\boldsymbol{Q}) \tag{3.99}$$

for all measurable set $\mathcal{A}$ of $(u+1)\times(u+1)$ matrices. Using Varadhan's theorem [25], the integral (3.98) is found to converge as $K\to\infty$ to

$$\sup_{\boldsymbol{Q}}[\beta^{-1}\,G^{(u)}(\boldsymbol{Q}) - I^{(u)}(\boldsymbol{Q})]. \tag{3.100}$$

Seeking the extremum over a $(u+1)^2$-dimensional space is a hard problem. The technique to circumvent this is to assume *replica symmetry*, namely, that the supremum in $\boldsymbol{Q}$ is symmetric over all replicated dimensions. The resulting supremum is then over merely a few parameters, and the free energy can be obtained.

The replica method is also used to calculate (3.91). Clearly, $\boldsymbol{X}_0$—$(\boldsymbol{Y}, \boldsymbol{S})$—$[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_u]$ is a Markov chain. The moments (3.91) are equivalent to

$$\lim_{K\to\infty} \mathsf{E}\left\{ X_{0k}^j \prod_{m=1}^{i} X_{mk} \right\} \tag{3.101}$$

which can be readily evaluated by working with a modified partition function akin to (3.95).

Detailed replica analysis of the real-valued channel is carried out in Section 3.4. The complex-valued counterpart is discussed in Section 3.5. As mentioned in Section 3.1, while we assume the replica trick and replica symmetry to be valid as well as the self-averaging property, their fully rigorous justification is still open mathematical physics.

3.4 Proofs Using the Replica Method

This section proves Claims 3.1–3.3 using the replica method. The free energy is calculated first and hence the spectral efficiency under joint decoding is derived. The joint moments of the input and multiuser detection output are then found and it is demonstrated that the CDMA channel can be effectively decoupled into single-user Gaussian channels. Thus the multiuser efficiency as well as the spectral efficiency under separate decoding is found.

3.4.1 Free Energy

We will find the free energy by (3.93) and then the spectral efficiency is trivial by (3.88). From (3.95),

$$\begin{aligned}
\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} &= \mathsf{E}\left\{\int p_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S}) Z^u(\boldsymbol{y},\boldsymbol{S})\,\mathrm{d}\boldsymbol{y}\right\} &(3.102)\\
&= \mathsf{E}\left\{\int p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_0,\boldsymbol{S}) \prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_a,\boldsymbol{S})\,\mathrm{d}\boldsymbol{y}\right\} &(3.103)
\end{aligned}$$

where the expectation in (3.103) is taken over the channel state matrix $\boldsymbol{S}$, the original symbol vector $\boldsymbol{X}_0$ (i.i.d. entries with distribution p_X), and the replicated symbols $\boldsymbol{X}_a$, $a = 1, \ldots, u$ (i.i.d. entries with distribution q_X). Note that $\boldsymbol{S}$, $\boldsymbol{X}_0$ and $\boldsymbol{X}_a$ are independent. Let $\underline{\boldsymbol{X}} = [\boldsymbol{X}_0, \ldots, \boldsymbol{X}_u]$. Plugging (3.16) and (3.17) into (3.103),

$$\begin{aligned}
\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} =& \mathsf{E}\left\{\int (2\pi)^{-\frac{L}{2}} (2\pi\sigma^2)^{-\frac{uL}{2}} \exp\left[-\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{X}_0\|^2\right]\right.\\
&\left.\times \prod_{a=1}^{u} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{X}_a\|^2\right]\,\mathrm{d}\boldsymbol{y}\right\}.
\end{aligned} \tag{3.104}$$

We glean from the fact that the L dimensions of the CDMA channel are independent and statistically identical, and write (3.104) as

$$\begin{aligned}
\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} =& \mathsf{E}\left\{\left[\left(2\pi\sigma^2\right)^{-\frac{u}{2}} \int \mathsf{E}\left\{\exp\left[-\frac{1}{2}\left(y-\tilde{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{X}_0\right)^2\right]\right.\right.\right.\\
&\left.\left.\left.\times \prod_{a=1}^{u} \exp\left[-\frac{1}{2\sigma^2}\left(y-\tilde{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{X}_a\right)^2\right]\right|\boldsymbol{\Gamma},\underline{\boldsymbol{X}}\right\}\frac{\mathrm{d}y}{\sqrt{2\pi}}\right]^L\right\}
\end{aligned} \tag{3.105}$$

where the inner expectation in (3.105) is taken over $\tilde{\boldsymbol{S}} = [S_1, \dots, S_K]$, a vector of i.i.d. random variables each taking the same distribution as the random spreading chips S_{nk}. Define the following variables:

$$V_a = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sqrt{\mathsf{snr}_k}\, S_k X_{ak}, \quad a = 0, 1, \dots, u. \tag{3.106}$$

Clearly, (3.105) can be rewritten as

$$\mathsf{E}\left\{ Z^u(\boldsymbol{Y}, \boldsymbol{S}) \right\} = \mathsf{E}\left\{ \exp\left[L\, G_K^{(u)}\left(\boldsymbol{\Gamma}, \underline{\boldsymbol{X}}\right) \right] \right\} \tag{3.107}$$

where

$$\begin{aligned} G_K^{(u)}\left(\boldsymbol{\Gamma}, \underline{\boldsymbol{X}}\right) = -\, & \frac{u}{2} \log\left(2\pi\sigma^2\right) + \log \int \mathsf{E}\left\{ \exp\left[-\frac{1}{2}\left(y - \sqrt{\beta}\, V_0 \right)^2 \right] \right. \\ & \left. \times \prod_{a=1}^{u} \exp\left[-\frac{1}{2\sigma^2} \left(y - \sqrt{\beta}\, V_a \right)^2 \right] \middle| \boldsymbol{\Gamma}, \underline{\boldsymbol{X}} \right\} \frac{\mathrm{d}y}{\sqrt{2\pi}}. \end{aligned} \tag{3.108}$$

Note that given $\boldsymbol{\Gamma}$ and $\underline{\boldsymbol{X}}$, each V_a is a sum of K weighted i.i.d. random chips, and hence converges to a Gaussian random variable as $K \to \infty$. In fact, due to a generalization of the central limit theorem, $\boldsymbol{V}$ converges to a zero-mean Gaussian random vector with covariance matrix $\boldsymbol{Q}$ where

$$Q_{ab} = \mathsf{E}\left\{ V_a V_b \mid \boldsymbol{\Gamma}, \underline{\boldsymbol{X}} \right\} = \frac{1}{K} \sum_{k=1}^{K} \mathsf{snr}_k X_{ak} X_{bk}, \quad a, b = 0, \dots, u. \tag{3.109}$$

Note that although inexplicit in notation, Q_{ab} is a function of $\{\mathsf{snr}_k, X_{ak}, X_{bk}\}_{k=1}^{K}$. The reader is referred to Appendix A.4 for a justification of the asymptotic normality of $\boldsymbol{V}$ through the Edgeworth expansion. As a result,

$$\exp\left[G_K^{(u)}(\boldsymbol{\Gamma}, \underline{\boldsymbol{X}}) \right] = \exp\left[G^{(u)}(\boldsymbol{Q}) + \mathcal{O}(K^{-1}) \right] \tag{3.110}$$

where

$$\begin{aligned} G^{(u)}(\boldsymbol{Q}) = -\, & \frac{u}{2} \log\left(2\pi\sigma^2\right) + \log \int \mathsf{E}\left\{ \exp\left[\left(y - \sqrt{\beta}\, \tilde{V}_0 \right)^2 \right] \right. \\ & \left. \times \prod_{a=1}^{u} \exp\left[-\frac{1}{2\sigma^2} \left(y - \sqrt{\beta}\, \tilde{V}_a \right)^2 \right] \middle| \boldsymbol{Q} \right\} \frac{\mathrm{d}y}{\sqrt{2\pi}}, \end{aligned} \tag{3.111}$$

in which $\tilde{\boldsymbol{V}}$ is a Gaussian random vector with covariance matrix $\boldsymbol{Q}$. By (3.107) and (3.110),

$$\begin{aligned} \frac{1}{K} \log \mathsf{E}\left\{ Z^u(\boldsymbol{Y}, \boldsymbol{S}) \right\} &= \frac{1}{K} \log \mathsf{E}\left\{ \exp\left[L \left(G^{(u)}(\boldsymbol{Q}) + \mathcal{O}\left(K^{-1}\right) \right) \right] \right\} && (3.112) \\ &= \frac{1}{K} \log \int \exp\left[K\, \beta^{-1}\, G^{(u)}(\boldsymbol{Q}) \right] \, \mathrm{d}\mu_K^{(u)}(\boldsymbol{Q}) + \mathcal{O}(K^{-1}) && (3.113) \end{aligned}$$

where the expectation over the replicated symbols is rewritten as an integral over the probability measure of the correlation matrix $\boldsymbol{Q}$, which is expressed as

$$\mu_K^{(u)}(\boldsymbol{Q}) = \mathsf{E}\left\{ \prod_{0 \le a \le b}^{u} \delta\left(\frac{1}{K} \sum_{k=1}^{K} \mathsf{snr}_k X_{ak} X_{bk} - Q_{ab} \right) \right\} \tag{3.114}$$

where $\delta(\cdot)$ is the Dirac function. By Cramér's theorem, the probability measure of the empirical means Q_{ab} defined by (3.109) satisfies, as $K \to \infty$, the large deviations property with some rate function $I^{(u)}(\boldsymbol{Q})$ [25]. Note the factor K in the exponent in the integral in (3.113). As $K \to \infty$, the integral is dominated by the maximum of the overall effect of the exponent and the rate of the measure on which the integral takes place. Precisely, by Varadhan's theorem [25],

$$\lim_{K\to\infty} \frac{1}{K} \log \mathsf{E}\left\{Z^u(\boldsymbol{Y}, \boldsymbol{S})\right\} = \sup_{\boldsymbol{Q}} [\beta^{-1} G^{(u)}(\boldsymbol{Q}) - I^{(u)}(\boldsymbol{Q})] \tag{3.115}$$

where the supremum is over all possible $\boldsymbol{Q}$ that can be obtained from varying X_{ak} in (3.109).

Let the moment generating function be defined as

$$M^{(u)}(\tilde{\boldsymbol{Q}}) = \mathsf{E}\left\{\exp\left[\mathsf{snr}\boldsymbol{X}^\top \tilde{\boldsymbol{Q}} \boldsymbol{X}\right]\right\} \tag{3.116}$$

where $\tilde{\boldsymbol{Q}}$ is a $(u+1) \times (u+1)$ symmetric matrix and the expectation in (3.116) is taken over independent random variables $\mathsf{snr} \sim P_{\mathsf{snr}}$, $X_0 \sim p_X$ and $X_1, \dots, X_u \sim q_X$. The rate of the measure $\mu_K^{(u)}$ is given by the Legendre-Fenchel transform of the cumulant generating function (logarithm of the moment generating function) [25]:

$$\begin{aligned} I^{(u)}(\boldsymbol{Q}) &= \sup_{\tilde{\boldsymbol{Q}}} I^{(u)}(\boldsymbol{Q}, \tilde{\boldsymbol{Q}}) && (3.117) \\ &= \sup_{\tilde{\boldsymbol{Q}}} \left[\mathsf{tr}\left\{\tilde{\boldsymbol{Q}}\boldsymbol{Q}\right\} - \log M^{(u)}(\tilde{\boldsymbol{Q}})\right] && (3.118) \end{aligned}$$

where the supremum is taken with respect to the symmetric matrix $\tilde{\boldsymbol{Q}}$.

In Appendix A.5, (3.111) is evaluated to obtain

$$G^{(u)}(\boldsymbol{Q}) = -\frac{1}{2} \log\det(\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{Q}) - \frac{1}{2}\log\left(1 + \frac{u}{\sigma^2}\right) - \frac{u}{2}\log\left(2\pi\sigma^2\right) \tag{3.119}$$

where $\boldsymbol{\Sigma}$ is a $(u+1) \times (u+1)$ matrix:[5]

$$\boldsymbol{\Sigma} = \frac{\beta}{\sigma^2(\sigma^2+u)} \left[\begin{array}{c:ccccc} u\sigma^2 & -\sigma^2 & -\sigma^2 & \dots & -\sigma^2 \\ \hdashline -\sigma^2 & \sigma^2+u-1 & -1 & \dots & -1 \\ -\sigma^2 & -1 & \sigma^2+u-1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -1 \\ -\sigma^2 & -1 & \dots & -1 & \sigma^2+u-1 \end{array}\right]. \tag{3.120}$$

Note that the lower right $u \times u$ block of the matrix in (3.120) is $(\sigma^2+u)\boldsymbol{I}_u - \boldsymbol{E}_u$ where $\boldsymbol{I}_u$ denotes the $u \times u$ identity matrix and $\boldsymbol{E}_u$ denotes a $u \times u$ matrix whose entries are all 1. It is clear that $\boldsymbol{\Sigma}$ is invariant if two nonzero indexes are interchanged, i.e., $\boldsymbol{\Sigma}$ is symmetric in the replicas.

[5] For convenience, the index number of all $(u+1) \times (u+1)$ matrices in this chapter starts from 0.

By (3.115)–(3.119), one has

$$\begin{aligned}
&\lim_{K\to\infty} \frac{1}{K} \log \mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} \\
&= \sup_{\boldsymbol{Q}} \left\{ \beta^{-1}\, G^{(u)}(\boldsymbol{Q}) - \sup_{\tilde{\boldsymbol{Q}}} \left[\mathsf{tr}\left\{\tilde{\boldsymbol{Q}}\boldsymbol{Q}\right\} - \log M^{(u)}(\tilde{\boldsymbol{Q}}) \right] \right\} && (3.121)\\
&= \sup_{\boldsymbol{Q}} \inf_{\tilde{\boldsymbol{Q}}} T^{(u)}(\boldsymbol{Q},\tilde{\boldsymbol{Q}}) && (3.122)
\end{aligned}$$

where

$$\begin{aligned}
T^{(u)}(\boldsymbol{Q},\tilde{\boldsymbol{Q}}) = &-\frac{1}{2\beta}\log\det(\boldsymbol{I}+\boldsymbol{\Sigma}\boldsymbol{Q}) - \mathsf{tr}\left\{\tilde{\boldsymbol{Q}}\boldsymbol{Q}\right\} + \log \mathsf{E}\left\{\exp\left[\mathsf{snr}\boldsymbol{X}^\top\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\} \\
&-\frac{1}{2\beta}\log\left(1+\frac{u}{\sigma^2}\right) - \frac{u}{2\beta}\log\left(2\pi\sigma^2\right).
\end{aligned} \tag{3.123}$$

For an arbitrary $\boldsymbol{Q}$, we first seek the point of zero gradient with respect to $\tilde{\boldsymbol{Q}}$ and find that for a given $\boldsymbol{Q}$, the extremum in $\tilde{\boldsymbol{Q}}$ satisfies

$$\boldsymbol{Q} = \frac{\mathsf{E}\left\{\mathsf{snr}\boldsymbol{X}\boldsymbol{X}^\top \exp\left[\mathsf{snr}\boldsymbol{X}^\top\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}{\mathsf{E}\left\{\exp\left[\mathsf{snr}\boldsymbol{X}^\top\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}. \tag{3.124}$$

Let $\tilde{\boldsymbol{Q}}^*(\boldsymbol{Q})$ be a solution to (3.124), which is a function of $\boldsymbol{Q}$. Assuming that $\tilde{\boldsymbol{Q}}^*(\boldsymbol{Q})$ is sufficiently smooth, we then seek the point of zero gradient of $T^{(u)}\left(\boldsymbol{Q},\tilde{\boldsymbol{Q}}^*(\boldsymbol{Q})\right)$ with respect to $\boldsymbol{Q}$.[6] By virtue of the relationship (3.124), one finds that the derivative of $\tilde{\boldsymbol{Q}}^*$ with respect to $\boldsymbol{Q}$ is inconsequential, and the extremum in $\boldsymbol{Q}$ satisfies

$$\tilde{\boldsymbol{Q}} = -\frac{1}{\beta}\left(\boldsymbol{\Sigma}^{-1}+\boldsymbol{Q}\right)^{-1}. \tag{3.125}$$

It is interesting to note from the resulting joint equations (3.124)–(3.125) that the order in which the supremum and infimum are taken in (3.122) can be exchanged without harm. The solution $\left(\boldsymbol{Q}^*,\tilde{\boldsymbol{Q}}^*\right)$ is in fact a saddle point of $T^{(u)}$. Notice that (3.124) can also be expressed as

$$\boldsymbol{Q} = \mathsf{E}\left\{\mathsf{snr}\boldsymbol{X}\boldsymbol{X}^\top \,\middle|\, \tilde{\boldsymbol{Q}}\right\} \tag{3.126}$$

where the expectation is over an appropriately defined measure $p_{\boldsymbol{X},\mathsf{snr}|\tilde{\boldsymbol{Q}}}$ dependent on $\tilde{\boldsymbol{Q}}$.

Solving joint equations (3.124) and (3.125) directly is prohibitive except in the simplest cases such as q_X being Gaussian. In the general case, because of the symmetry in the matrix $\boldsymbol{\Sigma}$ (3.120), we postulate that the solution to the joint equations satisfies *replica symmetry*, namely, both $\boldsymbol{Q}^*$ and $\tilde{\boldsymbol{Q}}^*$ are invariant if two nonzero replica indexes are interchanged. In

[6]The formula in footnote 2 on page 10 and the following identity is useful:

$$\frac{\partial\boldsymbol{Q}^{-1}}{\partial x} = -\boldsymbol{Q}^{-1}\frac{\partial\boldsymbol{Q}}{\partial x}\boldsymbol{Q}^{-1}.$$

other words, the extremum can be written as

$$\boldsymbol{Q}^* = \begin{bmatrix} r & m & m & \dots & m \\ m & p & q & \dots & q \\ m & q & p & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & q \\ m & q & \dots & q & p \end{bmatrix}, \qquad \tilde{\boldsymbol{Q}}^* = \begin{bmatrix} c & d & d & \dots & d \\ d & g & f & \dots & f \\ d & f & g & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & f \\ d & f & \dots & f & g \end{bmatrix} \tag{3.127}$$

where r, m, p, q, c, d, f, g are some real numbers.

It can be shown that replica symmetry holds in case that the postulated prior q_X is Gaussian. Under equal-power binary input and individually optimal detection, Tanaka showed also the stability of the replica-symmetric solution against replica-symmetry-breaking (RSB) if it is a "stable" solution (i.e., when the parameters satisfies certain condition) [96]. Thus, the replica-symmetric solution is at least a local maximum in such cases. In other cases, stability of replica symmetry can be broken [51]. Unfortunately, there is no known general condition for replica symmetry to hold. In this work we assume replica symmetry to hold and limit ourselves to replica-symmetric solution. We believe that in case replica symmetry is not a valid assumption, such solutions are a good approximation to the actual one. A justification of the replica symmetry assumption is relegated to future work.

