Two Multi-Terminal Communication Problems: Distributed Estimation and Source-Channel Broadcast

Vinod Prabhakaran

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2007-154
http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-154.html

December 17, 2007

Two Multi-Terminal Communication Problems: Distributed Estimation and Source-Channel Broadcast

by

Vinod Malathi Prabhakaran

B.Tech. (University of Kerala, Thiruvananthapuram, India) 1999
M.Eng. (Indian Institute of Science, Bangalore, India) 2001

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering-Electrical Engineering and Computer Sciences

and the Designated Emphasis

in

Communication, Computation, and Statistics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Kannan Ramchandran, Co-chair
Professor David Tse, Co-chair
Professor Venkatachalam Anantharam
Professor Bin Yu

Fall 2007

The dissertation of Vinod Malathi Prabhakaran is approved:

Co-chair Date

Co-chair Date

Date

Date

University of California, Berkeley

Fall 2007

Two Multi-Terminal Communication Problems: Distributed Estimation and Source-Channel Broadcast

by

Vinod Malathi Prabhakaran

Abstract

Two Multi-Terminal Communication Problems: Distributed Estimation and Source-Channel Broadcast

by

Vinod Malathi Prabhakaran

Doctor of Philosophy in Engineering-Electrical Engineering and Computer Sciences

Designated Emphasis in Communication, Computation, and Statistics

University of California, Berkeley

Professors Kannan Ramchandran and David Tse, Co-chairs

A major success of information theory is a more or less satisfactory understanding of point-to-point communication (where a single source of information wants to convey this information to a single sink). However, many questions remain open when there is more than one source or sink of information. This dissertation addresses two classes of such problems.

The first set of problems is motivated by wireless sensor networks where a large number of sensors collect potentially noisy data about an underlying physical phenomenon and communicate a compressed form of the data to a fusion centre which attempts to reconstruct some aspects of the original source of data. One model for

studying such problems is the so-called *CEO model* proposed by previous researchers where the sensors communicate only to the fusion centre and not with one another. We derive the optimal trade-off between the quality of reconstruction at the fusion centre and the communication burden on the sensors under this model for the Gaussian case. We demonstrate the robustness of the scheme to sensor failures. We also study improved models to explore the potential for co-operation between the sensors through interactions and drive bounds on performance.

The motivation for our second set of problems is the migration of audio and video broadcasting from analog transmission schemes to the more spectrum efficient digital schemes. We first address the question of this migration itself and derive the optimal trade-off between the qualities at which the new digital and the legacy analog receivers can be served simultaneously without consuming any additional spectrum. Then, for an all-digital system, we design efficient transmission schemes which can offer differentiated levels of service in terms of fidelity of source reconstruction to users based on the quality of their channels from the broadcast station.

Professor Kannan Ramchandran
Dissertation Committee Co-chair

Professor David Tse
Dissertation Committee Co-chair

To my parents

and

my teachers.

Contents

List of Figures

Acknowledgments

I wish to acknowledge my deep debt of gratitude to the co-chairs of my dissertation committee, Professors Kannan Ramchandran and David Tse. They exposed me to a wide range of ideas, and helped me at every stage of my doctoral research by listening to me patiently and asking the right questions. Their vast intellects, limitless patience, and exceptional teaching and mentoring abilities will continue to inspire me.

I would like to thank the other members of my dissertation committee Professors Venkat Anantharam and Bin Yu who also served on my qualifying exam committee and gave very valuable feedback on my research. It is a pleasure to acknowledge the many discussions I had with Rohit Puri. Part II of this dissertation are the fruits of my collaboration with him. I also benefited from discussions with Sandeep Pradhan, Prakash Ishwar, Pramod Viswanath, and Aaron Wagner at different stages of my research.

I would also like to thank all my friends, in particular, Murali Ambati, Mariko Azuma, Christophe Choumert, Pulkit Grover, Animesh Kumar, Mani Narayanan, Rahul Tandra, Asako Toda, June Wang, Wei Wang, Darren Yamaguchi, Chuohao Yeo, and everybody at the Wireless Foundations, who made my life in Berkeley tremendously enjoyable.

My brother and my parents have been a constant source of encouragement without whose love and support I would not have been able to make it this far. I am deeply indebted to them.

Chapter 1

Introduction

The short history of information theory since it was founded by Shannon in 1948 is marked by many successes. Remarkably, many of the more profound and impactful of these successes can found be in Shannon's seminal paper [41] which established the field. Arguably, the theory of point-to-point communication (where a single source of information wants to convey this information to a single sink) was covered more or less single-handedly by Shannon himself in a sequence of papers [41–43]. However, many questions remain open when there is more than one source or sink of information. We consider two classes of such problems. In the first class, the multiplicity is in the sources of information with just one sink consuming this information, whereas in the second class of problems there is a single source of information, but multiple sinks are interested in receiving the information.

We are motivated by wireless sensor networks in studying our first set of problems.

Wireless sensor networks raise many challenges to a system designer. Typically, the sensing devices make low-quality measurements, operate with limited communication resources (e.g. power, bandwidth), and suffer from poor reliability (sensor-failures, power-cycling). The fact that the cost of communication is typically much higher than the cost of processing [21] calls for the use of distributed or locally-collaborative designs. On the positive side, sensors are typically inexpensive devices which allows for a large-scale deployment resulting in a high degree of redundancy in the measurements which can be exploited. There have been many noteworthy efforts at building protocols and infrastructure for sensor networks [2, 19, 27]. Our interest here is in understanding the fundamental limits of performance of and optimal designs for such networks.

One of the earliest attempts at studying such problems appears to be by Tenny and Sandell in 1981 [47]. The approach is called distributed detection (see [9, 49, 53] and references therein). A simple form of the model studied involves many sensors making an observation each and helping a fusion centre decide among many hypotheses about the environment in which the sensors reside by sending it messages. The main design question is: what should the sensors send to the fusion centre if there are constraints on how much the sensors can send? This is in contrast to a centralised detection scheme where the fusion centre will have direct access to all the observations. Many authors have addressed various versions of this model (see references above).

The approach we will follow is information theoretic. We treat the networked sensing problem as a multi-terminal rate distortion question. This was introduced by Berger, Zhang, and Viswanathan in a paper titled "The CEO Problem" [8]. The CEO in the title stands for the "chief estimation officer" who serves as the fusion centre mentioned above. The sensors (which are termed "agents" in the paper) make a sequence of observations and send a message each to the CEO whose objective is to reconstruct an underlying sequence which influences the observations of the agents, but is not necessarily visible to any of the agents. The objective of the CEO problem is to find the set of sizes (or rates) of the messages that the sensors need to send such that at least a certain quality of reproduction is achieved by the CEO. As pointed out earlier, typically, in wireless sensor networks, the cost of communication is larger than the cost of processing at the sensor. This makes the question posed by the CEO problem a relevant one for wireless sensor networks. The key difference with respect to distributed detection theory is that there are no delay constraints on the operation of the CEO. This is in contrast to the distributed detection problem where the delay constraint on the fusion centre can be thought of as being less than one sampling interval of the sensors. Thus the results here are more relevant for sensing applications where the delay constraints are not stringent (e.g., habitat monitoring [45], monitoring of the health of structures such as bridges [29]. See [39] for many more).

The second set of problems we address is motivated by the migration of audio and video broadcasting from analog transmission schemes to the more spectrum efficient digital schemes. This migration is stimulated as much, if not more, by a desire to free up spectrum for use by other services such as the future generations of wireless data service as by a desire to deliver improved reception and additional services such as interactive TV (in which, for instance, the viewers can choose the camera angle). The questions we address are two fold. First, we consider the process of this migration itself. Switching over the transmission format from analog to digital will render the large number of analog receivers useless. The current solution is to *simulcast* the content in the existing analog format and in the digital format in addition on a separate part of the spectrum. This is however redundant. We ask what is the optimal trade-off of quality we can offer to the legacy analog receivers and the digital receivers without occupying any more of the spectrum than the existing analog system.

In a broadcast system, different users experience different channel qualities. Offering differentiated levels of service in terms of the quality of source reproduction which the users can subscribe to based on the quality of their channels has been a long standing challenge. Techniques like scalable coding with unequal error protection are practical attempts at achieving this goal [48]. We address this problem under a simple, but relevant model and attempt to find efficient schemes for simultaneously satisfying users with different channel qualities.

1.1 Outline and contributions

In Part I we discuss the CEO problem and some variations motivated by sensor networks. In chapter 3, we first derive for the Gaussian version of the CEO problem, the optimal set (called the rate-region) of rates at which the sensors may operate in order to achieve a prescribed quality of estimation under the mean-squared error (MSE) distortion criterion. This result was obtained independently of Oohama who also published the same results [35, 36]. We then address the question of robustness of the optimal strategy to random sensor failures. We derive the optimal qualities of estimation at the CEO under different sensor-failure scenarios if the system is constrained to operate optimally when a certain minimum number of sensors are active.

In a practical sensor network deployment, it may be possible for the sensors to co-operate with one another, and for the central unit to provide feedback to the sensors. In Chapter 4, we extend the CEO model to consider different forms of co-operation and feedback. We derive an upper-bound on the total rate of all sensors, and use it to show that the jointly Gaussian distribution is a least-friendly distribution to co-operation in the sense that no reduction in the total rate of communication can be achieved over the CEO setting through feedback and co-operation.

In Part II, we consider the problem of sending coloured Gaussian sources over Gaussian channels to a heterogeneous set of receivers. In Chapter 6, motivated by the practical application of seamless “in-band” digital upgrade of legacy analog transmission systems, we study the problem of transmitting a Gaussian source with memory to a digital and a linear-analog receiver, over an arbitrarily coloured, non-degraded Gaussian broadcast channel. We characterise the set of all achievable MSE distortion pairs at the two receivers given a power constraint at the transmitter. Further, we show a constructive hybrid uncoded-coded scheme consisting of the cascade of source coding with side information and channel coding with side information systems that can achieve the entire power-distortion region associated with the problem. An interesting operating point in this region is one where the digital receiver obtains the classical point-to-point optimal quality and the analog receiver attains the best possible simultaneously achievable distortion.

We consider the problem of transmission of information sources with memory over a broadcast channel in Chapter 7. Using a parallel Gaussian source model and an additive white Gaussian noise broadcast channel model, we consider the problem of information source transmission with the goal of attaining trade-offs in MSE distortion achieved by individual receivers under a transmission power constraint. We propose a hybrid constructive framework consisting of building blocks derived from a number of tools from multi-user information theory – source coding with side information, channel coding with side information, successive refinement source coding and superposition broadcast channel coding. When specialised to the case of a white source with bandwidth mismatched to the channel bandwidth which has been studied, our scheme performs as well or in some cases better than all currently known strategies.

Part I

The CEO Problem

Chapter 2

Introduction

A canonical problem in the area of sensor networks is that of estimating a physical process (e.g. temperature) from noisy observations made by a large number of unreliable, resource-constrained sensors. The CEO problem is an information theoretic abstraction of this remote estimation problem. It models a group of say L sensors who observe independently corrupted versions of an underlying hidden memoryless source which a central unit (or the "CEO") is interested in estimating. The sensors encode long blocks of their observations into bits without co-operating with one another and communicate the encoded bits to the CEO over rate-constrained noiseless channels. We are interested in characterising the set of bitrates (L-tuples) of the channels which can support a desired fidelity at which the source is reproduced at the decoder. We will call this set the *rate region* for the specified fidelity. In the CEO problem, no constraints are placed on the delay at which the CEO reproduces the source.

The most basic result of this type is due to Shannon [41] who showed what the most parsimonious representation of an independent and identically distributed (i.i.d.) sequence is. His result states that an i.i.d. sequence[1] $X(1), X(2), \ldots, X(n)$ where each random variable has a distribution p_X over an alphabet $\mathcal{X}$ can be compressed losslessly to its *entropy* $H(X) = \sum_x -p_X(x)\log(p_X(x))$ bits per source symbol when n is large, and this is also the lowest bit rate possible. Shannon [43] generalised this to the case of lossy compression. If the source decoder produces a reconstruction of $\hat{X}(1), \ldots, \hat{X}(n)$ in a reconstruction alphabet $\hat{\mathcal{X}}$, and the distortion of this reconstructed sequence is measured using a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}_+$, the objective is to satisfy

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[d(X(k), \hat{X}(k))\right] \leq D,$$

where D is the target distortion. Shannon showed that the least possible rate of compression required to achieve this objective is

$$R(D) \stackrel{\text{def}}{=} \inf_{p_{\hat{X}|X}} I(X; \hat{X}),$$

where the infimum is over all conditional distributions $p_{\hat{X}|X}$ which satisfy $\mathbb{E}[d(X, \hat{X})] \leq D$. Here $I(X; \hat{X})$ is the *mutual information* between X and $\hat{X}$ defined as $I(X; \hat{X}) = H(X) + H(\hat{X}) - H(X, \hat{X})$.