Under replica symmetry, (3.119) is evaluated in Appendix A.5 to obtain

$$\begin{aligned} G^{(u)}\left(\boldsymbol{Q}^*\right) = & -\frac{u}{2}\log\left(2\pi\sigma^2\right) - \frac{u-1}{2}\log\left[1 + \frac{\beta}{\sigma^2}(p-q)\right] \\ & -\frac{1}{2}\log\left[1 + \frac{\beta}{\sigma^2}(p-q) + \frac{u}{\sigma^2}(1+\beta(r-2m+q))\right]. \end{aligned} \tag{3.128}$$

The moment generating function (3.116) is evaluated as

$$\begin{aligned} & M^{(u)}(\tilde{\boldsymbol{Q}}^*) \\ = \; & \mathsf{E}\left\{\exp\left[\mathsf{snr}\left(2d\sum_{a=1}^{u} X_0 X_a + 2f\sum_{0<a<b}^{u} X_a X_b + cX_0^2 + g\sum_{a=1}^{u} X_a^2\right)\right]\right\} \end{aligned} \tag{3.129}$$

$$\begin{aligned} = \; & \mathsf{E}\left\{\exp\left[\mathsf{snr}\left(\frac{d}{\sqrt{f}}X_0 + \sqrt{f}\sum_{a=1}^{u} X_a\right)^2\right.\right. \\ & \left.\left. + \left(c - \frac{d^2}{f}\right)\mathsf{snr}X_0^2 + (g-f)\mathsf{snr}\sum_{a=1}^{u} X_a^2\right]\right\}, \end{aligned} \tag{3.130}$$

where X_0 takes the actual input distribution p_X while X takes the postulated input distribution q_X. The expectation (3.130) with respect to $X_0, \dots, X_u$ can be decoupled using a variant of the Hubbard-Stratonovich transform [47]:

$$e^{x^2} = \sqrt{\frac{\eta}{2\pi}}\int \exp\left[-\frac{\eta}{2}z^2 + \sqrt{2\eta}\,xz\right]\,\mathrm{d}z, \quad \forall x, \eta. \tag{3.131}$$

Using (3.131) with $\eta = 2d^2/f$,

$$\begin{aligned} M^{(u)}(\tilde{\boldsymbol{Q}}^*) = \mathsf{E}\Bigg\{ & \sqrt{\frac{d^2}{f\pi}}\int \exp\left[-\frac{d^2}{f}z^2 + 2\sqrt{\mathsf{snr}}\left(\frac{d^2}{f}X_0 + d\sum_{a=1}^{u} X_a\right)z\right. \\ & \left. + \left(c - \frac{d^2}{f}\right)\mathsf{snr}X_0^2 + (g-f)\mathsf{snr}\sum_{a=1}^{u} X_a^2\right]\,\mathrm{d}z\Bigg\}. \end{aligned} \tag{3.132}$$

Since $X_0, \dots, X_u$ and snr are independent, the rate of the measure (3.118) under replica symmetry is obtained as

$$\begin{aligned} I^{(u)}\left(\boldsymbol{Q}^*\right) =& rc + upg + 2umd + u(u-1)qf \\ & - \log \mathsf{E}\left\{ \int \sqrt{\frac{d^2}{f\pi}} \mathsf{E}\left\{ \exp\left[-\frac{d^2}{f}\left(z - \sqrt{\mathsf{snr}}X_0\right)^2 + c\,\mathsf{snr}X_0^2 \right] \,\middle|\, \mathsf{snr} \right\} \right. \\ & \left. \times \left[\mathsf{E}_q\left\{ \exp\left[2d\sqrt{\mathsf{snr}}Xz + (g-f)\mathsf{snr}X^2 \right] \,\middle|\, \mathsf{snr} \right\} \right]^u \mathrm{d}z \right\} \end{aligned} \tag{3.133}$$

by also noticing that

$$\mathsf{tr}\left\{ \tilde{\boldsymbol{Q}}^* \boldsymbol{Q}^* \right\} = rc + upg + 2umd + u(u-1)qf. \tag{3.134}$$

The free energy is then found by (3.93) and (3.115):

$$\mathcal{F} = -\lim_{u\to 0} \frac{\partial}{\partial u} \left[\beta^{-1} G^{(u)}\left(\boldsymbol{Q}^*\right) - I^{(u)}\left(\boldsymbol{Q}^*\right) \right]. \tag{3.135}$$

The eight parameters (r, m, p, q, c, d, f, g) are the solution to the joint equations (3.124)–(3.125) under replica symmetry that minimizes $\mathcal{F}$. It is interesting to note that as functions of u, the derivative of each of the eight parameters with respect to u vanishes as $u \to 0$. Thus for the purpose of the free energy (3.135), it suffices to find the extremum of $\left[\beta^{-1}G^{(u)} - I^{(u)}\right]$ at $u = 0$. Using (3.125), it can be shown that at $u = 0$,

$$c = 0, \tag{3.136a}$$

$$d = \frac{1}{2[\sigma^2 + \beta(p-q)]}, \tag{3.136b}$$

$$f = \frac{1 + \beta(r - 2m + q)}{2[\sigma^2 + \beta(p-q)]^2}, \tag{3.136c}$$

$$g = f - d. \tag{3.136d}$$

The parameters r, m, p, q can be determined from (3.126) by studying the measure $p_{\boldsymbol{X},\mathsf{snr}|\tilde{\boldsymbol{Q}}}$ under replica symmetry and $u \to 0$. For that purpose, define two useful parameters:

$$\eta = \frac{2d^2}{f}, \tag{3.137a}$$

$$\xi = 2d. \tag{3.137b}$$

Noticing that $c = 0$, $g - f = -d$, (3.132) can be written as

$$\begin{aligned} M^{(u)}(\tilde{\boldsymbol{Q}}^*) =& \mathsf{E}\left\{ \sqrt{\frac{\eta}{2\pi}} \int \exp\left[-\frac{\eta}{2}\left(z - \sqrt{\mathsf{snr}}X_0\right)^2 \right] \right. \\ & \left. \times \left[\mathsf{E}_q\left\{ \exp\left[-\frac{\xi}{2}z^2 - \frac{\xi}{2}\left(z - \sqrt{\mathsf{snr}}X\right)^2 \right] \,\middle|\, \mathsf{snr} \right\} \right]^u \mathrm{d}z \right\}. \end{aligned} \tag{3.138}$$

It is clear that

$$\lim_{u\to 0} M^{(u)}(\tilde{\boldsymbol{Q}}^*) = 1. \tag{3.139}$$

Hence by (3.124), as $u \to 0$,

$$Q^*_{ab} = \mathsf{E}\left\{ \mathsf{snr}X_a X_b \mid \tilde{\boldsymbol{Q}}^* \right\} \to \mathsf{E}\left\{ \mathsf{snr}X_a X_b \exp\left[\boldsymbol{X}^\top \tilde{\boldsymbol{Q}}^* \boldsymbol{X} \right] \right\}. \tag{3.140}$$

We now give a useful representation for the parameters r, m, p, q defined in (3.127). Note that as $u \to 0$,

$$\mathsf{E}\left\{\mathsf{snr} X_0 X_1 \exp\left[\boldsymbol{X}^\top \tilde{\boldsymbol{Q}}^* \boldsymbol{X}\right]\right\} = \mathsf{E}\left\{\mathsf{snr} X_0 \int \sqrt{\frac{\eta}{2\pi}} \exp\left[-\frac{\eta}{2}\left(z - \sqrt{\mathsf{snr}} X_0\right)^2\right] \right. \\ \left. \times \frac{X_1 \sqrt{\frac{\xi}{2\pi}} \exp\left[-\frac{\xi}{2}\left(z - \sqrt{\mathsf{snr}} X_1\right)^2\right]}{\mathsf{E}_q\left\{\sqrt{\frac{\xi}{2\pi}} \exp\left[-\frac{\xi}{2}\left(z - \sqrt{\mathsf{snr}} X_1\right)^2\right] \,\middle|\, \mathsf{snr}\right\}} \, \mathrm{d}z \right\}. \tag{3.141}$$

Let two scalar Gaussian channels be defined as in Section 3.2.4, i.e., $p_{Z|X,\mathsf{snr};\eta}$ is given by (3.19) and similarly,

$$q_{Z|X,\mathsf{snr};\xi}(z|x, \mathsf{snr}; \xi) = \sqrt{\frac{\xi}{2\pi}} \exp\left[-\frac{\xi}{2}\left(z - \sqrt{\mathsf{snr}} x\right)^2\right]. \tag{3.142}$$

Assuming that the input distribution to the channel $q_{Z|X,\mathsf{snr};\xi}$ is q_X, a posterior probability distribution $q_{X|Z,\mathsf{snr};\xi}$ is induced. The retrochannel is then characterized by $q_{X|Z,\mathsf{snr};\xi}$. The posterior mean with respect to the measure q is therefore given by (3.37). The scalar Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$, the retrochannel $q_{X|Z,\mathsf{snr};\xi}$ and the PME are depicted in Figure 3.6. Then, (3.141) can be simplified to obtain

$$\begin{aligned} Q_{01}^* &= \mathsf{E}\left\{\mathsf{snr} X_0 X_1 \exp\left[\boldsymbol{X}^\top \tilde{\boldsymbol{Q}}^* \boldsymbol{X}\right]\right\} && (3.143) \\ &= \mathsf{E}\left\{\mathsf{snr}\, X_0 \left\langle X \right\rangle_q\right\} && (3.144) \end{aligned}$$

where $\langle X \rangle_q$, defined in (3.37), is the generalized posterior mean estimate under the posterior probability distribution $q_{X|Z,\mathsf{snr};\xi}$. Similarly, (3.140) can be evaluated for all indexes (a, b) yielding together with (3.127):

$$\begin{aligned} r &= Q_{00}^* = \mathsf{E}\left\{\mathsf{snr}\, X_0^2\right\} = \mathsf{E}\left\{\mathsf{snr}\right\}, && (3.145\text{a}) \\ m &= Q_{01}^* = \mathsf{E}\left\{\mathsf{snr}\, X_0 \left\langle X \right\rangle_q\right\}, && (3.145\text{b}) \\ p &= Q_{11}^* = \mathsf{E}\left\{\mathsf{snr}\, \left\langle X^2 \right\rangle_q\right\}, && (3.145\text{c}) \\ q &= Q_{12}^* = \mathsf{E}\left\{\mathsf{snr}\, (\left\langle X \right\rangle_q)^2\right\}. && (3.145\text{d}) \end{aligned}$$

In summary, under replica symmetry, the parameters c, d, f, g are given by (3.136) as functions of r, m, p, q, which are in turn determined by the statistics of the two channels (3.19) and (3.142) parameterized by $\eta = 2d^2/f$ and $\xi = 2d$. It is not difficult to see that

$$\begin{aligned} r - 2m + q &= \mathsf{E}\left\{\mathsf{snr}\left(X_0 - \left\langle X \right\rangle_q\right)^2\right\}, && (3.146\text{a}) \\ p - q &= \mathsf{E}\left\{\mathsf{snr}\left(X - \left\langle X \right\rangle_q\right)^2\right\}. && (3.146\text{b}) \end{aligned}$$

Using (3.136) and (3.137), it can be checked that

$$\begin{aligned} p - q &= \frac{1}{\beta}\left(\frac{1}{\xi} - \sigma^2\right), && (3.147\text{a}) \\ r - 2m + q &= \frac{1}{\beta}\left(\frac{1}{\eta} - 1\right). && (3.147\text{b}) \end{aligned}$$

Thus $G^{(u)}$ and $I^{(u)}$ given by (3.128) and (3.133) can be expressed in η and ξ. Using (3.135), the free energy is found as (3.42), where (η, ξ) satisfies fixed-point equations

$$\eta^{-1} = 1 + \beta \, \mathsf{E}\left\{\mathsf{snr}\left(X_0 - \langle X \rangle_q\right)^2\right\}, \tag{3.148a}$$

$$\xi^{-1} = \sigma^2 + \beta \, \mathsf{E}\left\{\mathsf{snr}\left(X - \langle X \rangle_q\right)^2\right\}, \tag{3.148b}$$

where $X_0 \sim p_X$ is the input to the scalar Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$, X is the output of the retrochannel $q_{X|Z,\mathsf{snr};\xi}$ induced by a postulated channel $q_{Z|X,\mathsf{snr};\xi}$ with input distribution q_X, and $\langle X \rangle_q$ is the generalized posterior mean estimate according to $q_{X|Z,\mathsf{snr};\xi}$. In case of multiple solutions to (3.148), (η, ξ) is chosen as the solution that gives the minimum free energy $\mathcal{F}$. By defining $\mathsf{mse}(\mathsf{snr}; \eta, \xi)$ and $\mathsf{var}(\mathsf{snr}; \eta, \xi)$ as (3.38) and (3.39), the coupled equations (3.136) and (3.145) can be summarized to establish the key coupled fixed-point equations (3.41).

From (3.19), it is clear that η is the inverse noise variance of the scalar Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$. In Section 3.4.2, we will show that the single-user channel seen at the output of the multiuser detector front end is equivalent to the scalar Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$ concatenated with a generalized PME decision function. Therefore, the parameter η is the degradation factor in the effective SNR due to the MAI, which is termed the multiuser efficiency. More on this later.

In the special case where the postulated measure q is identical to the actual measure p (i.e., $q_X = p_X$ and $\sigma = 1$), we choose $\eta = \xi$ as argued in Section 3.2.5. Using (3.88), the total spectral efficiency is

$$\begin{aligned} \mathsf{C}_{\text{joint}} = &- \beta \, \mathsf{E}\left\{\int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta) \, \log p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta) \, \mathrm{d}\, z\right\} \\ &- \frac{\beta}{2} \log \frac{2\pi e}{\eta} + \frac{1}{2}(\eta - 1 - \log \eta), \end{aligned} \tag{3.149}$$

where η satisfies

$$\eta + \eta \beta \, \mathsf{E}\left\{\mathsf{snr}\left[1 - \int \frac{[p_1(z, \mathsf{snr}; \eta)]^2}{p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta)} \, \mathrm{d}\, z\right]\right\} = 1. \tag{3.150}$$

Equation (3.30) in Claim 3.2 is thus established.

3.4.2 Joint Moments

Let us study the distribution of the generalized PME output

$$\langle X_k \rangle_q = \mathsf{E}_q\left\{X_k \mid \boldsymbol{Y}, \boldsymbol{S}\right\} \tag{3.151}$$

conditioned on X_k being transmitted with signal-to-noise ratio snr_k. For this purpose we calculate the joint moments of X_k and $\langle X_k \rangle_q$ using the replica method. It turns out that $\langle X_k \rangle_q$ is as if X_k were corrupted by Gaussian noise and then mapped by a decision function.

Consider the CDMA channel, the generalized PME and a retrochannel as depicted in Figure 3.9. Let $\underline{\boldsymbol{X}}_a = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_u]$.

Lemma 3.1 *Let $\boldsymbol{X}_0 \to (\boldsymbol{Y}, \boldsymbol{S}) \to \underline{\boldsymbol{X}}_a$ be the Markov chain as depicted in Figure 3.9. Then for every k,*

$$\mathsf{E}\left\{X_{0k}^j \, \langle X_k \rangle_q^i\right\} = \mathsf{E}\left\{X_{0k}^j \prod_{m=1}^{i} X_{mk}\right\}. \tag{3.152}$$

Proof: Since the replicas are i.i.d., $\langle X_{ak} \rangle_q = \langle X_k \rangle_q$ for all $a = 1, \ldots, u$. Hence,

$$\begin{aligned}
\mathsf{E}\left\{X_{0k}^j \, \langle X_k \rangle_q^i\right\} &= \mathsf{E}\left\{X_{0k}^j \prod_{m=1}^{i} \langle X_{mk} \rangle_q\right\} && (3.153)\\
&= \mathsf{E}\left\{X_{0k}^j \prod_{m=1}^{i} \mathsf{E}_q\left\{X_{mk} \mid \boldsymbol{Y}, \boldsymbol{S}\right\}\right\} && (3.154)\\
&= \mathsf{E}\left\{X_{0k}^j \, \mathsf{E}_q\left\{\prod_{m=1}^{i} X_{mk} \,\middle|\, \boldsymbol{Y}, \boldsymbol{S}\right\}\right\} && (3.155)\\
&= \mathsf{E}\left\{X_{0k}^j \prod_{m=1}^{i} X_{mk}\right\}, && (3.156)
\end{aligned}$$

where (3.155)–(3.156) are by conditional independence of X_{ak} due to Markov property. ■

The joint moments of $\langle X_k \rangle$ and X_{0k} have now been converted into expectations under the replicated system. The following lemma allows us to determine the expected value of a function of the replicated symbols by considering a modified partition function akin to (3.95).

Lemma 3.2 *Given an arbitrary function $f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)$, define*

$$Z^{(u)}(\boldsymbol{y}, \boldsymbol{S}, \boldsymbol{x}_0; h) = \mathsf{E}_q\left\{\exp\left[h\, f(\boldsymbol{x}_0, \underline{\boldsymbol{X}}_a)\right] \prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_a, \boldsymbol{S}) \,\middle|\, \boldsymbol{S}\right\}. \tag{3.157}$$

If $\mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a) \mid \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0\right\}$ is not dependent on u, then

$$\mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)\right\} = \lim_{u \to 0} \frac{\partial}{\partial h} \log \mathsf{E}\left\{Z^{(u)}(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0; h)\right\}\Big|_{h=0}. \tag{3.158}$$

Proof: See Appendix A.6. ■

For the function $f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)$ to have influence on the free energy, it must grow at least linearly with K. Assume that $f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)$ involves user 1 through $K_1 = \alpha_1 K$ where $0 < \alpha_1 < 1$ is fixed as $K \to \infty$:

$$f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a) = \sum_{k=1}^{K_1} X_{0k}^j \prod_{m=1}^{i} X_{mk}. \tag{3.159}$$

It is also assumed that user 1 through K_1 take the same signal-to-noise ratio snr_1. We will finally take the limit $\alpha_1 \to 0$ so that the constraint of equal SNR from the first K_1 users is superfluous. By Lemma 3.1, the moments can be written as

$$\frac{1}{K_1} \sum_{k=1}^{K_1} \mathsf{E}\left\{X_{0k}^j \, \langle X_k \rangle^i\right\} = \frac{1}{K_1} \mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)\right\}. \tag{3.160}$$

Note that

$$\mathsf{E}\left\{ f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a) \mid \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0 \right\} = \mathsf{E}\left\{ \sum_{k=1}^{K_1} X_{0k}^j \, X_k^i \,\middle|\, \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0 \right\} \tag{3.161}$$

is not dependent on u. By Lemma 3.2, the moment (3.160) can be obtained as

$$\lim_{u \to 0} \frac{\partial}{\partial h} \frac{1}{\alpha_1 K} \log \mathsf{E}\left\{ Z^{(u)}(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0; h) \right\}\Big|_{h=0} \tag{3.162}$$

where

$$\begin{aligned} Z^{(u)}(\boldsymbol{y}, \boldsymbol{S}, \boldsymbol{x}_0; h) =& \mathsf{E}_q \left\{ \exp\left[h \sum_{k=1}^{K_1} x_{0k}^j \prod_{m=1}^{i} X_{mk} \right] \right. \\ & \left. \times \ \left(2\pi\sigma^2\right)^{-\frac{uL}{2}} \prod_{a=1}^{u} \exp\left[-\frac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{X}_a\|^2 \right] \middle| \, \boldsymbol{S} \right\}. \end{aligned} \tag{3.163}$$