Unlike in the above settings, in sensing networks, typically there is more than one sensor making observations and often they are physically separated. This can be modelled by introducing multiple encoders into Shannon's model. Slepian and

[1] We will also call this a *memoryless source* X.

Wolf [44] considered this generalisation in the lossless setting. They showed that for two dependent discrete memoryless sources X_1 and X_2 observed by two physically separated encoders can be reconstructed losslessly at the decoder if the encoder observing X_1 compresses at a bit rate R_1 and the encoder observing X_2 compresses at rate R_2 provided (R_1, R_2) satisfy the following.

$$
\begin{aligned}
R_1 &\geq H(X_1|X_2), \\
R_2 &\geq H(X_2|X_1), \\
R_1 + R_2 &\geq H(X_1, X_2),
\end{aligned}
$$

where the conditional entropy is defined as $H(X_1|X_2) = H(X_1, X_2) - H(X_2)$. This set of rates which is the largest set of rates under which we can achieve the objective of the problem (which in this case is the lossless recovery of the sources at the decoder) is called the *rate region* of the problem. The lossy compression version of this problem setting where the objective of the decoder is to reconstruct the source(s) under a distortion criterion is called the multi-terminal (lossy) source coding problem [5]. The problem in general remains open, and only inner and outer bounds are available for the rate region of most versions of the problem. One noteworthy instance of the problem which has been fully resolved is when the encoder has a memoryless source X_1 over an alphabet $\mathcal{X}_1$ and the decoder has a dependent memoryless source X_2 over an alphabet $\mathcal{X}_2$. The decoder reconstruction may depend on the bits received from the encoder and the source observation $X_2(1), \ldots, X_2(n)$. Under a distortion constraint on the decoded sequence as in Shannon's problem above, the smallest bitrate needed was shown by Wyner and Ziv [57] to be

$$R_{X_1|X_2}(D) \stackrel{\text{def}}{=} \inf I(X_1; U|X_2), \tag{2.1}$$

where the infimisation is over all auxiliary random variables U over some alphabet $\mathcal{U}$ of cardinality no bigger than $|\mathcal{X}_1| + 1$ satisfying the Markov chain $U - X - Z$ and such that there is a function $f : \mathcal{X}_2 \times \mathcal{U} \to \hat{\mathcal{X}}_2$ satisfying $\mathbb{E}[d(X_1, f(X_2, U))] \leq D$.

The CEO problem is a special case of the general multi-terminal (lossy) source coding problem (sometimes also referred to as the many-help-one problem). The only other examples where the rate region had been completely characterised at the time of this work were due to Körner and Marton [30] and Berger and Yeung [7]. After this work was completed, Wagner and Anantharam [54] derived the currently best known outer bound on the rate region of the general multi-terminal source coding problem and provided another example where the rate region can be completely characterised. Perhaps more significantly, they showed that all known tight upper bounds are special cases of their result. As we mention below, using some of the results presented here, recently Tavildar, Wagner and Viswanath have provided yet more examples [46, 55] where the rate region can be fully characterised, most notably the two-user Gaussian multi-terminal problem for quadratic distortions which remained open from when Oohama [33] proposed a partial solution.

There is a natural achievable strategy for multi-terminal source coding. Each encoder quantises its observation as in the single-user source coding. Since the observations are dependent, the quantisation codewords produced by this at the different encoders are also dependent. Then, instead of communicating the indices of the quantised codewords as in single-user source coding, they send only a "hash" of the indices. The hashing is performed such that the decoder is able to recover the dependent codewords produced by the encoders from the hashes it receives. This hashing method is commonly referred to as *binning* and is due to Cover [13] who introduced it for the Slepian and Wolf's problem [44] discussed earlier. For an application to the discrete-alphabet (lossy) source coding case, see [5, 28, 50]. For the continuous case, which is of interest here, the technique appears in [22, 33, 56]. We will refer to this natural achievable strategy as the Berger-Tung-Housewright achievable strategy.

The CEO problem was first studied by Berger, Zhang, and Viswanathan [8] in the context of discrete memoryless sources and observations. They characterised the relationship between the sum-rate at which the sensors communicate to the CEO and the minimum error frequency at which the decoder can recover the hidden source, in the limit of $L \to \infty$ and high sum-rates. Ideas from hypothesis testing were used to obtain the converse. Viswanathan and Berger [52] studied the memoryless Gaussian version of this problem under a mean-squared error (MSE) distortion criterion. This problem is commonly referred to as the *quadratic Gaussian* CEO problem. They showed that in the limit of a large number of sensors ($L \to \infty$) making observations of the same quality, the Berger-Tung-Housewright achievable strategy is optimal for high sum-rates. They computed upper and lower bounds for the sum-rate distortion product for this setting. The upper bound was derived based on the Berger-Tung-Housewright region while the derivation of the lower bound was an extension of the converse proof in [8] and used techniques from estimation theory. Oohama [34] studied the same sum-rate problem in the limit of large L, and showed that the sum-rate of the Berger-Tung-Housewright achievable region is tight (at all sum-rates). Multi-user information theory methods (esp. entropy power inequality [14, pg. 496]) were used to prove the converse.

The entire rate region of the quadratic Gaussian CEO problem for a finite number of sensors (not all necessarily having the same quality of observations) was characterised independently by Oohama [35, 36] and in the work which forms part of the next chapter. The result is that the Berger-Tung-Housewright achievable region is indeed tight. This result has proved instrumental in resolving the long standing two-user Gaussian distributed source coding problem [55]. We also note here that the quadratic Gaussian CEO result has been extended by Tavildar, Viswanath and Wagner [46] to a more general setting where the observations are not necessarily independent conditioned on the quantity of interest, but obey a Markov relationship which they call the "tree-structure" condition.

As mentioned earlier, the CEO problem can be thought of as an abstraction of the networked sensing problem. However, in a sensor network setting the sensors are prone to failures due to malfunction or temporary outages to conserve battery. As part of the next chapter, we use a simple model proposed in [24] to study the robustness of the natural achievable strategy (quantisation followed by binning) to sensor failures. Our result is that the natural achievable strategy is optimal for the model considered in [24].

As pointed out earlier, in the CEO problem, the sensors are not allowed to co-operate or communicate with each other when producing their encoded bits. However, this may not be very realistic in a sensor network scenario where the sensors may be in a position to send messages to each other. In chapter 4, we consider models where sensors are allowed to communicate with each other. In one setup we consider, we allow the sensors to broadcast their messages. All the other sensors and the CEO receive this broadcast. Each broadcast is counted only once towards the sum-rate. The communication may now happen over many rounds with each sensor sending out one message in each round. The message is allowed to be a function of the observation made by the sensor producing it and all the messages which were transmitted by the other sensors up to that point. We will call this problem the CEO problem with *sensor broadcast*.

We also consider a related problem where the sensors do not broadcast, but they are allowed to interact with each other over rate constrained links. Here also the quantity of interest will be the overall sum-rate which is the sum of the rates at which the sensors communicate with each other and to the CEO. We will call this the CEO problem with *sensor interaction*.

We first show using an example that compared to the CEO problem, both sensor broadcast and sensor interaction setups can lead to a reduction in the sum-rate required for producing the same quality of reproduction at the CEO. We then derive a "single-letter" lower bound for the sum-rate of the CEO problem with sensor broadcast. We would like to point out that the form of our lower bound has similarities to the lower bound for multi-terminal source coding problems in [54].

This lower bound is shown to be tight for the Gaussian version of the problem. It turns out that the optimal coding scheme for the Gaussian CEO problem is also optimal for this problem. In other words, the sensors do not have to make use of their access to messages from some of the other sensors. This means that for the Gaussian problem the additional flexibility of sensor broadcast does not lead to a reduction in the sum-rate.

An interesting consequence of this is that the optimal scheme for the CEO problem (which involves the sensors not communicating with each other) is also optimal for the Gaussian CEO problem with sensor interaction. That is, interaction is not useful to reduce the sum-rate for the Gaussian CEO problem with sensor interaction.

Chapter 3

The Gaussian CEO Problem and a Robustness Result

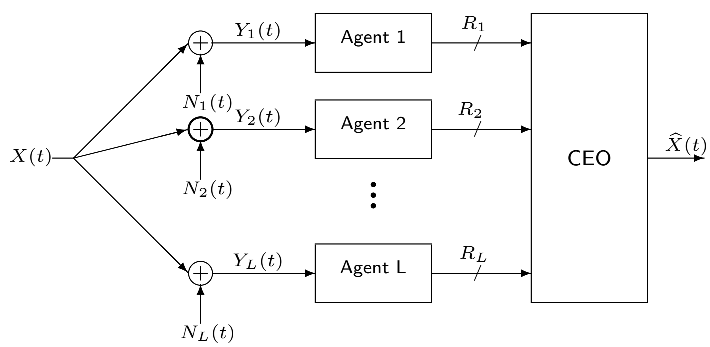


Figure 3.1: The Quadratic Gaussian CEO Problem

We begin by formulating the quadratic Gaussian CEO problem. We will use the following notation throughout this part of the thesis. Random variables and processes will be denoted by uppercase letters and their realisations by lowercase letters. n-

length sequences will be denoted by a superscript n [e.g. $X^n = (X(1), \ldots, X(n))$]. For a set $A \subseteq \{1, \ldots, L\}$, we will denote $(Y_k,\ k \in A)$ by $\mathbf{Y}_A$. Similarly, $\mathbf{Y}_A^n = (Y_k^n,\ k \in A)$. When $A = \{1, \ldots, L\}$, where L is the number of sources, we will drop the subscript as in $\mathbf{Y}^n = (Y_1^n, \ldots, Y_L^n)$ and $\mathbf{C} = (C_1, \ldots, C_L)$. Also, when $A = \{k\}$, we write Y_l^n for $Y_{\{l\}}^n$.

Let $X(t),\ t = 1, 2, \ldots$ be i.i.d Gaussian random variables with zero mean and variance σ_X^2. The decoder is interested in this sequence. But it cannot be observed directly. Instead, there are L sensors (encoders which cannot cooperate) who observe independently corrupted versions of this sequence. For $k = 1, \ldots, L$, sensor k observes $Y_k(t)$, a noisy version of the $X(t)$ sequence corrupted by independent additive white Gaussian noise.

$$Y_k(t) = X(t) + N_k(t),\ k = 1, \ldots, L,$$

where the noise sequence $N_k(t),\ t = 1, 2, \ldots$ are i.i.d Gaussian random variables independent of the X process with mean zero and variance $\sigma_{N_k}^2$. Also, the random processes N_k and N_j are independent for $j \neq k$.

For $k = 1, \ldots, L$, sensor k encodes its observation Y_k^n separately to $C_k = f_k^{(n)}(Y_k^n)$ where $f_k^{(n)} : \mathbb{R}^n \to \mathcal{C}_k^{(n)}$ is the encoding map. C_k is sent to the decoder at the rate $R_k \overset{\text{def}}{=} |\mathcal{C}_k^{(n)}|/n$. The decoder outputs an estimate $\widehat{X}^n = g^{(n)}\left(f_1^{(n)}(Y_1^n), \ldots, f_L^{(n)}(Y_L^n)\right)$ of X^n. Here $g^{(n)} : \mathcal{C}_1^{(n)} \times \ldots \times \mathcal{C}_L^{(n)} \to \mathbb{R}^n$ is the decoding map. The average distortion is

$$D^{(n)} = \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left((X(t) - \widehat{X}(t))^2\right). \tag{3.1}$$

For a target distortion D, a rate L-tuple $(R_1, \ldots, R_L)$ is said to be *achievable* if there are encoders $(f_1^{(n)}, \ldots, f_L^{(n)})$ working at these rates and decoder $g^{(n)}$ such that $D^{(n)} \leq D$ for some n. The closure of the set of all achievable rate L-tuples is called the rate-region and we denote it by $\mathcal{R}_\star(D) \subseteq \mathbb{R}_+^L$.

Let $\mathcal{R}_N(D)$ be the Berger-Tung-Housewright achievable region using Gaussian auxiliary random variables. We show that

$$\mathcal{R}_\mathcal{N}(D) = \bigcup_{(r_1,\ldots,r_L)\in\mathcal{F}(D)} \mathcal{R}_D(r_1, \ldots, r_L), \tag{3.2}$$

where

$$\mathcal{R}_D(r_1, \ldots, r_L) \stackrel{\text{def}}{=}$$

$$\left\{ \begin{array}{rl} (R_1, \ldots, R_L) : \ \sum_{k\in A} R_k & \geq \sum_{k\in A} r_k + \frac{1}{2}\log\frac{1}{D} \\ & \quad -\frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{k\in A^c} \frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}\right), \\ & \quad \forall \text{ non-empty } A \subseteq \{1, \ldots, L\} \end{array} \right\},$$

and

$$\mathcal{F}(D) \stackrel{\text{def}}{=} \left\{ (r_1, \ldots, r_L) \in \mathbb{R}_+^L : \frac{1}{\sigma_X^2} + \sum_{k=1}^{L} \frac{1-\exp(-2r_k)}{\sigma_{N_k}^2} = \frac{1}{D} \right\}.$$

As will become clear in the next section, r_k can be interpreted as the rate the k-th sensor spends in quantising its observation noise.

3.1 Main result

Our main result is

Theorem 3.1.1

$$\mathcal{R}_\star(D) = \mathcal{R}_\mathcal{N}(D).$$

Since $\mathcal{R}_\star(D)$ must be a convex set, the theorem implies that convexification is not required in (3.2), i.e., we do not need to time-share between points in different $\mathcal{R}_D$.

Proof

Achievability of theorem 3.1.1 Since Berger-Tung-Housewright achievable region has been discussed elsewhere in detail [22, 33, 34, 50], we will only provide a sketch along the lines of [51]. For each sensor, we define the auxiliary random variable $U_k = Y_k + W_k,\ k = 1, \ldots, L$, where $W_k \sim \mathcal{N}(0, \sigma^2_{W_k})$ are independently distributed and independent of $(\mathbf{Y}, X)$. $\sigma^2_{W_k} \geq 0$ are parameters which will be specified in terms of the target distortion D. To construct the random codebook for sensor k, draw $2^{nI(U_k;Y_k)+n\epsilon}$ n-length U_k vectors randomly according to the marginal of U_k and divide them equally (or *bin* them) into 2^{nR_k} bins. Encoding at sensor k involves picking a codeword U_k^n which is jointly typical [14] with the observed Y_k^n vector (with high probability there will be at least one such codeword for large enough n) and sending the bin index. It can be proved along the lines of the extended Markov lemma of [34] [50] that $X^n, \mathbf{Y^n}$, and $\mathbf{U^n}$ will be jointly typical with high probability because $U_j - Y_j - X - Y_k - U_k, k \neq j$ are Markov chains.

The decoder attempts to recover the codewords from the the specified bins. The codewords $\mathbf{U}^n$ chosen by the encoders are jointly typical. No other set of codewords in the specified bins will be jointly typical with high probability if

$$\sum_{k\in A}(I(Y_k;U_k)-R_k)^+ < \sum_{k\in A} h(U_k) - h(\mathbf{U}_A|\mathbf{U}_{A^c}),\ \forall A \subseteq \{1,\dots,L\} \tag{3.3}$$

where A^c is the complement of A in $\{1,\dots,L\}$, and $(x)^+$ stands for 0 if $x<0$ and x otherwise. Viswanath [51] has shown that these conditions are the same as[1]

$$\sum_{k\in A} R_k > f(A) \stackrel{\text{def}}{=} I(\mathbf{Y}_A;\mathbf{U}_A|\mathbf{U}_{A^c}), \forall A \subseteq \{1,\dots,L\}. \tag{3.4}$$

After recovering $\mathbf{U}^n$, the decoder reconstructs $\widehat{\mathbf{X}}^n$ by applying the following function component-wise

$$\hat{x} = g(\mathbf{u}) \stackrel{\text{def}}{=} \mathbb{E}(X|\mathbf{U}=\mathbf{u}).$$

Joint typicality of X^n and the codewords $\mathbf{U}^n$ implies that the mean squared distortion in estimating X is $\mathbb{E}\left((X-\mathbb{E}(X|\mathbf{U}))^2\right)$. We set this equal to the target distortion D. This implies that the valid choices of $\sigma^2_{W_k}$ must satisfy

$$\frac{1}{D} = \frac{1}{\sigma_X^2} + \sum_{k=1}^{L} \frac{1}{\sigma^2_{N_k}+\sigma^2_{W_k}}. \tag{3.5}$$

Let us define (only for the achievability part of the proof)

$$r_k \stackrel{\text{def}}{=} I(Y_k;U_k|X) = \frac{1}{2}\log\frac{\sigma^2_{N_k}+\sigma^2_{W_k}}{\sigma^2_{W_k}}.$$

We can interpret r_k as the rate the k-th sensor spends in quantising its observation noise. We will use r_k's as the parameters instead of $\sigma^2_{W_k}$. Note that for any choice

[1] For consistency, we define $f(\varnothing) \stackrel{\text{def}}{=} 0$.

of $(r_k \geq 0,\ k = 1, \ldots, L)$, we can find a corresponding $(\sigma^2_{W_k} \geq 0,\ k = 1, \ldots, L)$ and therefore, a set of auxiliary random variables. Then we can rewrite (3.5) in terms of r_k's as

$$\frac{1}{D} = \frac{1}{\sigma^2_X} + \sum_{k=1}^{L} \frac{1 - \exp(-2r_k)}{\sigma^2_{N_k}}, \tag{3.6}$$

and $f(\cdot)$ in (3.4) for all non-empty $A \subseteq \{1, \ldots, L\}$ as

$$\begin{aligned}
f(A) &= I(\mathbf{Y}_A; \mathbf{U}_A | \mathbf{U}_{A^c}) \\
&= I(\mathbf{Y}_A, X; \mathbf{U}_A | \mathbf{U}_{A^c}), && \because (X, \mathbf{U}_{A^c}) - \mathbf{Y}_A - \mathbf{U}_A \text{ is a Markov chain} \\
&= I(X; \mathbf{U}_A | \mathbf{U}_{A^c}) + I(\mathbf{Y}_A; \mathbf{U}_A | X), && \because (\mathbf{Y}_A, \mathbf{U}_A) - X - \mathbf{U}_{A^c} \text{ is a Markov chain} \\
&= I(X; \mathbf{U}_A | \mathbf{U}_{A^c}) + \sum_{k \in A} r_k, \\
& && \because (Y_k, U_k) - X - (\mathbf{Y}_{A-\{k\}}, \mathbf{U}_{A-\{k\}}) \text{ is a Markov chain} \\
&= h(X | \mathbf{U}_{A^c}) - h(X | \mathbf{U}) + \sum_{k \in A} r_k \\
&= -\frac{1}{2} \log \left(\frac{1}{\sigma^2_X} + \sum_{k \in A^c} \frac{1 - \exp(-2r_k)}{\sigma^2_{N_k}} \right) + \frac{1}{2} \log \frac{1}{D} + \sum_{k \in A} r_k.
\end{aligned}$$

Now (3.4) can be rewritten as, for all non-empty $A \subseteq \{1, \ldots, L\}$

$$\sum_{k \in A} R_k \geq f(A) = \frac{1}{2} \log \frac{1}{D} - \frac{1}{2} \log \left(\frac{1}{\sigma^2_X} + \sum_{k \in A^c} \frac{1 - \exp(-2r_k)}{\sigma^2_{N_k}} \right) + \sum_{k \in A} r_k. \tag{3.7}$$

Note that we have replaced the strict inequality with the weaker form since any rate L-tuple which satisfies all the weaker inequalities is also in the closure of the set of all achievable rates.

Equations (3.6) and (3.7) together give the achievability $\mathcal{R}_\star(D) \supseteq \mathcal{R}_\mathcal{N}(D)$.

Converse of theorem 3.1.1 Suppose $(R_1, \ldots, R_L) \in \mathcal{R}_\star(D)$. Then, for every $\epsilon > 0$, there is a large enough n for which there are blocklength-n encoders working at rates $(R_1 + \epsilon, \ldots, R_L + \epsilon)$ and a corresponding decoder such that their average distortion is less than or equal to D. Let $\mathbf{C} = (C_1, \ldots, C_L)$ denote all the messages produced by these encoders after observing an n-block. Let us define

$$r_k \stackrel{\text{def}}{=} \frac{1}{n} I(Y_k^n; C_k | X^n). \tag{3.8}$$

Note that this definition of r_k is for the proof of the converse and it is different from the definition of r_k in the proof of achievability. We will use the above definition (3.8) again in the next section.

For any $A \subseteq \{1, \ldots, L\}$,

$$\begin{aligned}
\sum_{k \in A} (R_k + \epsilon) \geq \sum_{k \in A} \frac{1}{n} H(C_k) &\geq \frac{1}{n} H(\mathbf{C}_A) \\
&\geq \frac{1}{n} H(\mathbf{C}_A | \mathbf{C}_{A^c}) \\
&\geq \frac{1}{n} I(\mathbf{Y}_A^n; \mathbf{C}_A | \mathbf{C}_{A^c}) \\
&\stackrel{\text{(a)}}{=} \frac{1}{n} I(\mathbf{Y}_A^n, X^n; \mathbf{C}_A | \mathbf{C}_{A^c}) \\
&\stackrel{\text{(b)}}{=} \frac{1}{n} I(X^n; \mathbf{C}_A | \mathbf{C}_{A^c}) + \sum_{k \in A} r_k \\
&= \frac{1}{n} I(X^n; \mathbf{C}) - \frac{1}{n} I(X^n; \mathbf{C}_{A^c}) + \sum_{k \in A} r_k,
\end{aligned} \tag{3.9}$$

where we used the fact that $X^n - (\mathbf{Y}_A^n, \mathbf{C}_{A^c}) - \mathbf{C}_A$ and

$(Y_k^n, C_k) - X^n - (C_1, \ldots, C_{k-1}, C_{k+1}, \ldots, C_L)$ are Markov chains to get (a) and (b) respectively.

We have a simple lower bound for the first term $I(X^n;\mathbf{C})/n$. This is due to Shannon and is referred to as Shannon lower bound [4,43].

$$\begin{aligned}
\frac{1}{n}I(X^n;\mathbf{C}) &\geq \frac{1}{n}I(X^n;\widehat{X}^n), \qquad \text{by data processing inequality} \\
&= \frac{1}{n}h(X^n) - \frac{1}{n}h\left(X^n - \widehat{X}^n|\widehat{X}^n\right) \\
&\geq \frac{1}{n}h(X^n) - \frac{1}{n}h(X^n - \widehat{X}^n) \\
&\geq \frac{1}{2}\log 2\pi e\sigma_X^2 - \frac{1}{2}\log 2\pi e D \\
&= \frac{1}{2}\log\frac{\sigma_X^2}{D},
\end{aligned} \tag{3.10}$$

where we used the fact that the Gaussian distribution maximises the differential entropy for fixed second moment to get the last inequality.

To bound the second term of (3.9) we will need the following lemma.

Lemma 3.1.2 *Let $A \subseteq \{1,\ldots,L\}$. Then*

$$\frac{1}{\sigma_X^2}\exp\left(\frac{2}{n}I(X^n;\mathbf{C}_A)\right) \leq \frac{1}{\sigma_X^2} + \sum_{k\in A}\frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}. \tag{3.11}$$

We prove this lemma in appendix A. The lemma can be interpreted as follows: for any quantisation strategy at the sensors, if we compute the rates r_k's which is in effect being used to quantise the observation noises by any set A of sensors, then the lemma gives an upper bound on the mutual information that can be derived about the hidden source from the messages produced by the sensors in set A. The key inequality used in proving the lemma is the entropy power inequality [14, pg. 496] which becomes tight under the jointly Gaussian quantisation strategy we used in the achievability proof as we will see. Using the bounds from (3.10) and (3.11) in (3.9), for all $A \subseteq \{1, \ldots, L\}$

$$\sum_{k \in A} R_k + \epsilon \geq \frac{1}{2} \log \frac{1}{D} - \frac{1}{2} \log \left(\frac{1}{\sigma_X^2} + \sum_{k \in A^c} \frac{1 - \exp(-2r_k)}{\sigma_{N_k}^2} \right) + \sum_{k \in A} r_k. \qquad (3.12)$$

When $A = \varphi$ in (3.12) (or equivalently from (3.10) and (3.11) with $A = \{1, \ldots, L\}$), we have the condition

$$\frac{1}{\sigma_X^2} + \sum_{k=1}^{L} \frac{1 - \exp(-2r_k)}{\sigma_{N_k}^2} \geq \frac{1}{D}. \qquad (3.13)$$

Since this is true for arbitrarily small positive values of ϵ, (3.12) with ϵ set to zero, and (3.13) along with the non-negativity constraints on r_k's, define an outer bound for $\mathcal{R}_\star(D)$. It is easy to observe that replacing the inequality in (3.13) with an equality will not change this outer bound. Thus we have $\mathcal{R}_\star(D) \subseteq \mathcal{R}_\mathcal{N}(D)$. □

Remarks:

- Since $\mathcal{R}_\star(D)$ must be convex, we can conclude that $\mathcal{R}_\mathcal{N}(D)$ is also convex. This can also be directly inferred by noting that the right hand side of (3.12) (with ϵ set to zero) is convex in $(r_1, \ldots, r_L)$ and that the set of $(r_1, \ldots, r_L) \in \mathbb{R}_+^L$ which satisfy (3.13) is a convex set.