Consider (3.163) as a partition function and the same techniques in Section 3.4.1 can be used to write

$$\lim_{K \to \infty} \frac{1}{K} \log \mathsf{E}\left\{ Z^{(u)}(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0; h) \right\} = \sup_{\boldsymbol{Q}} \left\{ \beta^{-1} G^{(u)}(\boldsymbol{Q}) - I^{(u)}(\boldsymbol{Q}; h) \right\} \tag{3.164}$$

where $G^{(u)}(\boldsymbol{Q})$ is given by (3.108) and $I^{(u)}(\boldsymbol{Q}; h)$ is the rate of the following measure

$$\mu_K^{(u)}(\boldsymbol{Q}; h) = \mathsf{E}\left\{ \exp\left[h \sum_{k=1}^{K_1} X_{0k}^j \prod_{m=1}^{i} X_{mk} \right] \prod_{0 \le a \le b}^{u} \delta\left(\sum_{k=1}^{K} \mathsf{snr}_k X_{ak} X_{bk} - K Q_{ab} \right) \right\}. \tag{3.165}$$

By the large deviations property, one finds the rate

$$\begin{aligned} I^{(u)}(\boldsymbol{Q}; h) =& \sup_{\tilde{\boldsymbol{Q}}} \left[\mathsf{tr}\left\{ \tilde{\boldsymbol{Q}} \boldsymbol{Q} \right\} - \log M^{(u)}(\tilde{\boldsymbol{Q}}) \right. \\ & \left. - \alpha_1 \left(\log M^{(u)}(\tilde{\boldsymbol{Q}}, \mathsf{snr}_1; h) - \log M^{(u)}(\tilde{\boldsymbol{Q}}, \mathsf{snr}_1; 0) \right) \right] \end{aligned} \tag{3.166}$$

where $M^{(u)}(\tilde{\boldsymbol{Q}})$ is defined in (3.116), and

$$M^{(u)}(\tilde{\boldsymbol{Q}}, \mathsf{snr}; h) = \mathsf{E}\left\{ \exp\left[h\, X_0^j \prod_{m=1}^{i} X_m \right] \exp\left[\mathsf{snr} \boldsymbol{X}^\top \tilde{\boldsymbol{Q}} \boldsymbol{X} \right] \,\middle|\, \mathsf{snr} \right\}. \tag{3.167}$$

From (3.164) and (3.166), taking the derivative in (3.162) with respect to h at $h = 0$ leaves us only one term

$$\frac{\partial}{\partial h} \log M^{(u)}(\tilde{\boldsymbol{Q}}, \mathsf{snr}_1; h)\Big|_{h=0} = \frac{\mathsf{E}\left\{ X_0^j \prod_{m=1}^{i} X_m \exp\left[\mathsf{snr}_1 \boldsymbol{X}^\top \tilde{\boldsymbol{Q}} \boldsymbol{X} \right] \right\}}{\mathsf{E}\left\{ \exp\left[\mathsf{snr}_1 \boldsymbol{X}^\top \tilde{\boldsymbol{Q}} \boldsymbol{X} \right] \right\}}. \tag{3.168}$$

Since

$$Z^{(u)}(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0; h)\Big|_{h=0} = Z^u(\boldsymbol{Y}, \boldsymbol{S}), \tag{3.169}$$

the $\tilde{\boldsymbol{Q}}$ in (3.168) that give the supremum in (3.166) at $h \to 0$ is exactly the $\tilde{\boldsymbol{Q}}$ that gives the supremum of (3.118). Hence replica symmetry is straightforward. By introducing the parameters (η, ξ) the same as in Section 3.4.1, (3.168) can be further evaluated as

$$\frac{\int \left(\sqrt{\frac{2\pi}{\xi}} e^{\frac{\xi z^2}{2}}\right)^u p_j(z, \mathsf{snr}_1; \eta)\, q_0^{u-i}(z, \mathsf{snr}_1; \xi)\, q_1^i(Z, \mathsf{snr}_1; \xi)\, \mathrm{d}z}{\int \left(\sqrt{\frac{2\pi}{\xi}} e^{\frac{\xi z^2}{2}}\right)^u p_0(z, \mathsf{snr}_1; \eta) q_0^u(Z, \mathsf{snr}_1; \xi)\, \mathrm{d}z} \tag{3.170}$$

where we have used

$$\mathsf{E}\left\{ X_0^i \exp\left[\eta\sqrt{\mathsf{snr}}\, X_0 z - \frac{1}{2}\eta \mathsf{snr} X_0^2\right] \middle|\, \mathsf{snr}\right\} = \sqrt{\frac{2\pi}{\eta}}\, e^{\frac{\eta z^2}{2}}\, p_i(z, \mathsf{snr}; \eta), \tag{3.171a}$$

$$\mathsf{E}\left\{ X^i \exp\left[\xi\sqrt{\mathsf{snr}}\, X z - \frac{1}{2}\xi \mathsf{snr} X^2\right] \middle|\, \mathsf{snr}\right\} = \sqrt{\frac{2\pi}{\xi}}\, e^{\frac{\xi z^2}{2}}\, q_i(z, \xi; \eta). \tag{3.171b}$$

Taking the limit $u \to 0$ in (3.170), one has

$$\lim_{K\to\infty} \frac{1}{K_1} \sum_{k=1}^{K_1} \mathsf{E}\left\{X_{0k}^j\, \langle X_k\rangle_q^i\right\} = \int p_j(Z, \mathsf{snr}_1; \eta) \left[\frac{q_1(Z, \mathsf{snr}_1; \xi)}{q_0(Z, \mathsf{snr}_1; \xi)}\right]^i \mathrm{d}z. \tag{3.172}$$

Letting $K_1 \to 1$ (thus $\alpha_1 \to 0$) so that the requirement that the first K_1 users take the same SNR becomes unnecessary, we have proved the following by noticing that the above arguments hold for every user.

Proposition 3.2 *Let the SNR of user k be $\mathsf{snr}_k = \mathsf{snr}$, then for all integers $i, j \geq 0$,*

$$\lim_{K\to\infty} \mathsf{E}\left\{X_{0k}^j\, \langle X_k\rangle^i\right\} = \int p_j(z, \mathsf{snr}; \eta) \left[\frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)}\right]^i \mathrm{d}z. \tag{3.173}$$

where η is the solution to (3.150).

It is easy to see from (3.173) and (3.23) that

$$\begin{aligned}
&\lim_{K\to\infty} \mathsf{E}\left\{(\langle X_k\rangle - X_{0k})^i\right\} \\
&= \sum_{j=0}^{i} \binom{i}{j} \int p_j(z, \mathsf{snr}; \eta) \left[\frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)}\right]^{i-j} \mathrm{d}z && (3.174)\\
&= \int \sum_{j=0}^{i} \binom{i}{j} \mathsf{E}\left\{X^i p_{Z|X,\mathsf{snr};\eta}(z \,|\, X, \mathsf{snr}; \eta) \,\middle|\, \mathsf{snr}\right\} \left[\frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)}\right]^{i-j} \mathrm{d}z && (3.175)\\
&= \int \mathsf{E}\left\{\left(\frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)} - X\right)^i p_{Z|X,\mathsf{snr};\eta}(z \,|\, X, \mathsf{snr}; \eta) \,\middle|\, \mathsf{snr}\right\} \mathrm{d}z && (3.176)\\
&= \mathsf{E}\left\{\left(\frac{q_1(Z, \mathsf{snr}; \xi)}{q_0(Z, \mathsf{snr}; \xi)} - X\right)^i \,\middle|\, \mathsf{snr}\right\} && (3.177)
\end{aligned}$$

Since (3.177) is true for all k, we can drop the user index in further discussion on separate decoding. Note that the moments are uniformly bounded. The distribution is thus uniquely determined by the moments due to the Carleman's Theorem [28, p. 227]. Thus, if we consider X as the input of a user with signal-to-noise ratio snr to the multiple-access channel and $\langle X \rangle$ as the output of the generalized PME, then the error $\langle X \rangle - X$ has identical distribution as

$$\frac{q_1(Z, \mathsf{snr}; \xi)}{q_0(Z, \mathsf{snr}; \xi)} - X \tag{3.178}$$

where (Z, X) has joint distribution $q_{Z,X|\mathsf{snr};\xi} = q_X\, q_{Z|X,\mathsf{snr};\xi}$. By (3.43),

$$\frac{q_1(\,\cdot\,, \mathsf{snr}; \xi)}{q_0(\,\cdot\,, \mathsf{snr}; \xi)} \tag{3.179}$$

is the decision function associated with the generalized posterior mean estimator. Clearly, the joint distribution of the input X to the multiple-access channel (3.4) and the output of the generalized PME $\langle X \rangle_q$ is identical to the joint distribution of the input to the single-user Gaussian channel (3.18) and the decision function (3.179) evaluated at the output Z of (3.18). Hence the decoupling principle.

The decision function can be ignored from both detection- and information-theoretic viewpoints due to its monotonicity:

Proposition 3.3 *The decision function (3.24) is strictly monotone increasing in z for all α.*

Proof: Let $(\cdot)'$ denote derivative with respect to z. One can show that

$$q_i'(z, \mathsf{snr}; \xi) = \xi\sqrt{\mathsf{snr}}\, q_{i+1}(z, \mathsf{snr}; \xi) - \xi z\, q_i(z, \mathsf{snr}; \xi) \quad i = 0, 1, \ldots \tag{3.180}$$

Clearly,

$$\left[\frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)}\right]' = \xi\sqrt{\mathsf{snr}}\, \frac{q_2(z, \mathsf{snr}; \xi) q_0(z, \mathsf{snr}; \xi) - [q_1(z, \mathsf{snr}; \xi)]^2}{[q_0(z, \mathsf{snr}; \xi)]^2}. \tag{3.181}$$

The numerator in (3.181) is positive by the Cauchy-Schwartz inequality. For the numerator in (3.181) to be 0, X must be a constant, which contradicts the assumption that X has zero mean and unit variance. Therefore, (3.24) is strictly increasing. ■

Collecting relevant results in the above, the equivalent single-user channel is then an additive Gaussian noise channel with input signal-to-noise ratio snr and noise variance η^{-1} as depicted in Figure 3.6. Claim 3.3 is thus proved. In the special case that the postulated measure q is identical to the actual measure p, Claim 3.3 reduces to Claim 3.1.

The single-user mutual information is now simply that of a Gaussian channel with input distribution p_X,

$$I(\eta\, \mathsf{snr}) = -\frac{1}{2} \log \frac{2\pi e}{\eta} - \int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta) \log p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta)\, \mathrm{d}z. \tag{3.182}$$

The overall spectral efficiency under separate decoding is therefore

$$\mathsf{C}_{\text{sep}} = \beta\, \mathsf{E}\left\{I(\eta\, \mathsf{snr})\right\}. \tag{3.183}$$

Hence the proof of (3.26) and (3.28). Claim 3.2 is proved by comparing (3.183) to (3.149).

3.5 Complex-valued Channels

Until now our discussion is based on a real-valued setting of the channel (3.4), namely, both the inputs X_k and the spreading chips S_{nk} take real values. In practice, particularly in carrier-modulated communications where spectral efficiency is a major concern, transmission in the complex domain must be addressed. Either the input symbols or the spreading chips or both can take values in the set of complex numbers. The channel model (3.4) is equivalent to the following real-valued channel:

$$\begin{bmatrix} \boldsymbol{Y}^{(\mathrm{r})} \\ \boldsymbol{Y}^{(\mathrm{i})} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}^{(\mathrm{r})} & -\boldsymbol{S}^{(\mathrm{i})} \\ \boldsymbol{S}^{(\mathrm{i})} & \boldsymbol{S}^{(\mathrm{r})} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}^{(\mathrm{r})} \\ \boldsymbol{X}^{(\mathrm{i})} \end{bmatrix} + \begin{bmatrix} \boldsymbol{N}^{(\mathrm{r})} \\ \boldsymbol{N}^{(\mathrm{i})} \end{bmatrix}, \tag{3.184}$$

where the superscripts (r) and (i) denote real and imaginary components respectively. Note that the previous analysis does not apply to (3.184) since the channel state matrix does not contain i.i.d. entries in this case.

If the inputs take complex values but the spreading is real-valued ($\boldsymbol{S}^{(\mathrm{i})} = 0$), the channel can be considered as two uses of the real-valued channel $\boldsymbol{S} = \boldsymbol{S}^{(\mathrm{r})}$, where the inputs $\boldsymbol{X}^{(\mathrm{r})}$ and $\boldsymbol{X}^{(\mathrm{i})}$ to the two channels may be dependent. Since independent inputs maximize the channel capacity, there is little reason to transmit dependent signals in the two subchannels. Thus the analysis of the real-valued channel in previous sections also applies to the case of independent in-phase and quadrature components, while the only change is that the spectral efficiency is the sum of that of the two subchannels.

If the spreading chips are complex-valued, the analysis in the previous sections can be modified to take this into account. For convenience it is assumed that the real and imaginary components of spreading chips, $S_{nk}^{(\mathrm{r})}$, $S_{nk}^{(\mathrm{i})}$ are i.i.d. with zero mean and unit variance. The noise vector has i.i.d. circularly symmetric Gaussian entries, i.e., $\mathsf{E}\{\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}}\} = 2\boldsymbol{I}$. Thus the conditional probability distribution of the true and the postulated CDMA channels are

$$p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = (2\pi)^{-L} \exp\left[-\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right], \tag{3.185}$$

and

$$q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = \left(2\pi\sigma^2\right)^{-L} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{3.186}$$

respectively. Also, the true and the postulated input distributions p_X and q_X have both zero-mean and unit variance, $\mathsf{E}\left\{|X|^2\right\} = \mathsf{E}_q\left\{|X|^2\right\} = 1$. Note that the in-phase and the quadrature components are intertwined due to complex spreading.

The replica analysis can be carried out similarly as in Section 3.4. In the following we highlight the major differences.

Given $(\boldsymbol{\Gamma}, \underline{\boldsymbol{X}})$, the variables defined in (3.106),

$$V_a = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sqrt{\mathsf{snr}_k}\, S_k X_{ak}, \quad a = 0, 1, \dots, u \tag{3.187}$$

have asymptotically independent real and imaginary components. Thus, $G_K^{(u)}$ can be evaluated to be 2 times that under real-valued channels with

$$Q_{ab} = \frac{1}{K} \sum_{k=1}^{K} \mathsf{snr}_k \mathrm{Re}\left\{X_{ak} X_{bk}^*\right\}, \quad a, b = 0, \dots, u. \tag{3.188}$$

The rate $I^{(u)}$ of the measure $\mu_K^{(u)}$ of $\boldsymbol{Q}$ is obtained as

$$I^{(u)}(\boldsymbol{Q}) = \sup_{\tilde{\boldsymbol{Q}}} \left[\mathsf{tr}\left\{ \tilde{\boldsymbol{Q}}\boldsymbol{Q} \right\} - \log \mathsf{E}\left\{ \exp\left[\mathsf{snr}\boldsymbol{X}^{\mathrm{H}}\tilde{\boldsymbol{Q}}\boldsymbol{X} \right] \right\} \right]. \tag{3.189}$$

As a result, the fixed-point joint equations for $\boldsymbol{Q}$ and $\tilde{\boldsymbol{Q}}$ are

$$\tilde{\boldsymbol{Q}} = -\frac{2}{\beta}\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{Q}\right)^{-1}, \tag{3.190a}$$

$$\boldsymbol{Q} = \frac{\mathsf{E}\left\{ \mathsf{snr}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\exp\left[\mathsf{snr}\boldsymbol{X}^{\mathrm{H}}\tilde{\boldsymbol{Q}}\boldsymbol{X} \right] \right\}}{\mathsf{E}\left\{ \exp\left[\mathsf{snr}\boldsymbol{X}^{\mathrm{H}}\tilde{\boldsymbol{Q}}\boldsymbol{X} \right] \right\}}. \tag{3.190b}$$

Under replica symmetry (3.127), the parameters (c, d, f, g) are found to be 2 times the values given in (3.136), and (r, m, p, q) are found the same as (3.145) except that all squares are replaced by squared norms. By defining two parameters (differs from (3.137) by a factor of 2):

$$\eta = \frac{d^2}{f}, \tag{3.191a}$$

$$\xi = d, \tag{3.191b}$$

we have the following result.

Claim 3.4 *Let the generalized multiuser posterior mean estimate of the complex-valued multiple-access channel (3.185) with complex-valued spreading be* $\langle \boldsymbol{X} \rangle_q$ *with a postulated input distribution* q_X *and noise level* σ. *Then, in the large-system limit, the distribution of the multiuser detection output* $\langle X_k \rangle_q$ *conditioned on* $X_k = x$ *being transmitted with signal-to-noise ratio* snr_k *is identical to the distribution of the generalized estimate* $\langle X \rangle_q$ *of a scalar complex Gaussian channel*

$$Z = \sqrt{\mathsf{snr}}\, X + \frac{1}{\sqrt{\eta}} N \tag{3.192}$$

conditioned on $X = x$ *being transmitted with input signal-to-noise ratio* $\mathsf{snr} = \mathsf{snr}_k$, *where* N *is circularly symmetric Gaussian with unit variance,* $\mathsf{E}\left\{|N|^2\right\} = 1$. *The multiuser efficiency* η *and the inverse noise variance* ξ *of the postulated scalar channel (3.186) satisfy the coupled equations:*

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{mse}(\mathsf{snr};\eta,\xi)\right\}, \tag{3.193a}$$

$$\xi^{-1} = \sigma^2 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathsf{var}(\mathsf{snr};\eta,\xi)\right\}, \tag{3.193b}$$

where the mean-square error of the generalized PME and the variance of the retrochannel are defined similarly as that of the real-valued channel, with the squares in (3.38) and (3.39) replaced by the squared norms. In case of multiple solutions to (3.41), (η, ξ) *are chosen to minimize the free energy:*

$$\begin{aligned}\mathcal{F} = &-\mathsf{E}\left\{ \int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta)\,\log q_{Z|\mathsf{snr};\xi}(z|\mathsf{snr};\xi)\,\mathrm{d}\,z \right\} \\ &+ \log\frac{\xi}{\pi} - \frac{\xi}{\eta} + \frac{\sigma^2\xi(\eta-\xi)}{\beta\eta} + \frac{1}{\beta}(\xi - 1 - \log\xi) + \frac{1}{\beta}\log(2\pi) + \frac{\xi}{\beta\eta}.\end{aligned} \tag{3.194}$$

Corollary 3.3 *For the complex-valued channel (3.185), the mutual information of the single-user channel seen at the generalized multiuser PME output for a user with signal-to-noise ratio* snr *is*

$$I(\eta\,\mathsf{snr}) = \mathsf{D}(p_{Z|X,\mathsf{snr};\eta}\,||\,p_{Z|\mathsf{snr};\eta}\,|\,p_X). \tag{3.195}$$

where η *is the multiuser efficiency given by Claim 3.4 and* $p_{Z|\mathsf{snr};\eta}$ *is the marginal probability distribution of the output of channel (3.192). The overall spectral efficiency under suboptimal separate decoding is*

$$\mathsf{C}_{\text{sep}}(\beta) = \beta\,\mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\}. \tag{3.196}$$

The spectral efficiency under optimal joint decoding is

$$\mathsf{C}_{\text{joint}}(\beta) = \beta\,\mathsf{E}\left\{\mathsf{D}(p_{Z|X,\mathsf{snr};\eta}\,||\,p_{Z|\mathsf{snr};\eta}\,|\,p_X)\right\} + \eta - 1 - \log\eta, \tag{3.197}$$

where η *is the PME multiuser efficiency by postulating a measure* q *that is identical to* p.