- We can explicitly compute the rate-region by solving the following optimisation problem

$$\min_{(R_1, \ldots, R_L) \in \mathcal{R}_\star(D)} \sum_{k=1}^{L} \alpha_k R_k$$

for all choices of $(\alpha_1, \ldots, \alpha_L) \in \mathbb{R}_+^L$. Without loss of generality assume that

$\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_L$. Recall that

$$\mathcal{R}_{\mathcal{N}}(D) = \bigcup_{(r_1,\ldots,r_L)\in\mathcal{F}(D)} \mathcal{R}_D(r_1,\ldots,r_L).$$

Therefore,

$$\min_{(R_1,\ldots,R_L)\in\mathcal{R}_\star(D)} \sum_{k=1}^{L} \alpha_k R_k = \min_{(r_1,\ldots,r_L)\in\mathcal{F}(D)} \min_{(R_1,\ldots,R_L)\in\mathcal{R}_D(r_1,\ldots,r_L)} \sum_{k=1}^{L} \alpha_k R_k.$$

The inner minimisation is easy to perform by noting that

$$\mathcal{R}_D(r_1,\ldots,r_L) \stackrel{\text{def}}{=} \left\{ \begin{array}{ll} (R_1,\ldots,R_L): & \sum_{k\in A} R_k \geq I(\mathbf{Y}_A;\mathbf{U}_A|\mathbf{U}_{A^c}), \\ & \forall \text{ non-empty } A \subseteq \{1,\ldots,L\} \end{array} \right\},$$

where the U_k's are defined as in section 3.1. The optimal choice of $(R_1,\ldots,R_L)$ for the inner minimisation is

$$\begin{aligned} R_k &= I(Y_k;U_k|U_{k+1},\ldots,U_L) \\ &= I(X;U_k|U_{k+1},\ldots,U_L) + I(Y_k;U_k|X) \\ &= I(X;U_k,U_{k+1},\ldots,U_L) - I(X;U_{k+1},\ldots,U_L) + I(Y_k;U_k|X) \\ &= \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i=k}^{L} \frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}\right) \\ &\qquad - \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i=k+1}^{L} \frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}\right) + r_k. \end{aligned}$$

We can now rewrite the optimisation as

$$\begin{aligned}
&\min_{(R_1,\ldots,R_L)\in\mathcal{R}_\star(D)} \sum_{k=1}^{L} \alpha_k R_k \\
&= \min_{(r_1,\ldots,r_L)\in\mathcal{F}(D)} \frac{\alpha_1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{k=1}^{L}\frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}\right) \\
&\qquad -\sum_{m=1}^{L-1}\frac{\alpha_m-\alpha_{m+1}}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{k=m+1}^{L}\frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}\right) \\
&\qquad -\frac{\alpha_L}{2}\log\frac{1}{\sigma_X^2}+\sum_{k=1}^{L}\alpha_k r_k. \qquad (3.14)
\end{aligned}$$

If we rewrite this in terms of parameters $p_k \stackrel{\text{def}}{=} \frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}$, $k = 1,\ldots,L$, and note that the first term is $\alpha_1 \log(1/D)/2$, we get a convex optimisation problem which is in a form that can be readily solved using Lagrange multipliers.

$$\begin{aligned}
\min_{p_1,\ldots,p_L} \quad & -\sum_{m=1}^{L-1}\frac{\alpha_m-\alpha_{m+1}}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{k=m+1}^{L}p_k\right)-\sum_{k=1}^{L}\frac{\alpha_k}{2}\log(1-p_k\sigma_{N_k}^2) \\
& -\frac{\alpha_1}{2}\log D+\frac{\alpha_L}{2}\log\sigma_X^2 \\
\text{subject to} \quad & \frac{1}{\sigma_X^2}+\sum_{k=1}^{L}p_k=\frac{1}{D}, \\
& p_k \geq 0,\ k=1,2,\ldots,L.
\end{aligned}$$

The explicit expression for the solution is cumbersome and is omitted.

An interesting case is the symmetric setting where all the sensors observe the source with the same signal-to-noise ratio ($\sigma_{N_k}^2 = \sigma_N^2$, $k = 1,\ldots,L$) and they all operate at the same rate, say R. For this case the optimal distortion $D_{\text{opt}_L}(R)$ that can be achieved is defined as

$$D_{\text{opt}_L}(R) = \inf\{D : (R,R,\ldots,R) \in \mathcal{R}_\star(D)\}. \qquad (3.15)$$

A direct consequence of theorem 3.1.1 is the following corollary which we state without proof.

Corollary 3.1.3

$$\frac{1}{D_{\mathrm{opt}_L}(R)} = \frac{1}{\sigma_X^2} + L\left(\frac{1-\exp(-2r)}{\sigma_N^2}\right), \tag{3.16}$$

where r is the unique solution of

$$R = \frac{1}{2L}\log\left(1 + L\frac{1-\exp(-2r)}{\sigma_N^2}\right) + r. \tag{3.17}$$

3.2 A robustness result

In this section we examine the robustness of the achievability scheme used in the last section to failure of the sensors. Let us consider the symmetric case of $\sigma_{N_1}^2 = \sigma_{N_2}^2 = \ldots = \sigma_{N_L}^2 = \sigma_N^2$, with all sensors allowed to work at the same maximum rate, say R. Suppose that the sensors are prone to failures (also see [24]). In particular, assume that we are given that at least K out of the L sensors will not fail. But we do not know beforehand how many sensors will fail (except that the number of failures is less than or equal to $L-K$) or which sensors will fail. Suppose we further insist that when exactly K sensors are active, the resulting distortion D_K be $D_{\mathrm{opt}_K}(R)$ which is defined as the lowest achievable distortion for the symmetric K-sensor CEO problem when all K sensors work at the same rate R [cf. (3.15)]. Corollary 3.1.3 gives us

$D_{\text{opt}_K}(R)$.

$$\frac{1}{D_{\text{opt}_K}(R)} = \frac{1}{\sigma_X^2} + K\left(\frac{1-\exp(-2r)}{\sigma_N^2}\right), \tag{3.18}$$

where r is the unique solution of

$$R = \frac{1}{2K}\log\left(1 + K\frac{1-\exp(-2r)}{\sigma_N^2}\right) + r. \tag{3.19}$$

At this point it is not clear that it would be possible to achieve this. But as we will see this is indeed possible.

Under these constraints, we want to characterise the distortions achievable when more than K sensors are active. If D_m denotes the average distortion when m sensors are active, where $m = K+1, \ldots, L$, then we want to characterise the set of distortions $(D_{K+1}, D_{K+2}, \ldots, D_L)$ achievable under the restriction that each sensor operates at rate R and $D_K = D_{\text{opt}_K}(R)$. The result is that there is no trade-off between these distortions and the optimal distortions are given by

$$\frac{1}{D_m(R)} = \frac{1}{\sigma_X^2} + m\left(\frac{1-\exp(-2r)}{\sigma_N^2}\right),\ m = K+1, K+2, \ldots, L, \tag{3.20}$$

where r is given by (3.19). Moreover, the optimal encoding strategy from last section (where now each sensor encodes as if for a symmetric K-user CEO problem) will achieve this optimal performance.

Formally, the rates of the encoders are defined as in the previous section. For each set of m sensors, we define a sequence of decoder maps (which will be used by the CEO when that set of sensors is active) indexed by the blocklength n. Let D_m^n be the smallest of the distortions obtained among all the decoders with a given blocklength n [cf. (3.1)]. This is done for each $m = K, K+1, \ldots, L$, and a set of distortions D_m, $m = K, K+1, \ldots, L$ is said to be achievable if for any $\epsilon > 0$, there is some n such that $D_m^n \leq D_m + \epsilon$, $m = K, K+1, \ldots, L$. The closure of the set of all these achievable $L-K+1$-tuples of distortions is a *distortion region* which is obviously of interest. However, a description of this region still remains open. Our result is a characterisation of the projection of this closure on the hyperplane $D_K = D_{\text{opt}_K}(R)$, and the result is that this projection is a point $(D_{\text{opt}_K}(R), D_{K+1}(R), D_{K+2}(R), \ldots . D_L(R))$, where $D_m(R)$, $m = K+1, \ldots, L$ are given by (3.20).

The achievability of this point is already the subject of a paper [24]. It also readily follows from the discussion of the achievability result in the last section. Following the development there, each sensor k generates a codebook for the auxiliary random variable $U_k = Y_k + W_k$ and bins it into $2^{nR_k} = 2^{nR}$ bins. Since the problem is symmetric, we will choose $\sigma^2_{W_k} = \sigma^2_W$ for all k. The main difference is that unlike there, now the decoder should be able to decode the codewords upon receiving from at least K sensors. This means that the analogue of (3.3) is that for all cardinality-K subsets B of $\{1, \ldots, L\}$,

$$\sum_{k \in A} (I(Y_k; U_k) - R_k)^+ < \sum_{k \in A} h(U_k) - h(\mathbf{U}_A | \mathbf{U}_{A-B}), \ \forall A \subseteq B,$$

But since $R_k = R$ and $\sigma^2_{W_k} = \sigma^2_W$, the condition for $A = B$ subsumes all others. Thus we need

$$KR > I(Y_B; U_B).$$

Note that if the decoder can decode the codewords of the active sensors when exactly K sensors are active, then it can clearly decode the codewords of the active sensors when more than K are active. The distortion D achieved when $m \geq K$ sensors are active is given by [cf. (3.5)]

$$\frac{1}{D_m} = \frac{1}{\sigma_X^2} + \frac{m}{\sigma_N^2 + \sigma_W^2}.$$

Defining r in terms of σ_W^2 as in the achievability proof from last section and manipulating the two previous equations we get [cf. (3.6) and (3.7)]

$$\frac{1}{D_m} = \frac{1}{\sigma_X^2} + K\left(\frac{1 - \exp(-2r)}{\sigma_N^2}\right),$$

is achievable for $m = K, K+1, \ldots, M$, for $r \geq 0$ satisfying

$$R = \frac{1}{2K} \log\left(1 + K\frac{1 - \exp(-2r)}{\sigma_N^2}\right) + r.$$

Since by corollary 3.1.3, $D_K = D_{\text{opt}_K}$, we can conclude the achievability.

The converse argument can be roughly summarised as follows. The requirement that $D_K = D_{\text{opt}_K}(R)$ should translate into tightness for all the inequalities we had in the converse arguments in the previous section (applied to the symmetric K-sensor CEO problem). In particular, we should find $r_k \stackrel{\text{def}}{=} \frac{1}{n} I(Y_k^n; C_k | X^n)$ to be the same for all k and equal to the optimal choice for the K-sensor problem, namely the r from (3.19). This fixes the rate spent on quantising the noise at the encoders. Now invoking lemma 3.1.2 should give us upper bounds on the mutual information (and in turn lower bounds on the distortion) for the case when $m > K$ sensors are active. Since the achievable strategy uses Gaussian quantisation at the encoders, these bounds given by lemma 3.1.2 are in fact attained.

The converse argument in some more detail is as follows. Let A be an arbitrary subset of $\{1, 2, \ldots, L\}$ of cardinality K. If only the sensors in A are active, the CEO is required to achieve a distortion of $D_{\text{opt}_K}(R)$. This means that for any $\epsilon > 0$, we have an n and blocklength-n encoders and decoders whose average distortion is $D_{\text{opt}_K}(R) + \epsilon$ upon receiving from sensors in any arbitrary set A of cardinality K. Then,

$$\begin{aligned}
KR &\geq \frac{1}{n} H(C_A) \\
&\geq \frac{1}{n} I(Y_A^n; C_A) \\
&= \frac{1}{n} I(X^n; C_A) + \sum_{k \in A} r_k \\
&\geq \frac{1}{2} \log \frac{\sigma_X^2}{D_{\text{opt}_K}(R) + \epsilon} + \sum_{k \in A} r_k,
\end{aligned}$$

where the equality is analogous to (3.9), and the last inequality follows from an analogue of (3.10). Let us define $g_k = (1 - \exp(-2r_k))/\sigma_N^2$. Then

$$\begin{aligned}
KR &\geq \frac{1}{2} \log \frac{\sigma_X^2}{D_{\text{opt}_K}(R) + \epsilon} - \sum_{k \in A} \frac{1}{2} \log \left(1 - \sigma_N^2 g_k\right) \\
&\geq \frac{1}{2} \log \frac{\sigma_X^2}{D_{\text{opt}_K}(R) + \epsilon} - \frac{K}{2} \log \left(1 - \sigma_N^2 \frac{1}{K} \sum_{k \in A} g_k\right), \qquad (3.21)
\end{aligned}$$

by concavity of $\log(1-\sigma_N^2 x)$ and Jensen's inequality. But from lemma 3.1.2, we have

$$\begin{aligned}\frac{1}{\sigma_X^2}+\sum_{k\in A} g_k &\geq \frac{1}{\sigma_X^2}\exp\left(\frac{2}{n}I(X^n;C_A)\right)\\ &\geq \frac{1}{D_{\mathrm{opt}_K}+\epsilon}\\ &\geq \frac{1}{\sigma_X^2}+K\left(\frac{1-\exp(-2r)}{\sigma_N^2}\right)+\delta_1(\epsilon),\end{aligned}$$

where $\delta_1(\epsilon)\to 0$ as $\epsilon\to 0$. We used the analogue of (3.10) in the second step, and (3.18) in the last step. Repeating the chain of inequalities which led up to (3.21) and substituting from the above for $\sum_{k\in A} g_k$, we get

$$\begin{aligned}KR &\geq \frac{1}{2}\log\frac{\sigma_X^2}{D_{\mathrm{opt}_K}(R)+\epsilon}+\sum_{k\in A} r_k\\ &\geq \frac{1}{2}\log\frac{\sigma_X^2}{D_{\mathrm{opt}_K}(R)+\epsilon}-\frac{K}{2}\log\left(1-\sigma_N^2\frac{1}{K}\sum_{k\in A} g_k\right)\\ &\geq \frac{1}{2}\log\frac{\sigma_X^2}{D_{\mathrm{opt}_K}(R)+\epsilon}+Kr+\delta_2(\epsilon)\\ &= KR+\delta_3(\epsilon),\end{aligned}$$

where $\delta_2(\epsilon)\to 0$ and $\delta_3(\epsilon)\to 0$ as $\epsilon\to 0$. In the last step above we used (3.18) and (3.19). This implies that

$$\sum_{k\in A} r_k = Kr+\delta_4(\epsilon), \tag{3.22}$$

where $\delta_4(\epsilon)\to 0$ as $\epsilon\to 0$.

Using lemma 3.1.2, for any $B\subseteq\{1,2,\ldots,L\}$ of cardinality $m\in\{K,K+1,\ldots,L\}$, we have

$$\frac{1}{\sigma_X^2}\exp\left(\frac{2}{n}I(X^n;\mathbf{C}_B)\right) \le \frac{1}{\sigma_X^2} + \sum_{k\in B}\frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}$$
$$\le \frac{1}{\sigma_X^2} + m\frac{1-\exp\left(-\frac{2}{m}\sum_{k\in B} r_k\right)}{\sigma_{N_k}^2}$$
$$\le \frac{1}{\sigma_X^2} + m\frac{1-\exp\left(-2r\right)}{\sigma_{N_k}^2} + \delta_5(\epsilon), \tag{3.23}$$

where $\delta_5(\epsilon) \to 0$ as $\epsilon \to 0$. In the last step we used (3.22) and

$$\sum_{k\in B} r_k = \frac{m/k}{\begin{pmatrix} m \\ k \end{pmatrix}} \sum_{A:\, A\subseteq B \text{ and } |A|=K} \sum_{k\in A} r_k.$$

Now by using the inequality [cf. (3.10)]

$$I(X^n;\mathbf{C}_B) \ge \frac{n}{2}\frac{\sigma_X^2}{D_m+\epsilon},$$

in (3.23), we get

$$\frac{1}{D_m} \le \frac{1}{\sigma_X^2} + m\frac{1-\exp\left(-2r\right)}{\sigma_{N_k}^2} + \delta_6(\epsilon),$$

where $\delta_6(\epsilon) \to 0$ as $\epsilon \to 0$. This completes the converse argument.

Appendix A

Proof of Lemma 3.1.2

The proof is based on an idea of Ozarow [37] which he used for solving the two-user multiple description coding problem. Oohama later adapted it for the converse proof in [34].

Let the minimum mean-squared error (MMSE) estimate of $X(t)$ from $\mathbf{Y}_A(t)$ be denoted by $S(t)$. Then $S(t) = \sum_{k\in A} \frac{\sigma^2_{\widetilde{N}}}{\sigma^2_{N_k}} Y_k(t)$, where $1/\sigma^2_{\widetilde{N}} = 1/\sigma^2_X + \sum_{k\in A} 1/\sigma^2_{N_k}$. i.e.,

$$X(t) = S(t) + \widetilde{N}(t),$$

where $\widetilde{N}(t)$ is i.i.d Gaussian with mean zero, variance $\sigma^2_{\widetilde{N}}$ and the processes $\widetilde{N}$ and $\mathbf{Y}$ are independent. It is easy to show that conditioned on $(X^n = x^n, \mathbf{C}_A = \mathbf{c}_A)$,

$Y_k^n,\ k \in A$ are independent. Thus, by entropy power inequality (EPI) [14, pg. 496],

$$\begin{aligned}\frac{1}{n}h(S^n|X^n = x^n, \mathbf{C}_A = \mathbf{c}_A) &\geq \frac{1}{2}\log\left(\sum_{k\in A}\exp\left(\frac{2}{n}h\left(\frac{\sigma^2_{\widetilde{N}}}{\sigma^2_{N_k}}Y_k^n\middle|X^n = x^n, \mathbf{C}_A = \mathbf{c}_A\right)\right)\right)\\ &= \frac{1}{2}\log\left(\sum_{k\in A}\exp\left(\frac{2}{n}h\left(\frac{\sigma^2_{\widetilde{N}}}{\sigma^2_{N_k}}Y_k^n\middle|X^n = x^n, C_k = c_k\right)\right)\right).\end{aligned}$$

We now take expectation over $(X^n, \mathbf{C}_A)$. Since log-sum-exp function is convex, we may apply Jensen's inequality to the right-hand side. This gives

$$\begin{aligned}\frac{1}{n}h(S^n|X^n, \mathbf{C}_A) &\geq \frac{1}{2}\log\left(\sum_{k\in A}\exp\left(\frac{2}{n}h\left(\frac{\sigma^2_{\widetilde{N}}}{\sigma^2_{N_k}}Y_k^n\middle|X^n, C_k\right)\right)\right)\\ &= \frac{1}{2}\log\left(\sigma^4_{\widetilde{N}}\sum_{k\in A}\frac{1}{\sigma^4_{N_k}}\exp\left(\frac{2}{n}h(Y_k^n|X^n) - \frac{2}{n}I(Y_k^n; C_k|X^n)\right)\right)\\ &= \frac{1}{2}\log\left(\sigma^4_{\widetilde{N}}\sum_{k\in A}\frac{2\pi e\sigma^2_{N_k}}{\sigma^4_{N_k}}\exp\left(-\frac{2}{n}I(Y_k^n; C_k|X^n)\right)\right).\end{aligned}$$

$$\Rightarrow \qquad \exp\left(\frac{2}{n}h(S^n|X^n, \mathbf{C}_A)\right) \geq 2\pi e\sigma^4_{\widetilde{N}}\left(\sum_{k\in A}\frac{\exp(-2r_k)}{\sigma^2_{N_k}}\right), \tag{A.1}$$

where we used the definition (3.8). Now recall that $X^n = S^n + \widetilde{N}^n$, where S^n and $\widetilde{N}^n$ are independent. It is easy to see that conditioned on the messages $\mathbf{C}_A = \mathbf{c}_A$, S^n and $\widetilde{N}^n$ are still independent and the marginal distribution of $\widetilde{N}^n$ remains the same. Applying EPI,

$$\begin{aligned}&\exp\left(\frac{2}{n}h(X^n|\mathbf{C}_A = \mathbf{c}_A)\right) \geq \exp\left(\frac{2}{n}h(S^n|\mathbf{C}_A = \mathbf{c}_A)\right) + \exp\left(\frac{2}{n}h(\widetilde{N}^n)\right)\\ \Rightarrow \quad &\frac{1}{n}h(X^n|\mathbf{C}_A = \mathbf{c}_A) \geq \frac{1}{2}\log\left(\exp\left(\frac{2}{n}h(S^n|\mathbf{C}_A = \mathbf{c}_A)\right) + 2\pi e\sigma^2_{\widetilde{N}}\right)\\ \overset{(a)}{\Rightarrow} \quad &\frac{1}{n}h(X^n|\mathbf{C}_A) \geq \frac{1}{2}\log\left(\exp\left(\frac{2}{n}h(S^n|\mathbf{C}_A)\right) + 2\pi e\sigma^2_{\widetilde{N}}\right)\\ \Rightarrow \quad &\exp\left(\frac{2}{n}h(X^n|\mathbf{C}_A)\right) \geq \exp\left(\frac{2}{n}h(S^n|\mathbf{C}_A)\right) + 2\pi e\sigma^2_{\widetilde{N}},\end{aligned}$$

where (a) follows from Jensen's inequality and convexity of $1/2 \log\left(e^{2x}+k\right)$ in x for $k \geq 0$. We can rewrite the above as

$$\begin{aligned}\exp\left(-\frac{2}{n}I(X^n;\mathbf{C}_A)\right) &= \exp\left(-\frac{2}{n}h(X^n)\right)\exp\left(\frac{2}{n}h(X^n|\mathbf{C}_A)\right)\\ &\geq \frac{1}{2\pi e\sigma_X^2}\exp\left(\frac{2}{n}h(S^n|\mathbf{C}_A)\right)+\frac{\sigma_{\widetilde{N}}^2}{\sigma_X^2}. \qquad \text{(A.2)}\end{aligned}$$

Also

$$h(S^n|\mathbf{C}_A) = h(S^n|X^n,\mathbf{C}_A) + I(X^n;S^n|\mathbf{C}_A),$$

and

$$I(X^n;S^n|\mathbf{C}_A) = I(X^n;S^n,\mathbf{C}_A) - I(X^n;\mathbf{C}_A) = I(X^n;S^n) - I(X^n;\mathbf{C}_A),$$

since $X^n - S^n - \mathbf{Y}_A^n - \mathbf{C}_A$ is a Markov chain.

$$\begin{aligned}\Rightarrow \exp\left(\frac{2}{n}h(S^n|\mathbf{C}_A)\right) &\geq \exp\left(\frac{2}{n}(h(S^n|X^n,\mathbf{C}_A)+I(X^n;S^n)-I(X^n;\mathbf{C}_A))\right)\\ &= \frac{\sigma_X^2}{\sigma_{\widetilde{N}}^2}\exp\left(\frac{2}{n}(h(S^n|X^n,\mathbf{C}_A)-I(X^n;\mathbf{C}_A))\right). \qquad \text{(A.3)}\end{aligned}$$

Substituting this in (A.2), we get

$$\exp\left(-\frac{2}{n}I(X^n;\mathbf{C}_A)\right) \geq \frac{1}{2\pi e\sigma_{\widetilde{N}}^2}\exp\left(\frac{2}{n}(h(S^n|X^n,\mathbf{C}_A)-I(X^n;\mathbf{C}_A))\right)+\frac{\sigma_{\widetilde{N}}^2}{\sigma_X^2}. \quad \text{(A.4)}$$

Using the lower bound for $h(S^n|X^n,\mathbf{C}_A)$ from (A.1)

$$\exp\left(-\frac{2}{n}I(X^n;\mathbf{C}_A)\right) \geq \sigma_{\widetilde{N}}^2\left(\sum_{k\in A}\frac{\exp(-2r_k)}{\sigma_{N_k}^2}\right)\exp\left(-\frac{2}{n}I(X^n;\mathbf{C}_A)\right)+\frac{\sigma_{\widetilde{N}}^2}{\sigma_X^2}.$$

which can be rewritten as

$$\frac{1}{\sigma_X^2}\exp\left(\frac{2}{n}I(X^n;\mathbf{C}_A)\right) \leq \frac{1}{\sigma_X^2}+\sum_{k\in A}\frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}. \qquad \text{(A.5)}$$

Chapter 4

The Interactive CEO Problem

The problem setting is the same as in the previous chapter. However, we will not restrict our attention to the quadratic Gaussian as we did there. The setup is as follows: Let $X(t),\ t = 1, 2, \ldots$ be a sequence of i.i.d. random variables over $\mathcal{X}$ which the decoder (CEO) is interested in. But it cannot be observed directly. Instead, there are L sensors (encoders which cannot directly co-operate) who observe independently corrupted versions of this sequence. For $k = 1, \ldots, L$, sensor k observes $Y_k(t)$, a noisy version of the $X(t)$ sequence corrupted by an independent memoryless channel $p_{Y_k|X}$. The channels are independent in the sense that conditioned on $X(t)$, the observations $Y_k(t)$ and $Y_j(t)$ are independent for $j \neq k$.