It is interesting to compare the performance of the real-valued channel and that of the complex-valued channel. We assume the in-phase and quadrature components of the input symbols are independent with identical distribution p'_X which has a variance of $\frac{1}{2}$. By Claim 3.4, the equivalent single-user channel (3.192) can also be regarded as two independent subchannels. The mean-square error and the variance in (3.193) are the sum of those of the subchannels. It can be checked that the performance of each subchannel is identical to that of the real-valued channel with input distribution p'_X normalized to unit variance. Note, however, that the complex spreading sequences take twice the energy of their real counterparts. In all, the efficiencies under complex-valued spreading are the same as those under real-valued spreading. This result simplifies the analysis of complex-valued channels such as those arise in multiantenna systems. On the other hand, if we have control over the channel state matrix, as in CDMA systems, complex-valued spreading should be avoided due to higher complexity with no performance gain.

We can also compare the real-valued and the complex-valued channel assuming the same real-valued input distribution. Under the complex-valued channel,

$$\begin{bmatrix}\boldsymbol{Y}^{(\mathrm{r})}\\ \boldsymbol{Y}^{(\mathrm{i})}\end{bmatrix} = \begin{bmatrix}\boldsymbol{S}^{(\mathrm{r})}\\ \boldsymbol{S}^{(\mathrm{i})}\end{bmatrix}\boldsymbol{X} + \begin{bmatrix}\boldsymbol{N}^{(\mathrm{r})}\\ \boldsymbol{N}^{(\mathrm{i})}\end{bmatrix}, \tag{3.198}$$

which is equivalent to transmitting the same $\boldsymbol{X}$ twice over two uses of real-valued channels. This is equivalent to having a real-valued channel with the load β halved.

3.6 Numerical Results

In Figures 3.10–3.11 we plot the simulated distribution of the posterior mean estimate and its corresponding "hidden" Gaussian statistic. Equal-power users with binary input are considered. We simulate CDMA systems of 4, 8, 12 and 16 users respectively. The load is fixed to $\beta = 2/3$ and the SNR is 2 dB. We collect the output decision statistics of the posterior mean estimator (i.e., the soft output of the individually optimal detector, $\langle X_k\rangle$) out of 1,000 experiments. A histogram of the statistic is obtained and then scaled to plot an estimate of the probability density function in Figure 3.10. We also apply the inverse

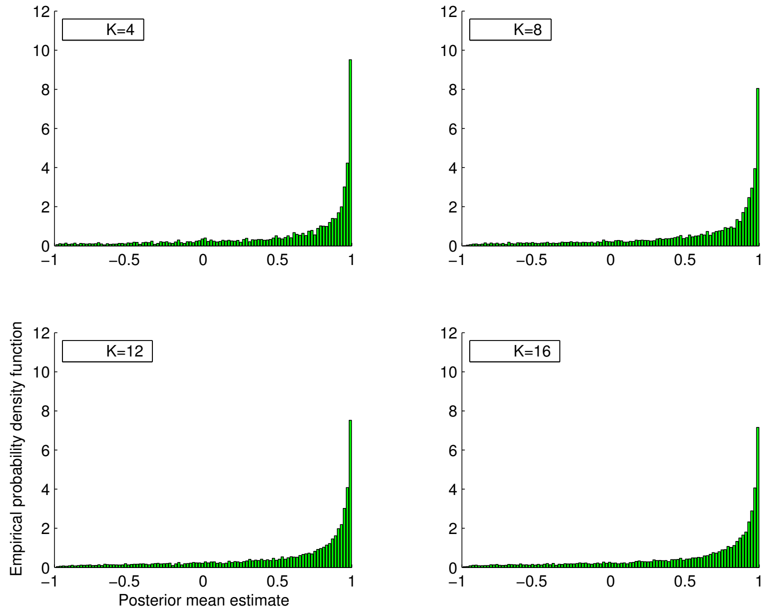

Figure 3.10: Simulated probability density function of the posterior mean estimates under binary input conditioned on "+1" being transmitted. Systems with 4, 8, 12 and 16 equal-power users are simulated with $\beta = 2/3$. The SNR is 2 dB.

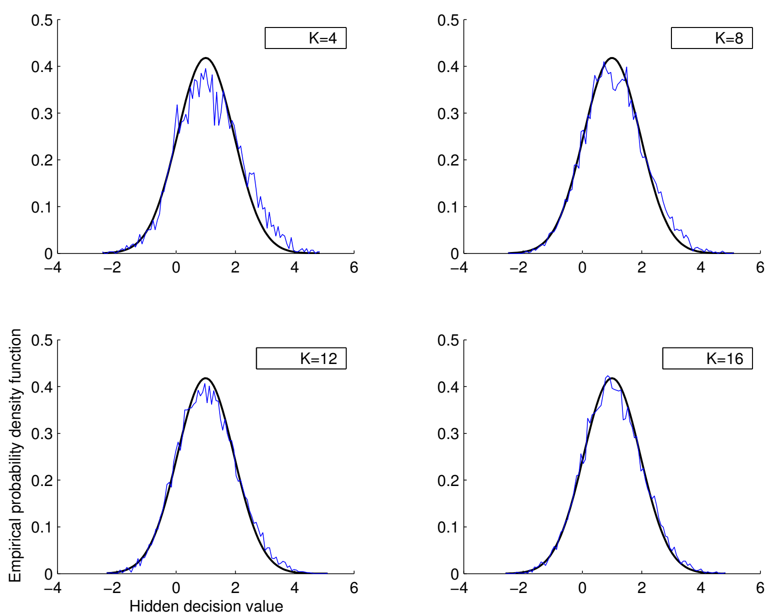

Figure 3.11: Simulated probability density function of the "hidden" Gaussian statistic recovered from the posterior mean estimates under binary input conditioned on "+1" being transmitted. Systems with 4, 8, 12 and 16 equal-power users are simulated with $\beta = 2/3$. The SNR is 2 dB. The asymptotic Gaussian distribution predicted by our theory is also plotted for comparison.

nonlinear decision function to recover the "hidden" Gaussian decision statistic (normalized so that its conditional mean value is X_k), which in this case is

$$Z_k = \frac{\tanh^{-1}(\langle X_k \rangle)}{\eta\, \mathsf{snr}_k}. \tag{3.199}$$

The probability density function of Z_k estimated from its histogram is then compared to the theoretically predicted Gaussian density function in Figure 3.11. It is clear that even though the PME output $\langle X_k \rangle$ takes a non-Gaussian distribution, the equivalent statistic Z_k converges to a Gaussian distribution as K becomes large. This result is particularly powerful considering that the "fit" to the Gaussian distribution is quite good even for a system with merely 8 users.

In Figures 3.12–3.23, the multiuser efficiency and the spectral efficiency are plotted as functions of the average SNR. We consider three input distributions, namely, real-valued Gaussian and binary inputs, and (complex-valued) 8PSK inputs. Under Gaussian and binary inputs, where real-valued spreading is considered, the multiuser efficiencies are given by (3.54) and (3.62) respectively, and the spectral efficiencies are given by (3.58)–(3.59), (2.17) and (3.63) respectively. Under 8PSK, where complex-valued spreading is assumed, the multiuser efficiency and the spectral efficiency are given by Claim 3.4 and Corollary 3.3 respectively. We also consider two SNR distributions: 1) equal SNR for all users (perfect power control), and 2) two groups of users of equal population with a power difference of 10 dB. We first assume a system load of $\beta = 1$ and then redo the experiments with $\beta = 3$.

In Figure 3.12, the multiuser efficiency under Gaussian inputs and linear MMSE detection is plotted as a function of the average SNR. The load is $\beta = 1$. We find the multiuser efficiencies decrease from 1 to 0 as the SNR increases. The monotonicity can be easily verified by inspecting the Tse-Hanly equation (3.54). Transmission with unbalanced power improves the multiuser efficiency. The corresponding spectral efficiencies of the system are plotted in Figure 3.13. Both joint decoding and separate decoding are considered. The gain in the spectral efficiency due to joint decoding is little under low SNR but significant under high SNR. Unbalanced SNR reduces the spectral efficiency, where under separate decoding the loss is almost negligible.

The multiuser efficiency under binary inputs and nonlinear MMSE (individually optimal) detection is plotted in Figure 3.14. The multiuser efficiency is not monotone. The multiuser efficiency converges to 1 for both diminishing SNR and infinite SNR. While for diminishing SNR this follows directly from the definition of multiuser efficiency, the convergence to unity as the SNR goes to infinity was shown in [98] for the case of binary inputs. A single dip is observed for the case of equal SNR while two dips are observed in the case of two SNRs of equal population with 10 dB difference in SNR (the gap is about 10 dB). The corresponding spectral efficiencies are plotted in Figure 3.15. The spectral efficiencies saturate to 1 bit/s/dimension at hight SNR. The difference between joint decoding and separate decoding is quite small for both very low and very high SNRs while it can be 25% at around 8 dB.

The multiuser efficiency under 8PSK inputs and nonlinear MMSE detection is plotted in Figure 3.16. The multiuser efficiency curve is slightly better than that for binary inputs. The corresponding spectral efficiencies are plotted in Figure 3.17. The spectral efficiencies saturate to 3 bit/s/dimension at hight SNR.

In Figures 3.18 through 3.23, we redo the previous experiments only with a different system load $\beta = 3$. The results are to be compared with Figures 3.12 through 3.17.

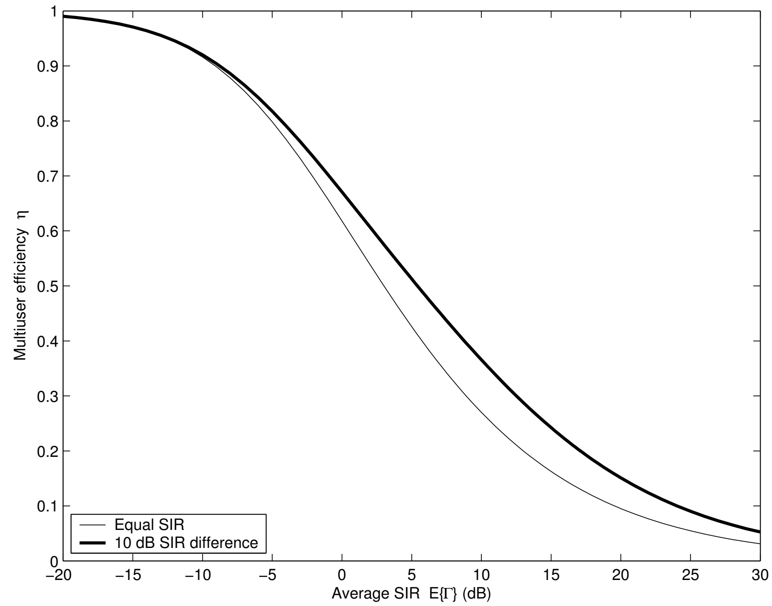

Figure 3.12: The multiuser efficiency vs. average SNR (Gaussian inputs, $\beta = 1$).

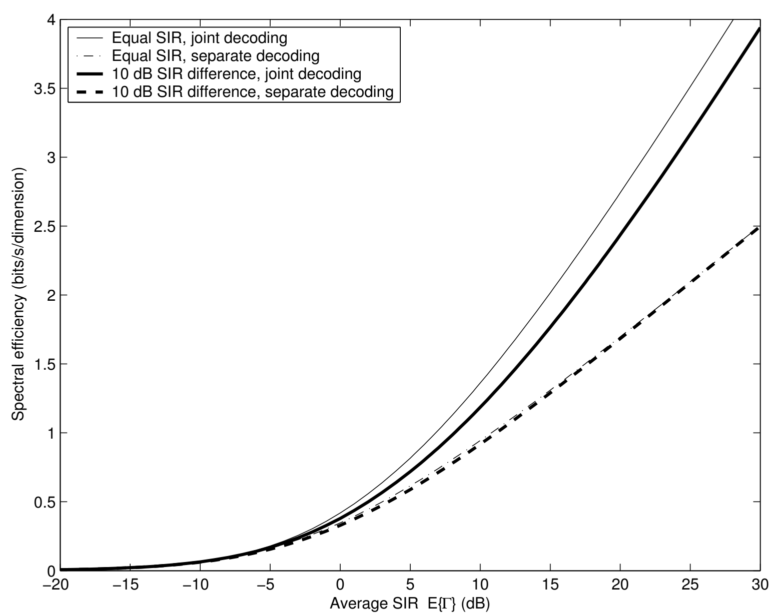

Figure 3.13: The spectral efficiency vs. average SNR (Gaussian inputs, $\beta = 1$).

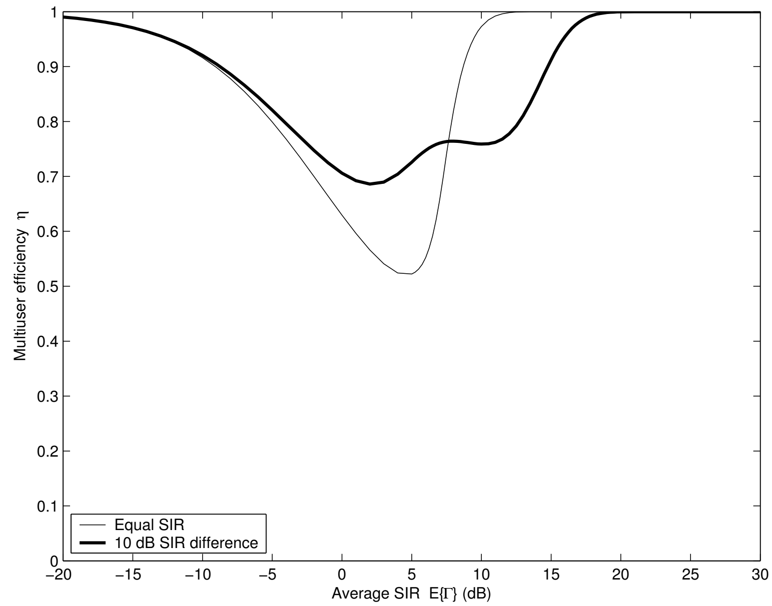

Figure 3.14: The multiuser efficiency vs. average SNR (binary inputs, $\beta = 1$).

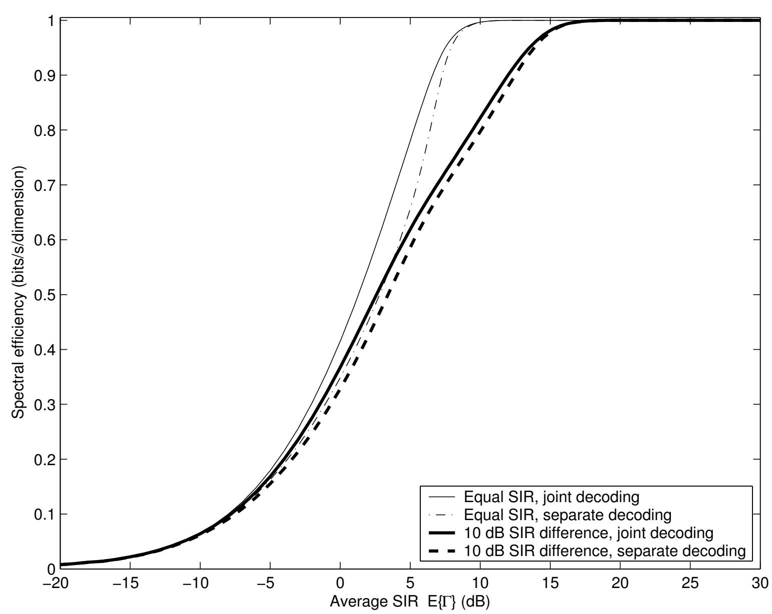

Figure 3.15: The spectral efficiency vs. average SNR (binary inputs, $\beta = 1$).

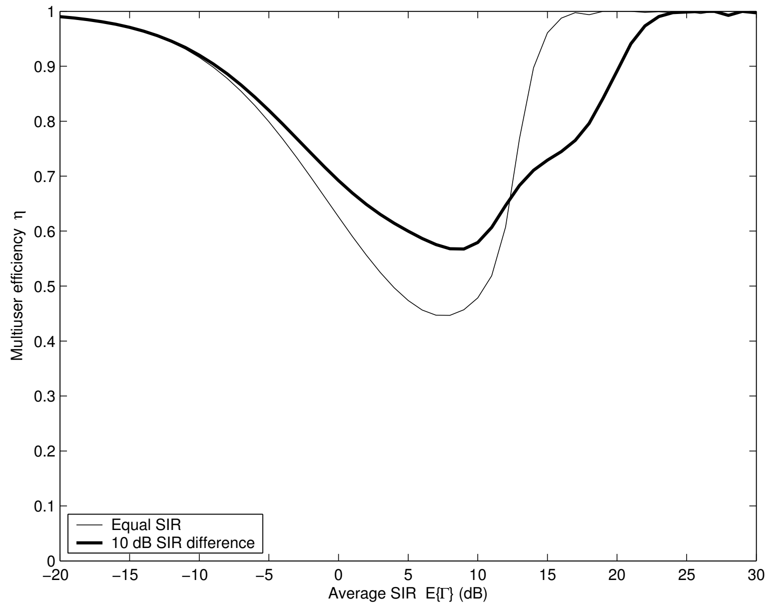

Figure 3.16: The multiuser efficiency vs. average SNR (8PSK inputs, $\beta = 1$).

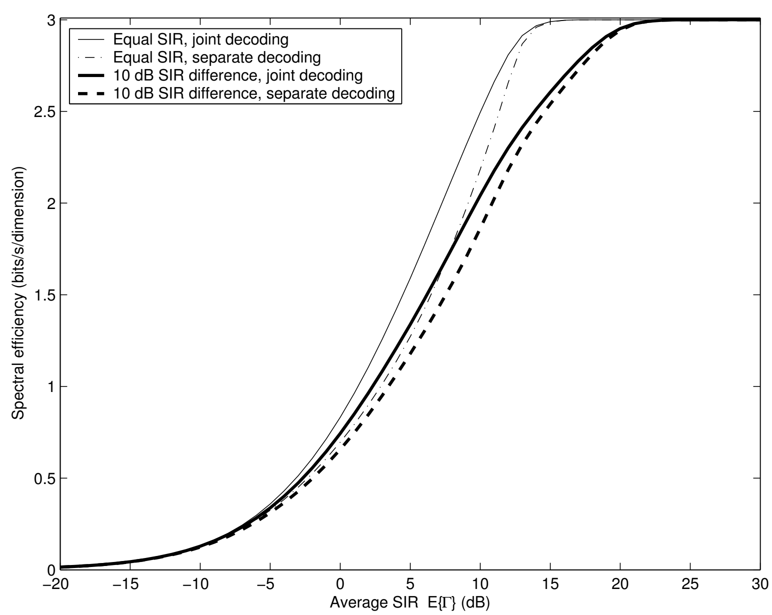

Figure 3.17: The spectral efficiency vs. average SNR (8PSK inputs, $\beta = 1$).

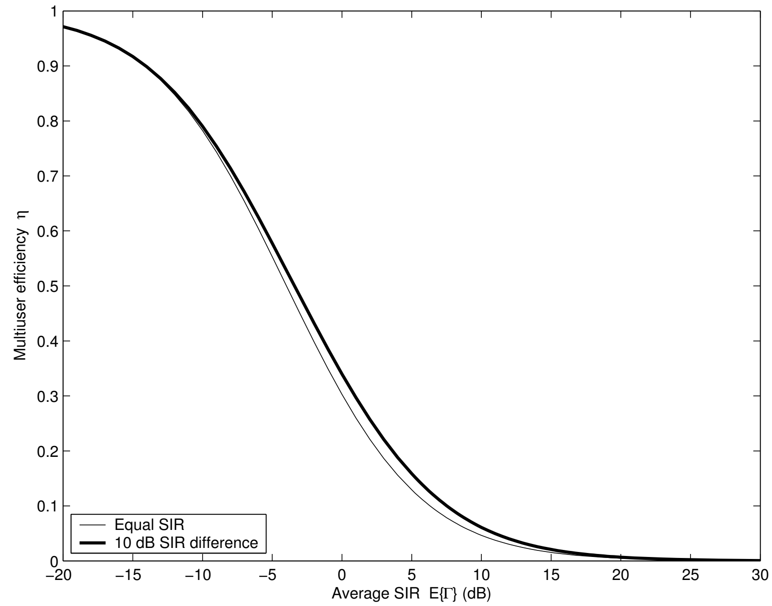

Figure 3.18: The multiuser efficiency vs. average SNR (Gaussian inputs, $\beta = 3$).

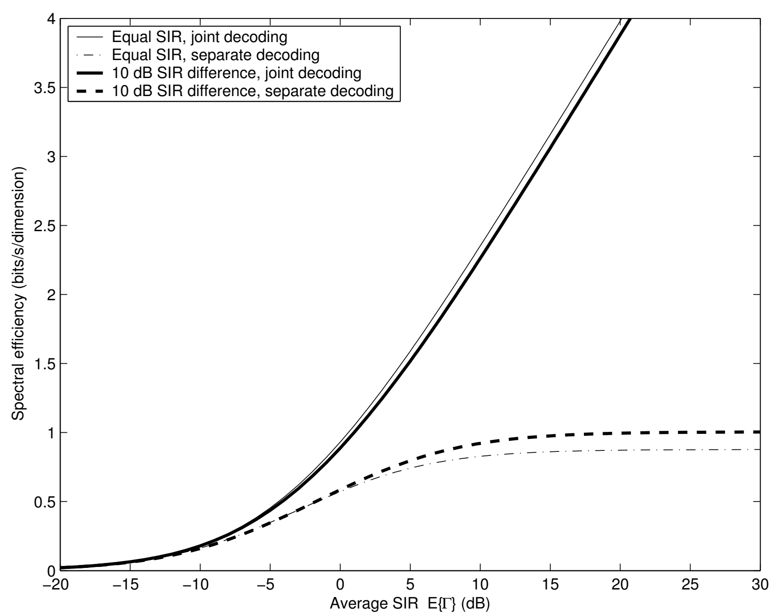

Figure 3.19: The spectral efficiency vs. average SNR (Gaussian inputs, $\beta = 3$).

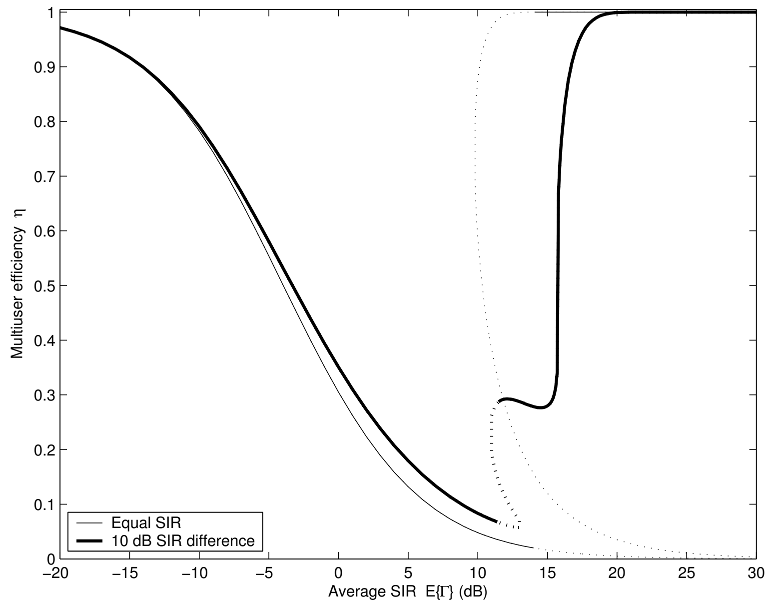

Figure 3.20: The multiuser efficiency vs. average SNR (binary inputs, $\beta = 3$).

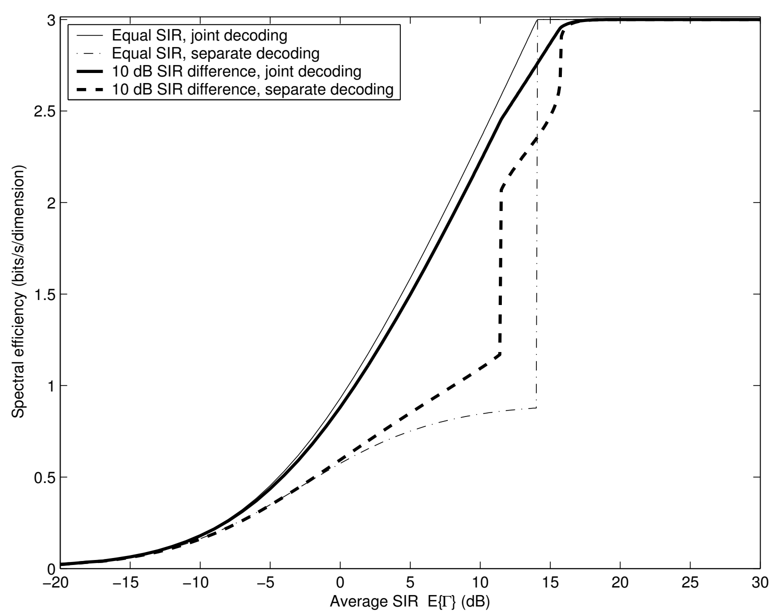

Figure 3.21: The spectral efficiency vs. average SNR (binary inputs, $\beta = 3$).

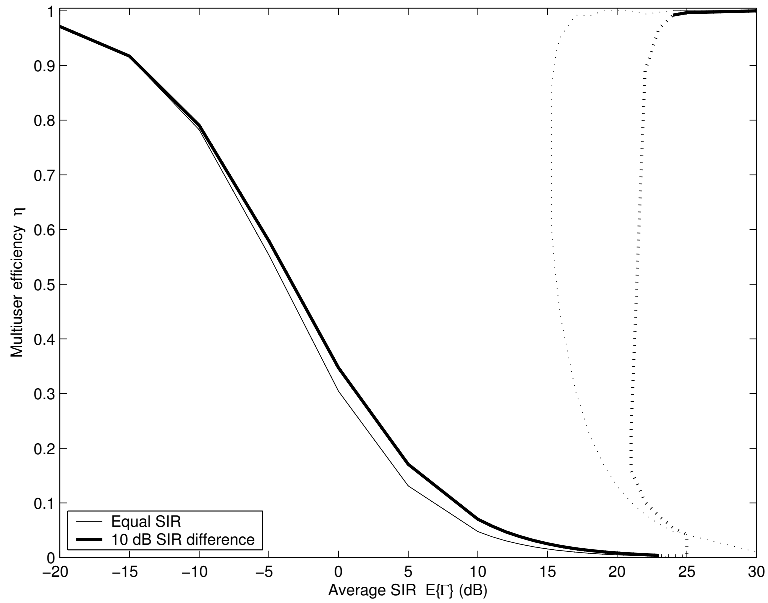

Figure 3.22: The multiuser efficiency vs. average SNR (8PSK inputs, $\beta = 3$).

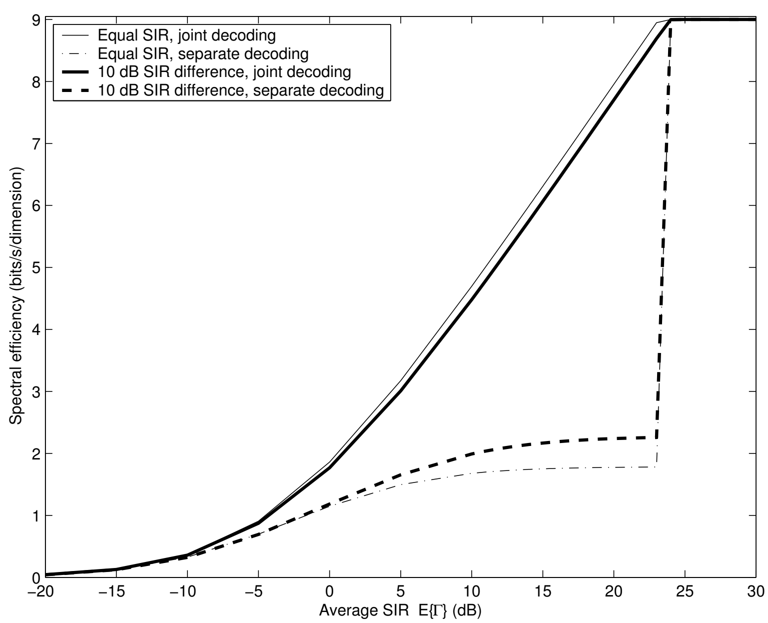

Figure 3.23: The spectral efficiency vs. average SNR (8PSK inputs, $\beta = 3$).

Under Gaussian inputs, the multiuser efficiency curves in Figure 3.18 take similar shape as in Figure 3.12, but are significantly lower due to the higher load. The corresponding spectral efficiencies are shown in Figure 3.19. It is clear that higher load results in higher spectrum usage under joint decoding. Separate decoding, however, is interference limited and saturates under high SNR (cf. [105, Figure 1]).

In Figure 3.20, we plot the multiuser efficiency under binary inputs. All solutions to the fixed-point equation (3.22) of the multiuser efficiency are shown. Under equal SNR, multiple solutions coexist for an average SNR of 10 dB or higher. If two groups of users with 10 dB difference in SNR, multiple solutions are seen in between 11 to 13 dB. The solution that minimizes the free energy is valid and is shown in solid lines, while invalid solutions are plotted using dotted lines. An almost 0 to 1 jump is observed under equal SNR and a much smaller jump is seen under unbalanced SNRs. This is known as phase transition in statistical physics. The asymptotics under equal SNR can be shown by taking the limit $\mathsf{snr} \to \infty$ in (3.62). Essentially, if $\eta\mathsf{snr} \to \infty$, then $\eta \to 1$; while if $\eta\mathsf{snr} \to \tau$ where τ is the solution to

$$\tau \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} [1 - \tanh(\tau - z\sqrt{\tau})] \, \mathrm{d}z = \frac{1}{\beta}, \tag{3.200}$$

then $\eta \to 0$. If $\beta > 2.085$, there exists a solution to (3.200) so that two solutions coexist for large SNR.

The spectral efficiency under binary inputs and $\beta = 3$ is shown in Figure 3.21. As a result of phase transition, one observes a jump to saturation in the spectral efficiency under equal-power binary inputs. The gain due to joint decoding can be significant in moderate SNRs. In case of two groups of users with 10 dB difference in SNR, the spectral efficiency curve also shows one jump and the loss due to separate decoding is reduced significantly for a small window of SNRs around the areas of phase transition (11–13 dB). Therefore, perfect power control may not be the best strategy in such cases.

Under 8PSK inputs, the multiuser efficiency and spectral efficiency curves in Figure 3.22 and 3.23 take similar shape as the curves under binary inputs in Figure 3.20 and 3.21. Phase transition causes jumps in both the multiuser efficiency and the spectral efficiency.

A comparison of Figures 3.19, 3.21 and 3.23 shows that under separate decoding, the spectral efficiency under Gaussian inputs saturates well below that of binary and 8PSK inputs. This implies that the nonlinear MMSE detector with separate decoding is not efficient in case of dense constellation.

3.7 Summary

The main contribution of this chapter is a simple characterization of the multiuser efficiency and spectral efficiency of CDMA multiuser detection under arbitrary input distribution and SNR (and/or flat fading) in the large-system limit. A broad family of multiuser detectors is studied under the name of generalized posterior mean estimators, which includes well-known detectors such as the matched filter, decorrelator, linear MMSE detector, maximum likelihood (jointly optimal) detector, and the individually optimal detector.

A key conclusion is the decoupling of a Gaussian CDMA channel concatenated with a generic multiuser detector front end (Claims 3.1 and 3.3). It is found that the multiuser detection output is a deterministic function of a hidden Gaussian statistic centered at the transmitted symbol. Hence the single-user channel seen at the multiuser detection output is equivalent to a Gaussian channel in which the overall effect of multiple-access interference is a degradation factor in the effective signal-to-interference ratio. This degradation factor, known as the multiuser efficiency, is the solution to a set of coupled fixed-point equations, and can be computed easily numerically if not analytically.

Another set of results, tightly related to the decoupling principle, are some general formulas for the large-system spectral efficiency of CDMA channels expressed in terms of the multiuser efficiency, both under joint and separate decoding (Corollaries 3.1 and 3.2). It is found that the decomposition of optimum spectral efficiency as a sum of single-user efficiencies and a joint decoding gain (Claim 3.2) applies under more general conditions than shown in [85], thereby validating Müller's conjecture [71]. A relationship between the spectral efficiencies under joint and separate decoding (Theorem 3.1) is an outcome of the central formula that links the mutual information and MMSE in Chapter 2.

From a practical viewpoint, this chapter presents new results on the efficiency of CDMA communication under arbitrary input signaling such as m-PSK and m-QAM and arbitrary power profile. More importantly, the results in this chapter allow the performance of multiuser detection to be characterized by a single parameter, the multiuser efficiency. The efficiency of spectrum usage is also easily quantified by means of this parameter. Thus, our results offer valuable insights in the design and analysis of CDMA systems.

The linear system in our study also models multiple-input multiple-output channels where the channel state is unknown at the transmitter. The results can thus be used to evaluate the output SNR or spectral efficiency of high-dimensional MIMO channels (such as multiple-antenna systems) with arbitrary signaling and various detection techniques.

Results in this chapter have been published in part in [38, 39, 42, 41, 40] and the most recent development has been accepted for publication (pending revision) [43]. The results also form the basis for some other work [67, 113].

Chapter 4

Conclusion and Future Work

This thesis presents theoretical findings on information transmission and signal estimation pertinent to additive Gaussian channels. The central quantities, the input-output mutual information and the minimum mean-square error, are shown to be tightly related in both single-user and multiuser settings.

A fundamental result in this thesis is that the derivative of the mutual information (nats) between the input and output of a Gaussian channel with respect to the signal-to-noise ratio is always equal to a half of the minimum mean-square error in estimating the input given the output. This relationship holds for arbitrary scalar and vector input signals, as well as for discrete-time and continuous-time noncausal MMSE estimation. Using information-theoretic results, a one-to-one correspondence invariant to the input signaling is also identified between the continuous-time filtering and smoothing MMSEs as functions of the signal-to-noise ratio. That is, the filtering MMSE at any SNR is the average of the smoothing MMSE from zero signal-to-noise ratio to SNR. In the discrete-time setting, the relationship between the mutual information and the causal MMSEs takes the form of inequalities, i.e., the mutual information is lower bounded by the filtering MMSE but upper bounded by the prediction MMSE.

The incremental channel approach, which is developed to study mutual information increase due to infinitesimal change in system parameters, is applicable to the entire family of channels the noise of which has independent increments, i.e., that is characterized by Lévy processes [5]. We are currently investigating the special case of Poisson channels, whose output is a counting process the rate of which is affine in the input process [34, 35].

The mutual information-MMSE relationship applies to many objects of engineering interests, including for instance random fields. Some applications of the relationship are presented in this thesis. We hope that the central formula may help to solve some open problems, in particular to evaluate the mutual information by way of finding the MMSE. Particularly interesting cases are the intersymbol interference channel and fading channels. It is also hoped that the new relationships may help to explain the ubiquitous MMSE estimator in capacity-achieving schemes, e.g., in ISI channels [12, 13], CDMA [33, 102], lattice codes [26], writing on dirty paper schemes [15], and diversity-multiplexing trade-off of MIMO channels [24]. An important question yet to be answered is whether the new relationships carry useful operational meanings.

In a multiuser setting, randomly spread code-division multiple access and multiuser detection are also studied in this thesis. The analysis is conducted in the large-system limit using the replica method developed in statistical physics. It is shown that Gaussian CDMA channels with a multiuser detector front end can be decoupled into scalar Gaussian channels where the multiple-access interference present in the detection outputs is effectively summarized by a single parameter called multiuser efficiency. This decoupling principle is true for arbitrary input distribution and flat fading. A unified framework is developed to analyze a broad family of multiuser detectors including the matched filter, decorrelator, linear MMSE detector, and optimum detectors, etc. Spectral efficiencies under both joint and separate decoding are also derived. The additive decomposition of optimum capacity as a sum of single-user capacities and a joint decoding gain holds for arbitrary inputs. Numerical results suggest that the performance of an 8-user system can be closely approximated by the large-system limit. Based on a general linear vector channel model, these results are also applicable to single-user MIMO channels such as in multiantenna systems.

The analysis through the replica method originally developed in statistical physics is quite unusual in the area of communications and information theory. Interesting problems for future research include validating the assumptions made in Chapter 3, namely, 1) the asymptotic equipartition property in matrix channel setting, 2) the "replica trick", and 3) replica symmetry. It also remains to see what are the conditions for replica symmetry to hold. Neither do we know much about the operational meanings of phase transition. It would also be interesting to see whether the replica method can be applied to solve open problems such as the input-output mutual information of ISI channels. A great challenge is to investigate whether the free energy (or mutual information) admits a solution without resorting to the replica method. A possible starter is to fix the specious interpretation in Section 3.2.6.