In the CEO problem with sensor broadcast, sensors take turns and communicate over many rounds. Without loss of generality, we will assume that in each round the sensors communicate in order, i.e., sensor 1, sensor 2, …, sensor L. The k-th sensor's message C_k^s in round s may depend on Y_k^n and all the messages produced by all the sensors up till then.

$$C_k^s = f_k^{s,(n)}\left(Y_k^n, C_1^1, C_2^1, \ldots, C_L^1, C_1^2, C_2^2, \ldots, C_L^2, \ldots, C_1^{s-1}, \ldots, C_L^{s-1}, C_1^s, \ldots, C_{k-1}^s\right),$$

where $f_k^{s,(n)}$ is the corresponding encoding map. The rate of this message is defined as $|\mathcal{C}_k^{s,(n)}|/n$ where $\mathcal{C}_k^{s,(n)}$ is the range space of $f_k^{s,(n)}$. The decoder forms the estimate $\widehat{X}^n = g^{(n)}(\mathbf{C})$, where $\mathbf{C}$ stands for the collection of all messages and $g^{(n)}$ is a decoder map from the space of all messages to the reconstruction alphabet $\widehat{\mathcal{X}}$. The average distortion is

$$D^{(n)} = \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left(d(X(t), \widehat{X}(t))\right),$$

where d is the distortion function. A sum-rate and target distortion pair (R, D) is said to be *achievable for the CEO problem with sensor broadcast* if for any $\epsilon > 0$ there are encoders which work at this sum-rate and a decoder such that $D^{(n)} \le D + \epsilon$ for sufficiently large n. We will denote by $R_{\text{broadcast}}(D)$ the minimum sum-rate required to achieve a target distortion D, i.e., $R_{\text{broadcast}}(D)$ is the infimum of all R such that (R, D) is achievable for the CEO problem with sensor broadcast.

Similarly, we can define the achievable sum-rate and distortion pairs for the CEO problem, where the messages produced by sensors will depend only on their block of observations. Then there is no need for considering multiple rounds of communication. The achievable pairs for the CEO problem with sensor interaction can also be defined similarly. Here, unlike in the sensor broadcast setup, the messages are received by only one recipient (either another sensor or the CEO). A sensor can produce messages dependent only on its observations and the messages it has received. Here we need to consider multiple rounds of communication. Let $R_{\text{CEO}}(D)$ and $R_{\text{interaction}}(D)$ be the minimum sum-rate required to attain a target distortion of D in the CEO problem and the CEO problem with sensor interaction respectively. It is easy to observe the following fact from the definitions

$$R_{\text{CEO}}(D) \geq R_{\text{interaction}}(D) \geq R_{\text{broadcast}}(D). \tag{4.1}$$

We consider an example which shows that the inequalities above can be strict. Let $T(t)$ be a sequence of i.i.d. random variables uniformly distributed over $\{1, 2\}$, and $W_1(t), W_2(t)$ be two independent sequences of i.i.d. random variables independent of $T(t)$ and uniformly distributed over $\{1, 2, 3, 4\}$. Let the hidden source be $X(t) = (T(t), W_T(t))$. Sensor 1 observes $Y_1(t) = T(t)$, and sensor 2 observes $Y_2(t) = (W_1(t), W_2(t))$. Clearly, conditioned on X, Y_1 and Y_2 are independent. Let us suppose that the CEO wants to recover the hidden source losslessly. It is easy to observe that in the CEO problem this is possible only if both sensors send essentially all of their observations to the CEO. This means $R_{\text{CEO}} = 5$ bits. However, in the CEO problem with interaction, sensor 1 can send its observation (at a rate of 1 bit) to the second sensor which can then use this to send W_T at a rate of 2 bits. Either sensor can send the observation of sensor 1 to the CEO at a rate of 1 bit. Therefore, $R_{\text{interaction}} \leq 4$ bits. Also, a rate of 3 bits cannot be achieved in this setup since the entropy rate of X itself is 3 bits and if the sum-rate is equal to 3 bits, the rate at which sensors send messages to each other is 0. It is easy to argue that when the sensors do not communicate at a non-zero rate, the sum-rate must equal the sum-rate of the CEO problem which is 5 bits. But, a sum-rate of 3 is achievable under sensor broadcast. The first sensor broadcasts its observation at a rate of 1 bit and using this information the second sensor broadcasts W_T at a rate of 2 bits. Thus we have $R_{\text{broadcast}} = 3 < R_{\text{interaction}}$.

4.1 A general lower bound

Let us define the function $R(D)$ for $D \in \mathbb{R}_+$ as

$$R(D) = \min_{U,W} I(X;U) + \sum_{k=1}^{L} I(Y_k;U|X,W), \tag{4.2}$$

where W is an auxiliary random variable independent of $(X, \mathbf{Y})$ and U is an auxiliary random variable which satisfies the following Markov chains: $X - (\mathbf{Y}, W) - U$ and $Y_i - (X, U, W) - (Y_j,\ j \neq i),\ \forall i$. Moreover, we require that there exist a function $f : \mathcal{U} \to \widehat{\mathcal{X}}$ such that $\mathbb{E}[d(X, f(U))] \leq D$.

Theorem 4.1.1

$$R_{\text{broadcast}}(D) \geq R(D).$$

Proof: If (R, D) is achievable for the CEO problem with sensor broadcast, we have

$$\begin{aligned}
nR &\geq H(\mathbf{C}) \\
&\geq I(X^n, \mathbf{Y}^n; \mathbf{C}) \\
&= I(X^n; \mathbf{C}) + I(\mathbf{Y}^n; \mathbf{C}|X^n) \\
&\stackrel{\text{(a)}}{=} I(X^n; \mathbf{C}) + \sum_{k=1}^{L} I(Y_k^n; \mathbf{C}|X^n) \\
&= \sum_{t=1}^{n} \left(I(X(t); \mathbf{C}|X^{t-1}) + \sum_{k=1}^{L} I(Y_k(t); \mathbf{C}|Y_k^{t-1}, X^n) \right) \\
&\geq \sum_{t=1}^{n} \left(I(X(t); \mathbf{C}X^{t-1}) + \sum_{k=1}^{L} I(Y_k(t); \mathbf{C}X^{t-1}|X(t)W(t)) \right) \\
&= \sum_{t=1}^{n} \left(I(X(t); U(t)) + \sum_{k=1}^{L} I(Y_k(t); U(t)|X(t)W(t)) \right)
\end{aligned}$$

where $W(t) = (X(t+1), \ldots, X(n))$ and $U(t) = (\mathbf{C}, X^{t-1})$. (a) follows from the fact that given $(X^n, \mathbf{C})$, the sensor observations $Y_k^n,\ k = 1, \ldots, L$ are independent. This is easy to verify along the lines of footnote 1 below. Note that $W(t)$ is independent of $(X(t), \mathbf{Y}(t))$. $U(t)$ satisfies the Markov conditions:[1] $X(t) - (\mathbf{Y}(t), W(t)) - U(t)$ and $Y_i(t) - (X(t), U(t), W(t)) - (Y_j(t),\ j \neq i),\ \forall i$. We take $f_t(U(t)) = f_t(\mathbf{C}, X^{t-1}) \stackrel{\text{def}}{=}$

[1]To see that the Markov chains hold, let us restrict ourselves to m rounds of transmissions. We have the joint distribution of (X^n, Y^n, U^n, W^n), or equivalently the joint distribution of $(X^n, Y^n, \mathbf{C})$ given by

$$\begin{aligned}
p(x^n, \mathbf{y}^n, \mathbf{c}) &= p(x^{t-1})p(x(t))p(x_{t+1}^n) \\
&\times \prod_{k=1}^{L} \left(\left(\prod_{r=1}^{n} p(y_k(r)|x(r)) \right) \left(\prod_{s=1}^{m} p(c_k^s|y_k^n, c_1^1, c_2^1, \ldots, c_L^1, c_1^2, c_2^2, \ldots, c_L^2, \ldots, c_1^{s-1}, \ldots, c_L^{s-1}, c_1^s, \ldots, c_{k-1}^s) \right) \right). \quad (4.3)
\end{aligned}$$

Let $\tilde{y}_k(t) = (y_k(1), y_k(2), \ldots, y_k(t-1), y_k(t+1), \ldots, y_k(n))$. The right hand side of (4.3) is of the form $f_1(x(t), \mathbf{y}(t)) f_2(u(t), \mathbf{y}(t), \tilde{\mathbf{y}}(t), w(t))$. This implies that conditioned on $(\mathbf{Y}(t), W(t))$, $X(t)$ and $U(t)$ are independent. In other words, $X(t) - (\mathbf{Y}(t), W(t)) - U(t)$ is a Markov chain. We can also

$g_t^{(n)}(\mathbf{C})$. The theorem now follows from the following lemma which is proved in appendix A.

Lemma 4.1.2 *$R(D)$ is a convex function of D.*

□

It is interesting to observe that this lower bound subsumes the sensor co-operation lower bound which is obtained by allowing all the sensors to pool their observations and act as one encoder. It is easy to show that the co-operation lower bound is given by

$$R_{\text{co-op}}(D) = \min_{\hat{X}} I(\mathbf{Y}; \hat{X}),$$

where the minimisation is over all $\hat{X}$ such that $X - \mathbf{Y} - \hat{X}$ is a Markov chain and $\mathbb{E}[d(X, \hat{X})] \leq D$. For any (U, W, f) satisfying the constraints following (4.2), let $\hat{X} = f(U)$. We can verify that $X - \mathbf{Y} - U$ is a Markov chain. Clearly, $X - \mathbf{Y} - \hat{X}$

---

write the right hand side of (4.3) in the form:

$$h(x(t), u(t), w(t)) \left(\prod_{k=1}^{L} p(y_k(t)|x(t)) p(\tilde{y}_k(t)|u(t), w(t)) g_k(y_k(t), \tilde{y}_k(t), u(t)) \right)$$

This implies that $Y_i(t) - (X(t), U(t), W(t)) - (Y_j(t),\ j \neq i)$ is a Markov chain.

is a Markov chain and $\mathbb{E}[d(X,\hat{X})] \le D$. Also,

$$\begin{aligned}
I(\mathbf{Y};\hat{X}) &\le I(\mathbf{Y};U), && \text{by data processing inequality} \\
&= I(X;U) + I(\mathbf{Y};U|X), && \text{since } X - \mathbf{Y} - U \text{ is a Markov chain} \\
&\le I(X;U) + I(\mathbf{Y};U|X,W), && \text{since } W \text{ is indep. of } (X,\mathbf{Y}) \\
&= I(X;U) + \sum_{k=1}^{L} I(Y_k;U|X,W), && \text{since } Y_i - (X,U,W) - Y_j \text{ and} \\
& && Y_i - X - Y_j \text{ are Markov chains if } i \ne j.
\end{aligned}$$

Thus the general lower bound subsumes the co-operation lower bound. Therefore, we can interpret this lower bound as a stronger form of the co-operation lower bound where the forms of co-operation allowed is restricted by the Markov condition $Y_i - (X,U,W) - (Y_j, j \ne i)$.