Appendix A

A.1 A Fifth Proof of Theorem 2.1

Proof: For simplicity, it is assumed that the order of expectation and derivative can be exchanged freely. A rigorous proof is relegated to Appendix A.2 where every such assumption is validated in the case of the more general vector model. The input-output conditional probability density function is given by (2.6). Let us define $p_i(y;\mathsf{snr})$ as in (2.103). Then

$$\begin{aligned} I(\mathsf{snr}) &= \mathsf{E}\left\{\log \frac{p_{Y|X;\mathsf{snr}}(Y|X;\mathsf{snr})}{p_{Y;\mathsf{snr}}(Y;\mathsf{snr})}\right\} && \text{(A.1)}\\ &= -\frac{1}{2}\log(2\pi e) - \int p_0(y;\mathsf{snr})\log p_0(y;\mathsf{snr})\,\mathrm{d}y. && \text{(A.2)} \end{aligned}$$

It is easy to check that

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}}p_i(y;\mathsf{snr}) = \frac{1}{2\sqrt{\mathsf{snr}}}yp_{i+1}(y;\mathsf{snr}) - \frac{1}{2}p_{i+2}(y;\mathsf{snr}) = -\frac{1}{2\sqrt{\mathsf{snr}}}\frac{\mathrm{d}}{\mathrm{d}y}p_{i+1}(y;\mathsf{snr}) \qquad \text{(A.3)}$$

as long as $p_{i+2}(y;\mathsf{snr})$ is well-defined. Therefore, from (A.2),

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}}I(\mathsf{snr}) &= -\int [\log p_0(y;\mathsf{snr})+1]\,\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}}p_0(y;\mathsf{snr})\,\mathrm{d}y && \text{(A.4)}\\ &= \frac{1}{2\sqrt{\mathsf{snr}}}\int \log p_0(y;\mathsf{snr})\,\frac{\mathrm{d}}{\mathrm{d}y}p_1(y;\mathsf{snr})\,\mathrm{d}y && \text{(A.5)}\\ &= -\frac{1}{2\sqrt{\mathsf{snr}}}\int \frac{p_1(y;\mathsf{snr})}{p_0(y;\mathsf{snr})}\,\frac{\mathrm{d}}{\mathrm{d}y}p_0(y;\mathsf{snr})\,\mathrm{d}y && \text{(A.6)}\\ &= -\frac{1}{2\sqrt{\mathsf{snr}}}\int \frac{p_1(y;\mathsf{snr})}{p_0(y;\mathsf{snr})}\,\left[\sqrt{\mathsf{snr}}\,p_1(y;\mathsf{snr}) - yp_0(y;\mathsf{snr})\right]\,\mathrm{d}y && \text{(A.7)}\\ &= \frac{1}{2\sqrt{\mathsf{snr}}}\int \frac{p_1(y;\mathsf{snr})}{p_0(y;\mathsf{snr})}\left[y - \sqrt{\mathsf{snr}}\,\frac{p_1(y;\mathsf{snr})}{p_0(y;\mathsf{snr})}\right]p_0(y;\mathsf{snr})\,\mathrm{d}y, && \text{(A.8)} \end{aligned}$$

where (A.6) is by integrating by parts. Note that the fraction in (A.8) is exactly the conditional-mean estimate:

$$\widehat{X}(y;\mathsf{snr}) = \frac{p_1(y;\mathsf{snr})}{p_0(y;\mathsf{snr})}. \qquad \text{(A.9)}$$

Therefore,

$$\frac{\mathrm{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = \frac{1}{2\sqrt{\mathsf{snr}}} \mathsf{E}\left\{ \mathsf{E}\left\{ X \mid Y; \mathsf{snr} \right\} \left[Y - \sqrt{\mathsf{snr}}\, \mathsf{E}\left\{ X \mid Y; \mathsf{snr} \right\} \right] \right\} \tag{A.10}$$
$$= \frac{1}{2} \mathsf{E}\left\{ X^2 - \left(\mathsf{E}\left\{ X \mid Y; \mathsf{snr} \right\} \right)^2 \right\} \tag{A.11}$$
$$= \frac{1}{2} \mathsf{mmse}(\mathsf{snr}). \tag{A.12}$$

■

Using the above technique, it is not difficult to find the derivative of the conditional-mean estimate $\widehat{X}(y; \mathsf{snr})$ (A.9) with respect to the SNR. In fact, one can find any derivative of the mutual information in this way.

A.2 A Fifth Proof of Theorem 2.2

Proof: The vector channel (2.18) has a Gaussian conditional density (2.19). The unconditional density of the channel output is given by (2.73), which is strictly positive for all $\boldsymbol{y}$. The mutual information can be written as

$$I(\mathsf{snr}) = -\frac{L}{2} \log(2\pi e) - \int p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr}) \log p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr})\, \mathrm{d}\boldsymbol{y}. \tag{A.13}$$

Hence,

$$\frac{\mathrm{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = -\int \left[\log p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr}) + 1 \right] \frac{\mathrm{d}}{\mathsf{dsnr}} p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr})\, \mathrm{d}\boldsymbol{y} \tag{A.14}$$
$$= -\int \left[\log p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr}) + 1 \right] \mathsf{E}\left\{ \frac{\mathrm{d}}{\mathsf{dsnr}} p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X}; \mathsf{snr}) \right\} \mathrm{d}\boldsymbol{y}, \tag{A.15}$$

where the order of taking the derivative and expectation in (A.15) can be exchanged by Lemma A.1, which is shown below in this Appendix. It is easy to check that

$$\frac{\mathrm{d}}{\mathsf{dsnr}} p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x}; \mathsf{snr}) = \frac{1}{2\sqrt{\mathsf{snr}}} (\boldsymbol{H}\boldsymbol{x})^\top \left(\boldsymbol{y} - \sqrt{\mathsf{snr}}\, \boldsymbol{H}\boldsymbol{x} \right) p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x}; \mathsf{snr}) \tag{A.16}$$
$$= -\frac{1}{2\sqrt{\mathsf{snr}}} (\boldsymbol{H}\boldsymbol{x})^\top \nabla p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x}; \mathsf{snr}). \tag{A.17}$$

Using (A.17), the right hand side of (A.15) can be written as

$$\frac{1}{\sqrt{\mathsf{snr}}} \mathsf{E}\left\{ (\boldsymbol{H}\boldsymbol{X})^\top \int \left[\log p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr}) + 1 \right] \nabla p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X}; \mathsf{snr})\, \mathrm{d}\boldsymbol{y} \right\}. \tag{A.18}$$

The integral in (A.18) can be carried out by parts to obtain

$$-\int p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X}; \mathsf{snr})\, \nabla \left[\log p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr}) + 1 \right] \mathrm{d}\boldsymbol{y}, \tag{A.19}$$

since for $\forall \boldsymbol{x}$,

$$p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x}; \mathsf{snr}) \left[\log p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y}; \mathsf{snr}) + 1 \right] \to 0 \quad \text{as } \|\boldsymbol{y}\| \to \infty. \tag{A.20}$$

Hence, the right hand side of (A.18) can be further evaluated as

$$-\frac{1}{\sqrt{\mathsf{snr}}}\int \mathsf{E}\left\{(\boldsymbol{H}\boldsymbol{X})^{\top}\frac{p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X};\mathsf{snr})}{p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y};\mathsf{snr})}\right\}\nabla p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y};\mathsf{snr})\,\mathrm{d}\boldsymbol{y} \tag{A.21}$$

where we have changed the order of the expectation with respect to $\boldsymbol{X}$ and the integral (i.e., expectation with respect to $\boldsymbol{Y}$). By (A.17) and Lemma A.2 (shown below in this Appendix), (A.21) can be further written as

$$\frac{1}{2\sqrt{\mathsf{snr}}}\int \mathsf{E}\left\{(\boldsymbol{H}\boldsymbol{X})^{\top}\,\middle|\,\boldsymbol{Y}=\boldsymbol{y};\mathsf{snr}\right\}\mathsf{E}\left\{\left(\boldsymbol{y}-\sqrt{\mathsf{snr}}\,\boldsymbol{H}\boldsymbol{X}\right)p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X};\mathsf{snr})\right\}\,\mathrm{d}\boldsymbol{y}. \tag{A.22}$$

Therefore, (A.15) can be rewritten as

$$\begin{aligned}
2\frac{\mathrm{d}}{\mathsf{dsnr}}I(\mathsf{snr}) &= \frac{1}{\sqrt{\mathsf{snr}}}\int \mathsf{E}\left\{(\boldsymbol{H}\boldsymbol{X})^{\top}\,\middle|\,\boldsymbol{Y}=\boldsymbol{y};\mathsf{snr}\right\} \\
&\qquad\times\mathsf{E}\left\{\boldsymbol{y}-\sqrt{\mathsf{snr}}\,\boldsymbol{H}\boldsymbol{X}\,\middle|\,\boldsymbol{Y}=\boldsymbol{y};\mathsf{snr}\right\}p_{\boldsymbol{Y};\mathsf{snr}}(\boldsymbol{y};\mathsf{snr})\,\mathrm{d}\boldsymbol{y} && \text{(A.23)}\\
&= \frac{1}{\sqrt{\mathsf{snr}}}\mathsf{E}\left\{\mathsf{E}\left\{(\boldsymbol{H}\boldsymbol{X})^{\top}\,\middle|\,\boldsymbol{Y};\mathsf{snr}\right\}\mathsf{E}\left\{\boldsymbol{Y}-\sqrt{\mathsf{snr}}\,\boldsymbol{H}\boldsymbol{X}\,\middle|\,\boldsymbol{Y};\mathsf{snr}\right\}\right\} && \text{(A.24)}\\
&= \frac{1}{\sqrt{\mathsf{snr}}}\mathsf{E}\left\{(\boldsymbol{H}\boldsymbol{X})^{\top}\boldsymbol{Y}\right\}-\mathsf{E}\left\{\left\|\mathsf{E}\left\{\boldsymbol{H}\boldsymbol{X}\mid\boldsymbol{Y};\mathsf{snr}\right\}\right\|^2\right\} && \text{(A.25)}\\
&= \mathsf{E}\left\{\|\boldsymbol{H}\boldsymbol{X}\|^2\right\}-\mathsf{E}\left\{\left\|\mathsf{E}\left\{\boldsymbol{H}\boldsymbol{X}\mid\boldsymbol{Y};\mathsf{snr}\right\}\right\|^2\right\} && \text{(A.26)}\\
&= \mathsf{E}\left\{\left\|\boldsymbol{H}\,\boldsymbol{X}-\boldsymbol{H}\,\mathsf{E}\left\{\boldsymbol{X}\mid\boldsymbol{Y};\mathsf{snr}\right\}\right\|^2\right\}. && \text{(A.27)}
\end{aligned}$$

■

The following two lemmas were needed to justify the exchange of expectation with respect to $P_{\boldsymbol{X}}$ and derivatives in the above proof of Theorem 2.2.

Lemma A.1 *If* $\mathsf{E}\|\boldsymbol{X}\|^2<\infty$, *then*

$$\frac{\mathrm{d}}{\mathsf{dsnr}}\mathsf{E}\left\{p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X};\mathsf{snr})\right\}=\mathsf{E}\left\{\frac{\mathrm{d}}{\mathsf{dsnr}}p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X};\mathsf{snr})\right\}. \tag{A.28}$$

Proof: Let

$$f_\delta(\boldsymbol{x},\boldsymbol{y},\mathsf{snr})=\frac{1}{\delta}\left[p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X};\mathsf{snr}+\delta)-p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X};\mathsf{snr})\right] \tag{A.29}$$

and

$$f(\boldsymbol{x},\boldsymbol{y},\mathsf{snr})=\frac{\mathrm{d}}{\mathsf{dsnr}}p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x};\mathsf{snr}). \tag{A.30}$$

Then, $\forall\,\boldsymbol{x},\boldsymbol{y},\mathsf{snr}$,

$$\lim_{\delta\to0}f_\delta(\boldsymbol{x},\boldsymbol{y},\mathsf{snr})=f(\boldsymbol{x},\boldsymbol{y},\mathsf{snr}). \tag{A.31}$$

Lemma A.1 is equivalent to

$$\lim_{\delta\to0}\int f_\delta(\boldsymbol{x},\boldsymbol{y},\mathsf{snr})P_{\boldsymbol{X}}(\,\mathrm{d}\boldsymbol{x})=\int f(\boldsymbol{x},\boldsymbol{y},\mathsf{snr})P_{\boldsymbol{X}}(\,\mathrm{d}\boldsymbol{x}). \tag{A.32}$$

Suppose we can show that for every $\delta,\boldsymbol{x},\boldsymbol{y}$ and snr,

$$|f_\delta(\boldsymbol{x},\boldsymbol{y},\mathsf{snr})|<\|\boldsymbol{H}\boldsymbol{x}\|^2+\frac{1}{\sqrt{\mathsf{snr}}}|\boldsymbol{y}^{\top}\boldsymbol{H}\boldsymbol{x}|. \tag{A.33}$$

Since the right hand side of (A.33) is integrable with respect to $P_{\boldsymbol{X}}$ by the assumption in the lemma, (A.32) holds by the Lebesgue convergence theorem [83]. Note that

$$f_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}) = (2\pi)^{-\frac{L}{2}} \frac{1}{\delta} \left\{ \exp\left[-\frac{1}{2} \|\boldsymbol{y} - \sqrt{\mathsf{snr} + \delta}\, \boldsymbol{H}\boldsymbol{x}\|^2 \right] - \exp\left[-\frac{1}{2} \|\boldsymbol{y} - \sqrt{\mathsf{snr}}\, \boldsymbol{H}\boldsymbol{x}\|^2 \right] \right\}. \tag{A.34}$$

If

$$\frac{1}{\delta} \leq \|\boldsymbol{H}\boldsymbol{x}\|^2 + \frac{1}{\sqrt{\mathsf{snr}}} |\boldsymbol{y}^\top \boldsymbol{H}\boldsymbol{x}|, \tag{A.35}$$

then (A.33) holds trivially. Otherwise,

$$\begin{aligned} |f_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr})| &< \frac{1}{\delta} \left| \exp\left[\frac{1}{2} \|\boldsymbol{y} - \sqrt{\mathsf{snr}}\, \boldsymbol{H}\boldsymbol{x}\|^2 - \frac{1}{2} \|\boldsymbol{y} - \sqrt{\mathsf{snr} + \delta}\, \boldsymbol{H}\boldsymbol{x}\|^2 \right] - 1 \right| && \text{(A.36)} \\ &< \frac{1}{2\delta} \left(\exp\left| \delta\, \|\boldsymbol{H}\boldsymbol{x}\|^2 - \left(\sqrt{\mathsf{snr} + \delta} - \sqrt{\mathsf{snr}} \right) \boldsymbol{y}^\top \boldsymbol{H}\boldsymbol{x} \right| - 1 \right) && \text{(A.37)} \\ &< \frac{1}{2\delta} \left(\exp\left[\delta \left(\|\boldsymbol{H}\boldsymbol{x}\|^2 + \frac{1}{\sqrt{\mathsf{snr}}} |\boldsymbol{y}^\top \boldsymbol{H}\boldsymbol{x}| \right) \right] - 1 \right). && \text{(A.38)} \end{aligned}$$

Using the fact

$$e^t - 1 < 2t, \quad \forall\, 0 \leq t < 1, \tag{A.39}$$

the inequality (A.33) holds for all $\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}$. ■

Lemma A.2 *If* $\mathsf{E}\boldsymbol{X}$ *exists, then for* $i = 1, \ldots, L$,

$$\frac{\partial}{\partial y_i} \mathsf{E}\left\{ p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{Y}|\boldsymbol{X}; \mathsf{snr}) \right\} = \mathsf{E}\left\{ \frac{\partial}{\partial y_i} p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{Y}|\boldsymbol{X}; \mathsf{snr}) \right\}. \tag{A.40}$$

Proof: Let

$$g(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}) = \frac{\partial}{\partial y_i} p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{x}; \mathsf{snr}) \tag{A.41}$$

and

$$g_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}) = \frac{1}{\delta} \left[p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y} + \delta\, \boldsymbol{e}_i | \boldsymbol{X}; \mathsf{snr}) - p_{\boldsymbol{Y}|\boldsymbol{X};\mathsf{snr}}(\boldsymbol{y}|\boldsymbol{X}; \mathsf{snr}) \right] \tag{A.42}$$

where $\boldsymbol{e}_i$ is a vector with all zero except on the i^{th} entry, which is 1. Then, $\forall \boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}$,

$$\lim_{\delta \to 0} g_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}) = g(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}). \tag{A.43}$$

Lemma A.2 is equivalent to

$$\lim_{\delta \to 0} \int g_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}) P_{\boldsymbol{X}}(\,\mathrm{d}\boldsymbol{x}) = \int g(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}) P_{\boldsymbol{X}}(\,\mathrm{d}\boldsymbol{x}). \tag{A.44}$$

If one can show that

$$|g_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr})| < |y_i| + 1 + \sqrt{\mathsf{snr}}\, |[\boldsymbol{H}\boldsymbol{x}]_i|, \tag{A.45}$$

then (A.44) holds by the Lebesgue convergence theorem since the right hand side of (A.45) is integrable with respect to $P_{\boldsymbol{X}}$ by assumption. Note that

$$g_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr}) = (2\pi)^{-\frac{L}{2}} \frac{1}{\delta} \left\{ \exp\left[-\frac{1}{2} \|\boldsymbol{y} + \delta\, \boldsymbol{e}_i - \sqrt{\mathsf{snr}}\, \boldsymbol{H}\boldsymbol{x}\|^2 \right] - \exp\left[-\frac{1}{2} \|\boldsymbol{y} - \sqrt{\mathsf{snr}}\, \boldsymbol{H}\boldsymbol{x}\|^2 \right] \right\}. \tag{A.46}$$

If
$$\frac{1}{\delta} \le |y_i| + 1 + \frac{1}{\sqrt{\mathsf{snr}}} |[\boldsymbol{Hx}]_i|, \tag{A.47}$$
then (A.45) holds trivially. Otherwise,
$$\begin{aligned}
|g_\delta(\boldsymbol{x}, \boldsymbol{y}, \mathsf{snr})| &< \frac{1}{2\delta}\left(\exp\left|\frac{1}{2}\|\boldsymbol{y} - \sqrt{\mathsf{snr}}\,\boldsymbol{Hx}\|^2 - \frac{1}{2}\|\boldsymbol{y} + \delta\,\boldsymbol{e}_i - \sqrt{\mathsf{snr}}\,\boldsymbol{Hx}\|^2\right| - 1\right) && \text{(A.48)}\\
&= \frac{1}{2\delta}\left(\exp\left|\frac{\delta}{2}\left(2y_i + \delta - 2\sqrt{\mathsf{snr}}\,[\boldsymbol{Hx}]_i\right)\right| - 1\right) && \text{(A.49)}\\
&< \frac{1}{2\delta}\left(\exp\left[\delta\left(|y_i| + 1 + \sqrt{\mathsf{snr}}\,|[\boldsymbol{Hx}]_i|\right)\right] - 1\right) && \text{(A.50)}\\
&< |y_i| + 1 + \sqrt{\mathsf{snr}}\,|[\boldsymbol{Hx}]_i|. && \text{(A.51)}
\end{aligned}$$

■

A.3 Proof of Lemma 2.5

Lemma 2.5 can be regarded as a consequence of Duncan's Theorem (Theorem 2.7). Consider the interval $[0, T]$. The mutual information can be expressed as a time-integral of the causal MMSE:
$$I\left(Z_0^T; Y_0^T\right) = \frac{\delta}{2}\int_0^T \mathsf{E}\left(Z_t - \mathsf{E}\left\{Z_t \mid Y_0^t; \delta\right\}\right)^2 \mathrm{d}t, \tag{A.52}$$
Notice that as the signal-to-noise ratio $\delta \to 0$, the observed signal Y_0^T becomes inconsequential in estimating the input signal. Indeed, the causal MMSE estimate converges to the unconditional-mean in mean-square sense:
$$\mathsf{E}\left\{Z_t \mid Y_0^t; \delta\right\} \to \mathsf{E}Z_t. \tag{A.53}$$
Putting (A.52) and (A.53) together proves Lemma 2.5.

In parallel with the development in Theorem 2.1, another reasoning of Lemma 2.5 from first principles without invoking Duncan's Theorem is presented in the following. In fact, Lemma 2.5 is established first in this paper so that a more intuitive proof of Duncan's Theorem is given in Section 2.3.3 using the idea of time-incremental channels.