The intuition behind this lower bound is the observation that the conditional independence of Y_i^n and Y_j^n for $i \ne j$ is preserved even if we condition on $(X^n, \mathbf{C})$ instead of X^n. This might suggest that a single letter lower bound of the form given in (4.2) with an empty W holds. However, this cannot be shown. To translate our observation into a single letter bound we have to introduce the auxiliary random variable W. Note that this is similar to the lower bound to rate-distortion functions for multi-terminal source coding derived by Wagner and Anantharam [54].

This lower bound is not tight in general[2]. However, it is tight for the important Gaussian model discussed below.

4.2 The quadratic Gaussian case

In the Gaussian case, $X(t)$ is a sequence of i.i.d. Gaussian random variables with mean 0 and variance σ_X^2. The k-th sensor observes $Y_k(t) = X(t) + N_k(t)$, where $N_k(t)$ is an i.i.d. sequence of additive Gaussian noise random variables with mean 0 and variance $\sigma_{N_k}^2$. Additive noise processes are independent across sensors. The distortion function is squared error, $d(x, \hat{x}) = (x - \hat{x})^2$. From the previous chapter, we can deduce that the sum-rate distortion trade-off for the CEO problem is [cf. (3.12)]

$$R_{\mathrm{CEO}}(D) = \frac{1}{2} \log \frac{\sigma_X^2}{D} + \sum_{k=1}^{L} r_k, \tag{4.4}$$

where $r_k \geq 0,\ k = 1, \ldots, L$, satisfy

$$\frac{1}{\sigma_X^2} + \sum_{k=1}^{L} \frac{1 - \exp(-2r_k)}{\sigma_{N_k}^2} = \frac{1}{D}.$$

[2]It is easy to construct an example where the lower bound is not tight. As noted above, the intuition behind our lower bound is that by restricting the forms of co-operation between sensors we can try to tighten the co-operation lower bound. The restriction we introduced is that conditioned on $(X^n, \mathbf{C})$, Y_i^n and Y_j^n are independent for $i \neq j$. We can defeat this restriction by redefining the problem so that the hidden source is made up of both X and the observations $\mathbf{Y}$, and the distortion function on the new composite source can be defined to take into account only the distortion on X. Clearly, this is a valid problem which is in effect no different from the original problem, but now the restriction that Y_i^n and Y_j^n be conditionally independent conditioned on $(X^n, \mathbf{C})$ becomes trivially true. Thus we have rendered the new lower bound equivalent to the co-operation lower bound which we know is too weak in general. This pathology notwithstanding, we believe that the lower bound is useful for appropriately defined problems.

To show that our lower bound is tight, we note that

$$\begin{aligned} I(X;U) \geq I(X;f(U)) &\geq h(X) - h(X - f(U)) \\ &\geq \frac{1}{2}\log 2\pi e \sigma_X^2 - \frac{1}{2}\log 2\pi e D = \frac{1}{2}\log\frac{\sigma_X^2}{D}. \end{aligned} \tag{4.5}$$

Consider the following lemma which is proved in appendix B

Lemma 4.2.1 *For auxiliary random variables satisfying the Markov conditions specified in the definition of $R(D)$,*

$$\frac{1}{\sigma_X^2}\exp\left(2I(X;U)\right) \leq \frac{1}{\sigma_X^2} + \sum_{k=1}^{L}\frac{1-\exp(-2r_k)}{\sigma_{N_k}^2},$$

where $r_k \overset{def}{=} I(Y_k;U|X,W)$.

From (4.5) and lemma 4.2.1 we observe that our lower bound $R(D)$ satisfies $R(D) \geq R_{\text{CEO}}(D)$. Since $(R_{\text{CEO}}(D), D)$ is achievable for the CEO problem (and therefore, for the CEO problem with sensor interaction) we have the following theorem

Theorem 4.2.2 *For the Gaussian case*

$$R_{\text{broadcast}}(D) = R_{\text{CEO}}(D).$$

Thus the bound is tight for the Gaussian case. Moreover, the optimal strategy is for the sensors to encode their observations independent of the messages from the other sensors. That is, the optimal strategy for the CEO problem (without sensor broadcast) is also optimal for the CEO problem with sensor broadcast. This implies that for the Gaussian case, no reduction in sum-rate can be obtained by allowing the sensors to broadcast messages.

An interesting consequence of this theorem which follows from (4.1) is that for the Gaussian case

Corollary 4.2.3

$$R_{\text{broadcast}}(D) = R_{\text{interaction}}(D) = R_{\text{CEO}}(D).$$

Thus, for the Gaussian case, the availability of inter-sensor communication does not lead to a reduction over the CEO setup in the sum-rate required for achieving a target mean squared error distortion at the CEO. The optimal strategy of communication remains unchanged from the Gaussian quantisation and binning strategy which is known to be optimal for the Gaussian CEO problem.

4.3 Discussion

It is easy to show that for the Gaussian problem with sensor communication (interaction or broadcast), if we restrict ourselves to the class of Gaussian quantisation (i.e., Gaussian auxiliary random variables) based Berger-Tung-Housewright strategies, no gain can be expected over the CEO setup. This is a consequence of the no-rate loss property [56] of Gaussian source coding with a jointly Gaussian side information at the decoder and squared error distortion function. The property is that the rate-distortion behaviour does not change even if the side information is made available to the encoder. Under Gaussian quantisation strategies, whether or not the sensors have access to messages from the other sensors does not affect the sum-rate at which the sensors have to send their messages (though it could affect their coding strategy) for producing the same fidelity at the decoder. However, this argument is not enough to prove the result in the previous section. We cannot rule out the possibility that by deviating from the Gaussian quantisation strategy we might gain from taking advantage of the inter-sensor communication and there is some trade-off between this advantage and the loss from using non-Gaussian quantisation. The converse proof asserts that no such trade-off exists and no achievable scheme for the Gaussian problem with sensor communication can out do an optimal coding scheme for the CEO setup.

The results hold even when we allow the CEO to have an observation $Z(t)$ governed by a memoryless channel $p_{Z|X,\mathbf{Y}}$ such that conditioned on (Z, X), the observations at the sensors $Y_1, \ldots, Y_L$ are independent. In the Gaussian version, this would include the case where $Z(t) = X(t) + N(t)$, where $N(t)$ is independent of $X(t)$ and all $N_k(t)$. Now we need to allow the CEO also to broadcast messages to all the sensors since this may, in general, lead to a reduction in the sum-rate. Under this setup a modified form of the lower bound can be derived along the same lines as Theorem 4.1.1. It can be shown that

$$R_{\text{broadcast}}(D) \geq \min_{U,W} I(X;U|Z) + \sum_{k=1}^{L} I(Y_k;U|X,Z,W),$$

where W is now independent of $(X, \mathbf{Y}, Z)$, and U satisfies the Markov chains: $(X, Z)-(\mathbf{Y}, W) - U$ and $Y_i - (X, U, W) - Y_j$ for all $i \neq j$. There should also exist a function $f : \mathcal{U} \times \mathcal{Z} \to \widehat{\mathcal{X}}$ such that $\mathbb{E}\left[d(X, f(U, Z))\right] \leq D$. For the Gaussian case it can also be shown as in section 4.2 that an optimal strategy will not exploit either the availability of inter-sensor communication or communication from the CEO. In this sense, we can view our result as a non-trivial generalisation of the no-rate loss property of Gaussian source coding with jointly Gaussian side information at the decoder and squared error distortion function.

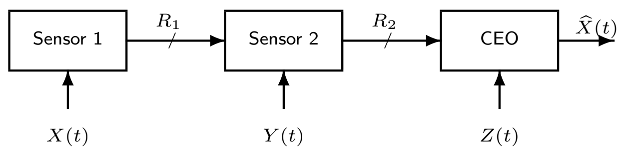


Figure 4.1: An Open Problem

It would be interesting to find the rate region of the problems where interaction is possible. This problem remains open. Perhaps one of the simpler problems in this class which is open even in the quadratic Gaussian setting is the one shown in Fig. 4.1. Here the CEO is interested in recovering the source $X(t)$ under a distortion function $d(.,.)$. Sensor 1 observes $X(t)$ noiselessly, but has no direct link to the CEO. Sensor 1 can communicate at rate R_1 to the intermediate sensor 2 who makes a noisy observation $Y(t)$ of source. Sensor 2 can communicate to the CEO at a rate R_2. The CEO itself has a noisy observation $Z(t)$ of the source. We also impose conditional independence of $Y(t)$ and $Z(t)$ conditioned on $X(t)$. This is in fact a special case of a problem in [15] which the authors call the "serial network." The achievable solution suggested by the authors is of the following form: Any (R_1, R_2, D) satisfying the following constraints is achievable

$$
\begin{aligned}
R_1 &\geq I(X;U) - I(Y;U), \\
R_2 &\geq I(Y,U;V) - I(Z;V) \\
D &= \mathbb{E}\left[d(X, f(Z,V))\right],
\end{aligned} \tag{4.6}
$$

where U, V are random variables satisfying the Markov chains $U - X - (Y, Z)$ and $V - (Y,U) - (X, Z)$, and f is some function whose range is the reproduction alphabet of the decoder. Intuitively, the solution above works as follows: Sensor 1 quantises the source to U-codeword and sends a bin index at rate R_1 so that sensor 2 can decode the U-codeword using its observation Y. Now, sensor 2 quantises all the information it has (namely U and Y) to a V-codeword and sends its bin index at rate R_2 so that the CEO can decode making use of its observation Z. Now the CEO forms an estimate $\hat{X}$ of the source using all the information it has, namely V and Z by computing $\hat{X} = f(Z, V)$.

In fact, an improved achievable solution is possible. We first note that when $R_1 \geq R_2$, the problem can be completely solved. The best possible distortion that is achievable under this condition is given by the optimal distortion of a Wyner-Ziv [57] problem of compressing the source $X(t)$ at a rate R_2 when the receiver has has side information $Z(t)$. This is given by (cf. (2.1))

$$
D_{X|Z}(R_2) \stackrel{\text{def}}{=} \inf \mathbb{E}\left[d(X, f(Z,U))\right],
$$

where the infimisation is over all random variables U satisfying the conditions stated after (2.1) and such that $I(X;U|Z) \leq R_2$. The optimal scheme involves sensor 1 performing the above Wyner-Ziv compression and sending the compressed bits, and the intermediate node simply forwarding them on to the CEO. The optimality becomes clear by noting that in the original problem, by giving sensor 1's observation to sensor 2 we can only decrease the value of the smallest achievable distortion and it is a simple matter to argue from Wyner and Ziv's converse that $D_{X|Z}(R_2)$ is indeed the smallest achievable distortion for this modified problem. However, the solution in (4.6) does not include this in general. This is because of the requantisation step at the intermediate sensor. In general, the intermediate sensor fails to take advantage of the fact that U-codeword is already quantised and treats it like an observation. This suggests that the following achievable solution.

$$\begin{aligned} R_1 &\geq I(X;U) - I(Y;U), \\ R_2 &\geq (I(X;U) - I(Z;U)) + (I(Y;V|U) - I(Z;V|U)), \\ D &= \mathbb{E}\left[d(X, f(Z,U,V))\right], \end{aligned} \tag{4.7}$$

where U and V satisfy the same Markov chains as earlier, and f is now a function of U as well. Here the intermediate sensor, upon decoding the U-codeword proceeds to send a bin index to the CEO so that the CEO may also decode the U-codeword making use of its observation Z. The rate at which the intermediate sensor needs to send this bin index for successful decoding by the CEO is $I(X;U) - I(Z;U) + \epsilon$. This is the first term in the expression for the lower bound for R_2 in the achievable solution.

Then the intermediate sensor quantise its observation Y to a V-codeword using a conditional codebook conditioned on the U-codeword which is now available at both the CEO and the intermediate sensor. A bin index of this codeword is sent at rate $I(Y;V|U) - I(Z;V|U)$ so that the CEO may use its knowledge of Z and U-codeword to decode the V-codeword. This strategy covers the $R_1 \geq R_2$ case mentioned above when it is optimal. But, in general, this may not subsume the region given by (4.6). However, for the quadratic Gaussian version of the problem, with jointly Gaussian choices for U, V, it is a simple computation to show that the strategy in (4.7) indeed subsumes the one in (4.6). However, whether or not the resulting achievable region for (R_1, R_2, D) is optimal remains open.