Proof: [Lemma 2.5] By definition (2.1), the mutual information is the expectation of the logarithm of the Radon-Nikodym derivative, which can be obtained by the chain rule as
$$\Phi = \frac{\mathrm{d}\mu_{YZ}}{\mathrm{d}\mu_Y\,\mathrm{d}\mu_Z} = \frac{\mathrm{d}\mu_{YZ}}{\mathrm{d}\mu_{WZ}}\left(\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_W}\right)^{-1}. \tag{A.54}$$

First assume that $\{Z_t\}$ is a bounded uniformly stepwise process, i.e., there exists a finite subdivision of $[0, T]$, $0 = t_0 < t_1 < \cdots < t_n = T$, and a finite constant M such that
$$Z_t(\omega) = Z_{t_i}(\omega), \quad t \in [t_i, t_{i+1}],\ i = 0, \dots, n-1, \tag{A.55}$$
and $Z_t(\omega) < M$, $\forall t \in [0, T]$. Let $\boldsymbol{Z} = [Z_{t_0}, \dots, Z_{t_n}]$, $\boldsymbol{Y} = [Y_{t_0}, \dots, Y_{t_n}]$, and $\boldsymbol{W} = [W_{t_0}, \dots, W_{t_n}]$ be $(n+1)$-dimensional vectors formed by the samples of the random processes. Then, the input-output conditional density is Gaussian:
$$p_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z}) = \prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi(t_{i+1} - t_i)}} \exp\left[-\frac{\left(y_{i+1} - y_i - \sqrt{\delta} z_i (t_{i+1} - t_i)\right)^2}{2(t_{i+1} - t_i)}\right]. \tag{A.56}$$

Easily,

$$\frac{p_{\boldsymbol{YZ}}(\boldsymbol{b},\boldsymbol{z})}{p_{\boldsymbol{WZ}}(\boldsymbol{b},\boldsymbol{z})} = \frac{p_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{b}|\boldsymbol{z})}{p_{\boldsymbol{W}}(\boldsymbol{b})} \tag{A.57}$$

$$= \exp\left[\sqrt{\delta}\sum_{i=0}^{n-1} z_i(b_{i+1}-b_i) - \frac{\delta}{2}\sum_{i=0}^{n-1} z_i^2(t_{i+1}-t_i)\right]. \tag{A.58}$$

Thus the Radon-Nikodym derivative can be established as

$$\frac{\mathrm{d}\mu_{YZ}}{\mathrm{d}\mu_{WZ}} = \exp\left[\sqrt{\delta}\int_0^T Z_t\ \mathrm{d}W_t - \frac{\delta}{2}\int_0^T Z_t^2\ \mathrm{d}t\right] \tag{A.59}$$

using the finite-dimensional likelihood ratios (A.58). It is clear that $\mu_{YZ} \ll \mu_{WZ}$.

For the case of a general finite-power process (not necessarily bounded) $\{Z_t\}$, a sequence of bounded uniformly stepwise processes which converge to the $\{Z_t\}$ in $L^2(\mathrm{d}t\,\mathrm{d}P)$ can be obtained. The Radon-Nikodym derivative (A.59) of the sequence of processes also converges. Absolutely continuity is preserved. Therefore, (A.59) holds for all such processes $\{Z_t\}$.

The derivative (A.59) can be re-written as

$$\frac{\mathrm{d}\mu_{YZ}}{\mathrm{d}\mu_{WZ}} = 1 + \sqrt{\delta}\int_0^T Z_t\ \mathrm{d}W_t + \frac{\delta}{2}\left[\left(\int_0^T Z_t\ \mathrm{d}W_t\right)^2 - \int_0^T Z_t^2\ \mathrm{d}t\right] + o(\delta). \tag{A.60}$$

By the independence of the processes $\{W_t\}$ and $\{Z_t\}$, the measure $\mu_{WZ} = \mu_W\mu_Z$. Thus integrating on the measure μ_Z gives

$$\frac{\mathrm{d}\mu_{Y}}{\mathrm{d}\mu_{W}} = 1 + \sqrt{\delta}\int_0^T \mathsf{E}Z_t\ \mathrm{d}W_t + \frac{\delta}{2}\left[\mathsf{E}_{\mu_Z}\left(\int_0^T Z_t\ \mathrm{d}W_t\right)^2 - \int_0^T \mathsf{E}Z_t^2\ \mathrm{d}t\right] + o(\delta). \tag{A.61}$$

Clearly, $\mu_Y \ll \mu_W$. Using (A.60), (A.61) and the chain rule (A.54), the Radon-Nikodym derivative Φ exists and is given by

$$\begin{aligned}\Phi &= 1 + \sqrt{\delta}\int_0^T Z_t - \mathsf{E}Z_t\ \mathrm{d}W_t + \frac{\delta}{2}\left[\left(\int_0^T Z_t\ \mathrm{d}W_t\right)^2 - \mathsf{E}_{\mu_Z}\left(\int_0^T Z_t\ \mathrm{d}W_t\right)^2\right.\\ &\qquad \left. -2\int_0^T \mathsf{E}Z_t\ \mathrm{d}W_t\int_0^T Z_t - \mathsf{E}Z_t\ \mathrm{d}W_t - \int_0^T Z_t^2\ \mathrm{d}t + \int_0^T \mathsf{E}Z_t^2\ \mathrm{d}t\right] + o(\delta) \qquad (\text{A.62})\\ &= 1 + \sqrt{\delta}\int_0^T Z_t - \mathsf{E}Z_t\ \mathrm{d}W_t + \frac{\delta}{2}\left[\left(\int_0^T Z_t - \mathsf{E}Z_t\ \mathrm{d}W_t\right)^2\right.\\ &\qquad \left. -\mathsf{E}_{\mu_Z}\left(\int_0^T Z_t - \mathsf{E}Z_t\ \mathrm{d}W_t\right)^2 - \int_0^T Z_t^2 - \mathsf{E}Z_t^2\ \mathrm{d}t\right] + o(\delta). \qquad (\text{A.63})\end{aligned}$$

Note that the mutual information is an expectation with respect to the measure μ_{YZ}. It can be written as

$$I\left(Z_0^T; Y_0^T\right) = \int \log\Phi'\ \mathrm{d}\mu_{YZ} \tag{A.64}$$

where Φ' is obtained from Φ (A.63) by substitute all occurrences of $\mathrm{d}W_t$ by $\mathrm{d}Y_t = \sqrt{\delta}\, Z_t + \mathrm{d}W_t$:

$$\begin{aligned}
\Phi' &= 1 + \sqrt{\delta} \int_0^T Z_t - \mathsf{E} Z_t \,\mathrm{d}Y_t + \frac{\delta}{2} \Bigg[\left(\int_0^T Z_t - \mathsf{E} Z_t \,\mathrm{d}Y_t \right)^2 \\
&\quad - \mathsf{E}_{\mu_Z} \left(\int_0^T Z_t - \mathsf{E} Z_t \,\mathrm{d}Y_t \right)^2 - \int_0^T Z_t^2 - \mathsf{E} Z_t^2 \,\mathrm{d}t \Bigg] + o(\delta) && \text{(A.65)} \\
&= 1 + \sqrt{\delta} \int_0^T Z_t - \mathsf{E} Z_t \,\mathrm{d}W_t + \frac{\delta}{2} \Bigg[\left(\int_0^T Z_t - \mathsf{E} Z_t \,\mathrm{d}W_t \right)^2 - \mathsf{E}_{\mu_Z} \left(\int_0^T Z_t - \mathsf{E} Z_t \,\mathrm{d}W_t \right)^2 \\
&\quad + \int_0^T (Z_t - \mathsf{E} Z_t)^2 \,\mathrm{d}t + \int_0^T \mathsf{E}(Z_t - \mathsf{E} Z_t)^2 \,\mathrm{d}t \Bigg] + o(\delta) && \text{(A.66)} \\
&= 1 + \sqrt{\delta} \int_0^T \tilde{Z}_t \,\mathrm{d}W_t + \frac{\delta}{2} \Bigg[\left(\int_0^T \tilde{Z}_t \,\mathrm{d}W_t \right)^2 \\
&\quad - \mathsf{E}_{\mu_Z} \left(\int_0^T \tilde{Z}_t \,\mathrm{d}W_t \right)^2 + \int_0^T \tilde{Z}_t^2 \,\mathrm{d}t + \int_0^T \mathsf{E} \tilde{Z}_t^2 \,\mathrm{d}t \Bigg] + o(\delta) && \text{(A.67)}
\end{aligned}$$

where $\tilde{Z}_t = Z_t - \mathsf{E} Z_t$. Hence

$$\log \Phi' = \sqrt{\delta} \int_0^T \tilde{Z}_t \,\mathrm{d}W_t + \frac{\delta}{2} \left[-\mathsf{E}_{\mu_Z} \left(\int_0^T \tilde{Z}_t \,\mathrm{d}W_t \right)^2 + \int_0^T \tilde{Z}_t^2 \,\mathrm{d}t + \int_0^T \mathsf{E} \tilde{Z}_t^2 \,\mathrm{d}t \right] + o(\delta). \tag{A.68}$$

Therefore, the mutual information is

$$\begin{aligned}
\mathsf{E} \log \Phi' &= \frac{\delta}{2} \left[-\mathsf{E} \left(\int_0^T \tilde{Z}_t \,\mathrm{d}W_t \right)^2 + 2 \int_0^T \mathsf{E} \tilde{Z}_t^2 \,\mathrm{d}t \right] + o(\delta) && \text{(A.69)} \\
&= \frac{\delta}{2} \left[- \int_0^T \mathsf{E} \tilde{Z}_t^2 \,\mathrm{d}t + 2 \int_0^T \mathsf{E} \tilde{Z}_t^2 \,\mathrm{d}t \right] + o(\delta) && \text{(A.70)} \\
&= \frac{\delta}{2} \int_0^T \mathsf{E} \tilde{Z}_t^2 \,\mathrm{d}t + o(\delta), && \text{(A.71)}
\end{aligned}$$

and the lemma is proved. ■

A.4 Asymptotic Joint Normality of $\{\boldsymbol{V}\}$

A proof of asymptotic joint normality in [96, Appendix B] using the Edgeworth expansion [66] of the probability density function is problematic since possible discrete nature of $\boldsymbol{V}$ may not allow a density function. A simple remedy is to use the cumulative distribution function (c.d.f.) instead. Odd order cumulants of $\boldsymbol{V}$ are all zero. The second- and fourth-order cumulants of $\boldsymbol{V}$ are given by

$$\kappa^{a,b} = \mathsf{E} \{ V_a V_b \mid \boldsymbol{\Gamma}, \underline{\boldsymbol{X}} \} = Q_{ab} \tag{A.72}$$

and

$$\begin{aligned}
\kappa^{a,b,c,d} &= \mathsf{E}\{V_aV_bV_cV_d \mid \mathbf{\Gamma}, \underline{\boldsymbol{X}}\} - \mathsf{E}\{V_aV_b \mid \mathbf{\Gamma}, \underline{\boldsymbol{X}}\}\ \mathsf{E}\{V_cV_d \mid \mathbf{\Gamma}, \underline{\boldsymbol{X}}\} \\
&\quad -\mathsf{E}\{V_aV_c \mid \mathbf{\Gamma}, \underline{\boldsymbol{X}}\}\ \mathsf{E}\{V_bV_d \mid \mathbf{\Gamma}, \underline{\boldsymbol{X}}\} \\
&\quad -\mathsf{E}\{V_aV_d \mid \mathbf{\Gamma}, \underline{\boldsymbol{X}}\}\ \mathsf{E}\{V_bV_c \mid \mathbf{\Gamma}, \underline{\boldsymbol{X}}\} \qquad \text{(A.73)} \\
&= -2K^{-2}\sum_{k=1}^{K} \mathsf{snr}_k^2 X_{ak}X_{bk}X_{ck}X_{dk}\mathsf{E}\left\{S_{nk}^4\right\} \qquad \text{(A.74)} \\
&= O(K^{-1}) \qquad \text{(A.75)}
\end{aligned}$$

for all $a, b, c, d = 0, \ldots, u$. Higher order cumulants are $O(K^{-2})$. Therefore, the joint c.d.f. of $\boldsymbol{V}$ allows an Edgeworth expansion,

$$F(\boldsymbol{v}) = F_0(\boldsymbol{v}) + \frac{1}{4!\,K}\sum_{a,b,c,d=0}^{u}\left(\frac{1}{K}\sum_{k=1}^{K}\mathsf{snr}_k^2 X_{ak}X_{bk}X_{ck}X_{dk}\right)\frac{\partial^4 F_0(\boldsymbol{v})}{\partial v_a \partial v_b \partial v_c \partial v_d} + O(K^{-2}) \qquad \text{(A.76)}$$

where F_0 is the c.d.f. of joint Gaussian variables with a covariance of $\boldsymbol{Q}$ [66]. In the limit of $K \to \infty$, the distribution F converges to the Gaussian distribution F_0.

A.5 Evaluation of $G^{(u)}$

From (3.111),

$$\begin{aligned}
\exp\left[G^{(u)}(\boldsymbol{Q})\right] &= \left(2\pi\sigma^2\right)^{-\frac{u}{2}} \mathsf{E}\left\{\int \exp\left[-\frac{1}{2}\left(y - \sqrt{\beta}\,\tilde{V}_0\right)^2\right]\right. \\
&\quad \left.\times \prod_{a=1}^{u} \exp\left[-\frac{1}{2\sigma^2}\left(y - \sqrt{\beta}\,\tilde{V}_a\right)^2\right]\frac{\mathrm{d}y}{\sqrt{2\pi}}\right|\boldsymbol{Q}\Bigg\} \qquad \text{(A.77)} \\
&= \left(2\pi\sigma^2\right)^{-\frac{u}{2}} \mathsf{E}\left\{\int \exp\left[-\frac{1}{2}\left(1 + \frac{u}{\sigma^2}\right)y^2 + \sqrt{\beta}\left(\tilde{V}_0 + \frac{1}{\sigma^2}\sum_{a=1}^{u}\tilde{V}_a\right)y\right.\right. \\
&\quad \left.\left. -\frac{1}{2}\beta\tilde{V}_0^2 - \frac{\beta}{2\sigma^2}\sum_{a=1}^{u}\tilde{V}_a^2\right]\frac{\mathrm{d}y}{\sqrt{2\pi}}\right|\boldsymbol{Q}\Bigg\} \qquad \text{(A.78)} \\
&= \left(2\pi\sigma^2\right)^{-\frac{u}{2}}\left(1 + \frac{u}{\sigma^2}\right)^{-\frac{1}{2}} \mathsf{E}\left\{\exp\left[\frac{\beta}{2\sigma^2(\sigma^2 + u)}\left(\sigma^2\tilde{V}_0 + \sum_{a=1}^{u}\tilde{V}_a\right)^2\right.\right. \\
&\quad \left.\left. -\frac{\beta}{2\sigma^2}\left(\sigma^2\tilde{V}_0^2 + \sum_{a=1}^{u}\tilde{V}_a^2\right)\right]\right|\boldsymbol{Q}\Bigg\} \qquad \text{(A.79)} \\
&= \left(2\pi\sigma^2\right)^{-\frac{u}{2}}\left(1 + \frac{u}{\sigma^2}\right)^{-\frac{1}{2}} \mathsf{E}\left\{\exp\left[-\frac{1}{2}\tilde{\boldsymbol{V}}^\top \mathbf{\Sigma}\tilde{\boldsymbol{V}}\right] \,\middle|\, \boldsymbol{Q}\right\} \qquad \text{(A.80)}
\end{aligned}$$

where $\mathbf{\Sigma}$ is a $(u+1) \times (u+1)$ matrix given by (3.120). Rewriting the expectation with respect to the Gaussian vector $\tilde{\boldsymbol{V}}$ as an integral, one finds that

$$\begin{aligned}\exp\left[G^{(u)}(\boldsymbol{Q})\right] &= \left(2\pi\sigma^2\right)^{-\frac{u}{2}}\left(1+\frac{u}{\sigma^2}\right)^{-\frac{1}{2}}(2\pi)^{-\frac{u+1}{2}}\left[\det\boldsymbol{Q}\right]^{-\frac{1}{2}} \\ &\quad\times\int\exp\left[\boldsymbol{v}^\top\left(\boldsymbol{Q}^{-1}+\mathbf{\Sigma}\right)\boldsymbol{v}\right]\,\mathrm{d}\boldsymbol{v} &\text{(A.81)}\\ &= \left(2\pi\sigma^2\right)^{-\frac{u}{2}}\left(1+\frac{u}{\sigma^2}\right)^{-\frac{1}{2}}\left[\det(\boldsymbol{I}+\mathbf{\Sigma}\boldsymbol{Q})\right]^{-\frac{1}{2}}, &\text{(A.82)}\end{aligned}$$

which is equivalent to (3.119).

Under the replica symmetry assumption (3.127), the determinant in (A.82) can be written as

$$\det(\boldsymbol{I}+\mathbf{\Sigma}\boldsymbol{Q}) = \frac{\sigma^2}{\sigma^2+u}\left[1+\frac{\beta}{\sigma^2}(p-q)\right]^{u-1}\left[1+\frac{\beta}{\sigma^2}(p-q)+\frac{u}{\sigma^2}(1+\beta(r-2m+q))\right] \tag{A.83}$$

and (3.128) follows.

A.6 Proof of Lemma 3.2

Proof: It is easy to see that

$$Z^{(u)}(\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0;h)\Big|_{h=0} = Z^u(\boldsymbol{Y},\boldsymbol{S}). \tag{A.84}$$

By taking the derivative and letting $h=0$, the right hand side of (3.158) is

$$\frac{1}{K}\lim_{u\to 0}\mathsf{E}\left\{\mathsf{E}_q\left\{f(\boldsymbol{X}_0,\underline{\boldsymbol{X}}'_a)\prod_{a=1}^{u}q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{X}'_a,\boldsymbol{S})\;\middle|\;\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0\right\}\right\}, \tag{A.85}$$

where $\underline{\boldsymbol{X}}'_a$ has the same statistics as $\underline{\boldsymbol{X}}_a$ (i.e., contains i.i.d. entries with distribution q_X) but independent of all other random variables. Also note that

$$q_{\underline{\boldsymbol{X}}_a|\boldsymbol{Y},\boldsymbol{S}}(\underline{\boldsymbol{X}}_a\,|\,\boldsymbol{Y},\boldsymbol{S}) = Z^{-u}(\boldsymbol{Y},\boldsymbol{S})\,q_{\underline{\boldsymbol{X}}_a}(\underline{\boldsymbol{X}}_a)\prod_{a=1}^{u}q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{Y}\,|\,\boldsymbol{X}_a,\boldsymbol{S}). \tag{A.86}$$

One can change the expectation over the replicas $\underline{\boldsymbol{X}}'_a$ independent of $(\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0)$ to an expectation over $\underline{\boldsymbol{X}}_a$ conditioned on $(\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0)$. Hence (A.85) can be further written as

$$\begin{aligned}&\frac{1}{K}\lim_{u\to 0}\mathsf{E}\left\{\mathsf{E}\left\{f(\boldsymbol{X}_0,\underline{\boldsymbol{X}}_a)|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0\right\}\,Z^u(\boldsymbol{Y},\boldsymbol{S})\right\}\\ &= \frac{1}{K}\mathsf{E}\left\{\mathsf{E}\left\{f(\boldsymbol{X}_0,\underline{\boldsymbol{X}}_a)|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0\right\}\right\} &\text{(A.87)}\\ &= \frac{1}{K}\mathsf{E}\left\{f(\boldsymbol{X}_0,\underline{\boldsymbol{X}}_a)\right\} &\text{(A.88)}\end{aligned}$$

where $Z^u(\boldsymbol{Y},\boldsymbol{S})$ can be dropped as $u\to 0$ in (A.87) since the conditional expectation in (A.87) is not dependent on u by the assumption in the lemma. ∎

A.7 Proof of Lemma 2.6

Proof: Let $Y = \sqrt{\mathsf{snr}}\, g(X) + N$. Since

$$0 \leq H(X) - I(X;Y) = H(X|Y), \tag{A.89}$$

it suffices to show that the uncertainty about X given Y vanishes as $\mathsf{snr} \to \infty$:

$$\lim_{\mathsf{snr}\to\infty} H(X|Y) = 0. \tag{A.90}$$

Assume first that X takes a finite number ($m < \infty$) of distinct values. Given Y, let $\widehat{X}_m$ be the decision for X that achieves the minimum probability of error, which is denoted by p. Then

$$H(X|Y) \leq H(X|\widehat{X}) \leq p \log(m-1) + H_2(p), \tag{A.91}$$

where $H_2(\cdot)$ stands for the binary entropy function, and the second inequality is due to Fano [17]. Since $p \to 0$ as $\mathsf{snr} \to \infty$, the right hand side of (A.91) vanishes.

In case X takes a countable number of values and that $H(X) < \infty$, for every natural number m, let U_m be an indicator which takes the value of 1 if X takes one of the m most likely values and 0 otherwise. Let $\widehat{X}_m$ be the function of Y which minimizes $\mathsf{P}\left\{X \neq \widehat{X}_m | U_m = 1\right\}$. Then for every m,

$$\begin{aligned}
H(X|Y) &\leq H(X|\widehat{X}_m) && \text{(A.92)}\\
&= H(X, U_m|\widehat{X}_m) && \text{(A.93)}\\
&= H(X|\widehat{X}_m, U_m) + H(U_m|\widehat{X}) && \text{(A.94)}\\
&\leq \mathsf{P}\{U_m = 1\} H(X|\widehat{X}, U_m = 1)\\
&\quad + \mathsf{P}\{U_m = 0\} H(X|\widehat{X}, U_m = 0) + H(U_m) && \text{(A.95)}\\
&\leq \mathsf{P}\{U_m = 1\} H(X|\widehat{X}, U_m = 1)\\
&\quad + \mathsf{P}\{U_m = 0\} H(X) + H_2(\mathsf{P}\{U_m = 0\}). && \text{(A.96)}
\end{aligned}$$

Conditioned on $U_m = 1$, the probability of error $\mathsf{P}\left\{X \neq \widehat{X}_m | U_m = 1\right\}$ vanishes as $\mathsf{snr} \to \infty$ by Fano's inequality. Therefore, for every m,

$$\lim_{\mathsf{snr}\to\infty} H(X|Y) \leq \mathsf{P}\{U_m = 0\} H(X) + H_2(\mathsf{P}\{U_m = 0\}). \tag{A.97}$$

The limit in (A.97) must be 0 since $\lim_{m\to\infty} \mathsf{P}\{U_m = 0\} = 0$. Thus (A.90) is also proved in this case.

In case $H(X) = \infty$, $H(X|U_m = 1) \to \infty$ as $m \to \infty$. For every m, the mutual information (expressed in the form of a divergence) converges:

$$\lim_{\mathsf{snr}\to\infty} \mathsf{D}\left(P_{Y|X,U_m=1} \| P_{Y|U_m=1} | P_{X|U_m=1}\right) = H(X|U_m = 1). \tag{A.98}$$

Therefore, the mutual information increases without bound as $\mathsf{snr} \to \infty$ by also noticing

$$I(X;Y) \geq I(X;Y|U_m) \geq \mathsf{P}\{U_m = 1\} \mathsf{D}\left(P_{Y|X,U_m=1} \| P_{Y|U_m=1} | P_{X|U_m=1}\right). \tag{A.99}$$

We have thus proved (2.196) in all cases. ■

References

[1] B. D. O. Anderson and S. Chirarattananon, "Smoothing as an improvement on filtering: A universal bound," *Electronics Letters*, vol. 7, pp. 524–525, Sept. 1971.

[2] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Commun.*, vol. 48, pp. 1347–1359, Aug. 2000.

[3] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review," *Statistica Sinica*, vol. 9, pp. 611–677, July 1999.

[4] I. Bar-David and S. Shamai, "On information transfer by envelop constrained signals over the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 34, pp. 371–379, May 1988.

[5] J. Bertoin, *Lévy Processes.* Cambridge University Press, 1996.

[6] R. E. Blahut, *Principles and Practice of Information Theory.* Reading, Mass.: Addison-Wesley, 1987.

[7] B. Z. Bobrovsky and M. Zakai, "A lower bound on the estimation error for certain diffusion processes," *IEEE Trans. Inform. Theory*, vol. 22, pp. 45–52, Jan. 1976.

[8] B. Z. Bobrovsky and M. Zakai, "Asymptotic a priori estimates for the error in the nonlinear filtering problem," *IEEE Trans. Inform. Theory*, vol. 28, pp. 371–376, March 1982.

[9] R. S. Bucy, "Information and filtering," *Information Sciences*, vol. 18, pp. 179–187, 1979.

[10] N. Chayat and S. Shamai, "Bounds on the capacity of intertransition-time-restricted binary signaling over an AWGN channel," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1992–2006, Sept. 1999.

[11] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *IEEE Trans. Inform. Theory*, vol. 50, pp. 245–258, Feb. 2004.

[12] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding— Part I: Equalization results, Part II: Coding results," *IEEE Trans. Commun.*, vol. 43, pp. 2582–2604, Oct. 1995.

[13] J. M. Cioffi and G. D. Forney, Jr., *Communications, Computation, Control and Signal Processing*, ch. 4, Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise. Kluwer Academic Publishers, 1997.

[14] F. Comets, "The martingale method for mean-field disordered systems at high temperature," in *Mathematical Aspects of Spin Glasses and Neural Networks*, (A. Bovier and P. Picco, eds.), pp. 91–113, Birkhäuser, 1998.

[15] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[16] M. H. M. Costa, "A new entropy power inequality," *IEEE Trans. Inform. Theory*, vol. 31, pp. 751–760, Nov. 1985.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York : Wiley, 1991.

[18] J. L. Doob, *Stochastic Processes.* New York: Wiley, 1990.

[19] V. Dotsenko, *The Theory of Spin Glasses and Neural Networks.* World Scientific, 1994.

[20] T. E. Duncan, "Evaluation of likelihood functions," *Information and Control*, vol. 13, pp. 62–74, 1968.

[21] T. E. Duncan, "On the calculation of mutual information," *SIAM Journal of Applied Mathematics*, vol. 19, pp. 215–220, July 1970.

[22] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, pp. 965–974, May 1975.

[23] Y. C. Eldar and A. M. Chan, "On the asymptotic performance of the decorrelator," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2309–2313, Sept. 2003.

[24] H. ElGamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 968–985, June 2004.

[25] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics.* Vol. 271 of *A series of comprehensive studies in mathematics*, Springer-Verlag, 1985.

[26] U. Erez and R. Zamir, "Lattice decoding can achieve $\frac{1}{2}\log(1+SNR)$ on the AWGN channel using nested codes," *IEEE Trans. Inform. Theory*, 2001. submitted.

[27] R. Esposito, "On a relation between detection and estimation in decision theory," *Information and Control*, vol. 12, pp. 116–120, 1968.

[28] W. Feller, *An Introduction to Probability Theory and Its Applications.* Vol. II, John Wiley & Sons, Inc., 2nd ed., 1971.

[29] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[30] R. G. Gallager, *Information Theory and Reliable Communication.* New York, Wiley, 1968.

[31] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proceedings 2002 IEEE Infocom*, New York, 2002.

[32] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2832–2836, Nov. 1998.

[33] T. Guess and M. K. Varanasi, "An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels," *IEEE Trans. Inform. Theory*, 2004. To appear.

[34] D. Guo, S. Shamai, and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," in *to be presented at IEEE Information Theory Workshop*, San Antonio, TX, USA, Oct. 2004.

[35] D. Guo, S. Shamai, and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *in preparation for IEEE Trans. on Inform. Theory*, 2004.

[36] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *submitted to IEEE Trans. on Inform. Theory*, 2004.

[37] D. Guo, S. Shamai, and S. Verdú, "Mutual information and MMSE in Gaussian channels," in *Proceedings 2004 IEEE International Symposium on Information Theory*, p. 347, Chicago, IL, USA, June 2004.

[38] D. Guo and S. Verdú, "Minimum probability of error of many-user CDMA without power control," in *Proceedings 2002 IEEE International Symposium on Information Theory*, p. 188, Lausanne, Switzerland, June 2002.

[39] D. Guo and S. Verdú, "Multiuser detection and statistical mechanics," in *Communications, Information and Network Security*, (V. Bhargava, H. V. Poor, V. Tarokh, and S. Yoon, eds.), ch. 13, pp. 229–277, Kluwer Academic Publishers, 2002.

[40] D. Guo and S. Verdú, "Decoupling of CDMA multiuser detection via the replica method," in *Proceedings 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2003.

[41] D. Guo and S. Verdú, "Replica analysis of CDMA spectral efficiency," in *Proceedings 2003 IEEE Information Theory Workshop*, Paris, France, March 2003.

[42] D. Guo and S. Verdú, "Spectral efficiency of large-system CDMA via statistical physics," in *Conference on Information Sciences and Systems*, Baltimore, MD, USA, March 2003.

[43] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, 2004. Submitted for publication.

[44] D. Guo, S. Verdú, and L. K. Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3080–3095, Dec. 2002.

[45] D. Guo, S. Shamai, and S. Verdú, "On information and estimation," Tech. Rep., Princeton University, 2004.

[46] C. P. Hatsell and L. W. Nolte, "Some geometric properties of the likelihood ratio," *IEEE Trans. Inform. Theory*, vol. 17, pp. 616–618, 1971.

[47] J. Hubbard, "Calculation of partition functions," *Physics Review Letters*, vol. 3, pp. 77–78, July 1959.

[48] A. G. Jaffer and S. C. Gupta, "On relations between detection and estimation of discrete time processes," *Information and Control*, vol. 20, pp. 46–54, 1972.

[49] Y. Kabashima and D. Saad, "Statistical mechanics of error-correcting codes," *Europhysics Letters*, vol. 45, no. 1, pp. 97–103, 1999.

[50] Y. Kabashima and D. Saad, "Statistical mechanics of low-density parity-check codes," *Journal of Physics A: Mathematical and General*, vol. 37, pp. R1–R43, 2004.

[51] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *Journal of Physics A: Mathematical and General*, vol. 36, pp. 11111–11121, 2003.

[52] T. T. Kadota, M. Zakai, and J. Ziv, "The capacity of a continuous memoryless channel with feedback," *IEEE Trans. Inform. Theory*, vol. 17, pp. 372–378, July 1971.

[53] T. T. Kadota, M. Zakai, and J. Ziv, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Trans. Inform. Theory*, vol. 17, pp. 368–371, July 1971.

[54] T. Kailath, "Adaptive matched filters," in *Mathematical Optimization Techniques*, (R. Bellman, ed.), Univ. of California Press, 1963.

[55] T. Kailath, "A note on least squares estimates from likelihood ratios," *Information and Control*, vol. 13, pp. 534–540, 1968.

[56] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 15, pp. 350–361, May 1969.

[57] T. Kailath, "A further note on a general likelihood formula for random signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 16, pp. 393–396, July 1970.

[58] T. Kailath, "The innovations approach to detection and estimation theory," *Proceedings of the IEEE*, vol. 58, pp. 680–695, May 1970.

[59] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2230–2259, Oct. 1998.

[60] A. N. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IEEE Trans. Inform. Theory*, vol. 2, pp. 102–108, Sept. 1956.

[61] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118–1134, May 2002.

[62] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes I: General Theory.* Springer, 2nd ed., 2001.

[63] R. S. Liptser, "Optimal encoding and decoding for transmission of a Gaussian Markov signal in a noiseless-feedback channel," *Problemy Peredachi Informatsii*, vol. 10, pp. 3–15, October-December 1974.

[64] E. Mayer-Wolf and M. Zakai, "On a formula relating the Shannon information to the Fisher information for the filtering problem," in *Lecture Notes in Control and Information Sciences*, pp. 164–171, Springer-Verlag, New York, 1983.

[65] R. R. Mazumdar and A. Bagchi, "On the relation between filter maps and correction factors in likelihood ratios," *IEEE Trans. Inform. Theory*, vol. 41, pp. 833–836, May 1995.

[66] P. McCullagh, *Tensor Methods in Statistics*. Chapman and Hall, 1987.

[67] F. Meshkati, D. Guo, H. V. Poor, S. Schwartz, and N. B. Mandayam, "A unified approach to power control for multiuser detectors," in *Proceedings 2nd International Workshop on Signal Processing for Wireless Communications*, London, England, June 2004.

[68] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. World Scientific, 1987.

[69] A. Montanari and N. Sourlas, "The statistical mechanics of turbo codes," *European Physical Journal B*, vol. 18, no. 1, pp. 107–119, 2000.

[70] R. R. Müller, "Multiuser receivers for randomly spread signals: Fundamental limits with and without decision-feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 268–283, Jan. 2001.

[71] R. R. Müller, "On channel capacity, uncoded error probability, ML-detection and spin glasses," in *Proceedings Workshop on Concepts in Information Theory*, pp. 79–81, Breisach, Germany, June 2002.

[72] R. R. Müller and W. Gerstacker, "On the capacity loss due to separation of detection and decoding in large CDMA systems," in *Proceedings IEEE Information Theory Workshop*, p. 222, Bangalore, India, Oct. 2002.

[73] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction. Number 111 in International Series of Monographs on Physics*, Oxford University Press, 2001.

[74] D. O. North, "An analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems," Tech. Rep. PTR-6C, RCA Lab., June 1943. (Reprinted in *Proc. IEEE*, vol. 51, pp. 1016–1027, July 1963.).

[75] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden-Day, 1964.

[76] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, 1994.

[77] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1567–1580, Aug. 2004.

[78] R. Price, "Optimum detection of random signals in noise, with application to scatter-multipath communication–I," *IRE Trans. Inform. Theory*, vol. 2, pp. 125–135, Dec. 1956.

[79] R. I. Price and J. P. E. Green, "A communication technique for multipath channels," *Proc. IRE*, vol. 46, pp. 555–570, March 1958.

[80] P. Rapajic, M. Honig, and G. Woodward, "Multiuser decision-feedback detection: Performance bounds and adaptive algorithms," in *Proceedings of 1998 IEEE International Symposium on Information Theory*, p. 34, Cambridge, MA USA, Aug. 1998.

[81] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Commun.*, vol. 48, pp. 1245–1248, Aug. 2000.

[82] G. J. Rodgers, "Comment: Spectral analysis of random matrices using the replica method," *Statistica Sinica*, vol. 9, pp. 662–667, 1999.

[83] H. L. Royden, *Real Analysis.* Macmillan, 1988.

[84] F. C. Schweppe, "Evaluation of likelihood functions for Gaussian signals," *IEEE Trans. Inform. Theory*, vol. 11, pp. 61–70, Jan. 1965.

[85] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1302–1327, May 2001.

[86] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[87] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, 1949.

[88] A. B. Shmelev, "Relationship between optimal filtering and interpolation of random signals in observations against a background of white Gaussian noise," *Radiotekhnika I Elektronika*, no. 1, pp. 86–89, 1985.

[89] N. Sourlas, "Spin-glass models as error-correcting codes," *Nature*, vol. 339, pp. 693–695, June 1989.

[90] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, pp. 101–112, 1959.

[91] Y. Steinberg, B. Z. Bobrovsky, and Z. Schuss, "Fixed-point smoothing of scalar diffusions II: The error of the optimal smoother," *SIAM J. Appl. Math*, vol. 61, pp. 1431–1444, 2001.

[92] R. L. Stratonovich and Y. G. Sosulin, "Optimal detection of a Markov process in noise," *Engrg. Cybernetics*, vol. 6, pp. 7–19, 1964.

[93] T. Tanaka, "Analysis of bit error probability of direct-sequence CDMA multiuser demodulators," in *Advances in Neural Information Processing Systems*, (T. K. Leen et al., ed.), pp. 315–321, The MIT Press, 2001.

[94] T. Tanaka, "Average-case analysis of multiuser detectors," in *Proceedings 2001 IEEE International Symposium on Information Theory*, p. 287, Washington, D.C., June 2001.

[95] T. Tanaka, "Statistical mechanics of CDMA multiuser demodulation," *Europhysics Letters*, vol. 54, no. 4, pp. 540–546, 2001.

[96] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[97] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, Nov./Dec. 1999.

[98] D. N. C. Tse and S. Verdú, "Optimum asymptotic multiuser efficiency for randomly spread CDMA," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2718–2722, Nov. 2000.

[99] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, March 1999.

[100] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.

[101] J. van Mourik, D. Saad, and Y. Kabashima, "Critical noise levels for LPDC decoding," *Physics Review E*, vol. 66, no. 2, pp. 026705:1–8, 2002.

[102] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 1405–1409, Monterey, CA, Nov. 1997.

[103] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proceedings 24th Allerton Conference on Communication, Control and Computing*, pp. 1025–1034, Oct. 1986.

[104] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

[105] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, March 1999.

[106] S. Verdú, "Minimum probability of error for asynchronous multiple access communication systems," in *Proceedings of IEEE Military Communications Conference*, pp. 213–219, Nov. 1983.

[107] S. Verdú, *Optimum multiuser signal detection.* PhD thesis, Dept. of Electrical Engineering, University of Illinois at Urbana-Champaign, Aug. 1984. Technical Report T-151, Coordinated Science Laboratory.

[108] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 32, pp. 85–96, Jan. 1986.

[109] S. Verdú, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Inform. Theory*, vol. 34, pp. 890–897, Sept. 1986.

[110] S. Verdú, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, no. 3, pp. 303–312, 1989.

[111] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[112] S. Verdú, *Multiuser Detection.* Cambridge University Press, 1998.

[113] C.-K. Wen and K.-K. Wong, "Asymptotic spectral efficiency of spatially correlated MIMO multiple-access channels with arbitrary signaling inputs for joint and separate decoding," in *Proceedings 2004 IEEE International Symposium on Information Theory*, p. 282, Chicago, IL, USA, June 2004.

[114] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications.* New York: Wiley, 1949. (Originally published in Feb. 1942, as a classified Nat. Defense Res. Council rep.).

[115] W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," *J. SIAM Control, Ser. A*, vol. 2, no. 3, pp. 347–369, 1965.

[116] Y.-C. Yao, "Estimation of noisy telegraph processes: Nonlinear filtering versus nonlinear smoothing," *IEEE Trans. Inform. Theory*, vol. 31, pp. 444–446, May 1985.

[117] J. Yedidia, W. Freeman, and Y. Weiss, "Generalized belief propagation," in *Advances in Neural Information Processing Systems 13*, pp. 689–695, MIT Press, 2001.

[118] M. C. Yovits and J. L. Jackson, "Linear filter optimization with game theory considerations," *Proc. IRE*, vol. 43, no. 3, p. 376, 1955.

[119] M. Zakai and J. Ziv, "Lower and upper bounds on the optimal filtering error of certain diffusion processes," *IEEE Trans. Inform. Theory*, vol. 18, pp. 325–331, May 1972.

[120] M. Zakai, "On mutual information, likelihood-ratios and estimation error for the additive Gaussian channel," *preprint*, 2004.

Index