Appendix A

Proof of Lemma 4.1.2

To prove that $R(D)$ is a convex function of D, let $(p_{W_1}(w_1)p_{U_1|\mathbf{Y},W_1}(u_1|\mathbf{y},w_1), f_1)$ and $(p_{W_2}(w_2)p_{U_2|\mathbf{Y},W_2}(u_2|\mathbf{y},w_2), f_2)$ be the minimisers for target distortions D_1 and D_2 respectively. In order to satisfy the second Markov condition, the joint distributions should be of the form

$$p(x,\mathbf{y},w_i,u_i) = p(x)p(\mathbf{y}|x)p_{W_i}(w_i)p_{U_i|\mathbf{Y},W_i}(u_i|\mathbf{y},w_i) = \prod_{k=1}^{L} g_i^{(k)}(x,y_k,w_i,u_i).$$

Also $\mathbb{E}_i[d(X,f_i(U_i)] \leq D_i$.

We can take (W_1,U_1) and (W_2,U_2) to be conditionally independent given $(X,\mathbf{Y})$. Let us define $W=(S,W')$, where S is a random variable independent of $(X,\mathbf{Y},W_1,U_1,W_2,U_2)$ which takes on value 1 with probability α and 2 with probability $1-\alpha$ and $W'=W_S$. Let U be defined as $U=(S,U')$ where $U'=U_S$. Then

$$p_W((s,w'))p_{U|\mathbf{Y},W}((s,u)|\mathbf{y},(s,w)) = p_S(s)p_{W_s}(w')p_{U_s|\mathbf{Y},W_s}(u'|\mathbf{y},w').$$

To ensure that (W, U) defined above is a valid choice, we should show that conditioned on (X, U, W), Y_k's are independent.

$$\begin{aligned} p_{X,\mathbf{Y},W,U}(x, \mathbf{y}, (s, w'), (s, u')) &= p(x)p(\mathbf{y}|x)p_S(s)p_{W_s}(w')p_{U_s|\mathbf{Y},W_s}(u'|\mathbf{y}, w') \\ &= p_S(s)\prod_{k=1}^{L} g_s^{(k)}(x, y_k, w', u') \end{aligned}$$

We define $f((s, u')) = f_s(u')$. The resulting distortion is

$$\mathbb{E}[d(X, f(U))] = \alpha D_1 + (1-\alpha)D_2.$$

It is easy to verify that

$$I(X;U) + \sum_{k=1}^{L} I(Y_k; U|X, W) = \alpha R(D_1) + (1-\alpha)R(D_2).$$

This completes the proof of the lemma. □

Appendix B

Proof of Lemma 4.2.1

The proof is substantially similar to the proof of lemma 3.1.2. The main observation is that conditioned on the hidden source and all the messages produced by the sensors under the sensor broadcast setting, the observations of the sensors are conditionally independent. This allows us to continue to have the first application of entropy power inequality in the proof of lemma 3.1.2. The lemma here is in a single-letter form. The proof is given below for completeness.

Let the MMSE estimate of X from $\mathbf{Y}$ be denoted by S. Then $S = \sum_{k=1}^{L} \frac{\sigma_{\widetilde{N}}^2}{\sigma_{N_k}^2} Y_k(t)$, where $1/\sigma_{\widetilde{N}}^2 = 1/\sigma_X^2 + \sum_{k=1}^{L} 1/\sigma_{N_k}^2$. i.e.,

$$X = S + \widetilde{N},$$

where $\widetilde{N}$ is i.i.d Gaussian with mean zero, variance $\sigma_{\widetilde{N}}^2$, and $\widetilde{N}$ is independent of $\mathbf{Y}$.

Recall that, given (X, U, W), $Y_1, \ldots, Y_L$ are independent, Then, by entropy power inequality (EPI),

$$h(S|X=x,U=u,W=w) \geq \frac{1}{2}\log\left(\sum_{k=1}^{L}\exp\left(2h\left(\frac{\sigma_{\widetilde{N}}^2}{\sigma_{N_k}^2}Y_k\middle|X=x,U=u,W=w\right)\right)\right)$$

We now take expectation over (X,U,W). Since log-sum-exp function is convex, we may apply Jensen's inequality to the right-hand side. This gives

$$\begin{aligned} h(S|X,U,W) &\geq \frac{1}{2}\log\left(\sum_{k=1}^{L}\exp\left(2h\left(\frac{\sigma_{\widetilde{N}}^2}{\sigma_{N_k}^2}Y_k\middle|X,U,W\right)\right)\right) \\ &= \frac{1}{2}\log\left(\sigma_{\widetilde{N}}^4\sum_{k=1}^{L}\frac{1}{\sigma_{N_k}^4}\exp\left(2h(Y_k|X,W)-2I(Y_k;U|X,W)\right)\right) \\ &= \frac{1}{2}\log\left(\sigma_{\widetilde{N}}^4\sum_{k=1}^{L}\frac{2\pi e\sigma_{N_k}^2}{\sigma_{N_k}^4}\exp\left(-2I(Y_k;U|X,W)\right)\right), \end{aligned}$$

where we used the fact that W is independent of $(X,\mathbf{Y})$.

$$\Rightarrow \qquad \exp\left(2h(S|X,U,W)\right) \geq 2\pi e\sigma_{\widetilde{N}}^4\left(\sum_{k=1}^{L}\frac{\exp(-2r_k)}{\sigma_{N_k}^2}\right), \tag{B.1}$$

where we used the definition of r_k from lemma 4.2.1, $r_k = I(Y_k;U|X,W)$. Now recall that $X = S + \widetilde{N}$, where S and $\widetilde{N}$ are independent. From the Markov conditions it is easy to see that conditioned on $(U=u, W=w)$, S and $\widetilde{N}$ are still independent and the marginal distribution of $\widetilde{N}$ remains the same. Applying EPI,

$$\begin{aligned} & \exp\left(2h(X|U=u,W=w)\right) \geq \exp\left(2h(S|U=u,W=w)\right) + \exp\left(2h(\widetilde{N})\right) \\ \Rightarrow \quad & h(X|U=u,W=w) \geq \frac{1}{2}\log\left(\exp\left(2h(S|U=u,W=w)\right) + 2\pi e\sigma_{\widetilde{N}}^2\right) \\ \overset{(a)}{\Rightarrow} \quad & h(X|U,W) \geq \frac{1}{2}\log\left(\exp\left(2h(S|U,W)\right) + 2\pi e\sigma_{\widetilde{N}}^2\right) \\ \Rightarrow \quad & \exp\left(2h(X|U,W)\right) \geq \exp\left(2h(S|U,W)\right) + 2\pi e\sigma_{\widetilde{N}}^2, \end{aligned}$$

where (a) follows from Jensen's inequality and convexity of $1/2\log\left(e^{2x}+k\right)$ in x for $k \geq 0$. We can rewrite the above as

$$\begin{aligned}\exp\left(-2I(X;U,W)\right) &= \exp\left(-2h(X)\right)\exp\left(2h(X|U,W)\right)\\ &\geq \frac{1}{2\pi e\sigma_X^2}\exp\left(2h(S|U,W)\right)+\frac{\sigma_{\tilde{N}}^2}{\sigma_X^2}. \end{aligned} \tag{B.2}$$

Also

$$h(S|U,W) = h(S|X,U,W) + I(X;S|U,W),$$

and

$$I(X;S|U,W) = I(X;S,U,W) - I(X;U,W) = I(X;S) - I(X;U,W),$$

since $X - S - \mathbf{Y} - (U,W)$ is a Markov chain.

$$\begin{aligned}\Rightarrow \exp\left(2h(S|U,W)\right) &\geq \exp\left(2(h(S|X,U,W)+I(X;S)-I(X;U,W))\right)\\ &= \frac{\sigma_X^2}{\sigma_{\tilde{N}}^2}\exp\left(2(h(S|X,U,W)-I(X;U,W))\right). \end{aligned} \tag{B.3}$$

Substituting this in (B.2), we get

$$\exp\left(-2I(X;U,W)\right) \geq \frac{1}{2\pi e\sigma_{\tilde{N}}^2}\exp\left(2(h(S|X,U,W)-I(X;U,W))\right)+\frac{\sigma_{\tilde{N}}^2}{\sigma_X^2}. \tag{B.4}$$

Using the lower-bound for $h(S|X,U,W)$ from (B.1)

$$\exp\left(-2I(X;U,W)\right) \geq \sigma_{\tilde{N}}^2\left(\sum_{k=1}^{L}\frac{\exp(-2r_k)}{\sigma_{N_k}^2}\right)\exp\left(-2I(X;U,W)\right)+\frac{\sigma_{\tilde{N}}^2}{\sigma_X^2}.$$

which can be rewritten as

$$\frac{1}{\sigma_X^2}\exp\left(2I(X;U,W)\right) \leq \frac{1}{\sigma_X^2}+\sum_{k=1}^{L}\frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}. \tag{B.5}$$

Now we use $I(X;U) \leq I(X;U,W)$ to write

$$\frac{1}{\sigma_X^2} \exp\left(2I(X;U)\right) \leq \frac{1}{\sigma_X^2} + \sum_{k=1}^{L} \frac{1-\exp(-2r_k)}{\sigma_{N_k}^2}. \tag{B.6}$$

□

Part II

Source-Channel Coding: Gaussian Sources over Gaussian Channels

Chapter 5

Introduction

Practical analog audio-visual transmission systems (e.g., AM/FM radio and NTSC TV) are highly spectrum-inefficient. This inefficiency is primarily due to the fact that typical audio-visual data samples are redundant in the sense that they have high memory (spatial and temporal), but there are no efficient analog compression techniques. Further, digital transmission techniques are significantly more spectrum-efficient than their analog counterparts. This has spurred the interest in digital compression and transmission including the popularity of digital formats like MPEG-2 [25] and MP3 [26] for video and audio and perhaps the Federal Communications Commission (FCC) mandate to migrate television to all-digital transmission by year 2009. In this part, we address two questions arising from the broadcast of a common source to multiple receivers.

Our first question relates to the migration from an analog transmission system to an all-digital transmission system. One complication arises in this context is that of backward compatibility with the existing wide base of legacy analog systems. The current FCC solution is to have simulcast of identical broadcast content on separate analog and digital channels, which is however spectrally wasteful due to the duplication of identical signal content without exploiting the fact that the analog version can constitute valuable side-information for the digital stream. In the next chapter, we ask the question, what is the optimal way to seamlessly broadcast sources with memory? We would like seamlessness in two regards: (i) no extra digital spectrum; and (ii) the best possible *simultaneously* delivered qualities of reproduction of the source at a legacy analog receiver and a digital receiver. Motivated by their numerous modelling successes for multimedia data [3, 20, 23] as well as for various transmission channels of interest, we consider a parallel Gaussian model for the source and a non-degraded Gaussian broadcast model for the channel respectively, along with a linear model (in a sense to be described) for the legacy analog receiver.

For the case of a white Gaussian source and white Gaussian point-to-point channel, Goblick [18] recognised that the uncoded transmission strategy of merely scaling individual source samples so as to meet the average encoder power constraint, followed by optimal linear minimum mean-squared error (MMSE) estimation of the source samples from the channel observations at the receiver, results in the optimal delivered quality (in a mean-squared error sense). Since this receiver is also an analog receiver as we define later, and the encoding strategy does not depend on the channel noise variance, this analog (uncoded) transmission is optimal for our problem in the context of white source and white channel. However, when the source has memory, it can be shown that the optimal analog strategy is sub-optimal compared to the best digital (coded) strategy in general. Berger and Tufts [6] considered the problem of uncoded transmission of a source using pulse-amplitude modulation (PAM) with the receiver constrained to be a linear-time invariant (LTI) filter. They derived the optimal PAM pulse-shape and the optimal LTI filter. Through examples they demonstrate the gap between the optimal uncoded and coded schemes. In fact, in a point-to-point set-up, for a parallel Gaussian source with m source components and an AWGN channel model with an equal number of sub-channels, the loss in performance of the analog approach with respect to the digital approach for sufficiently large transmit powers can be shown to be

$$\frac{\text{Analog MSE Distortion}}{\text{Digital MSE Distortion}} = \left(\frac{(\sigma_1 + \sigma_2 + \ldots + \sigma_m)/m}{(\sigma_1\sigma_2 \ldots \sigma_m)^{1/m}}\right)^2 \tag{5.1}$$

where σ_j^2 denotes the variance of the j-th source component. We will show this simple fact in the next chapter. This gap can therefore be arbitrarily large.

Shamai, Verdú, and Zamir [40] also considered the problem of digital upgrade of analog transmission. In contrast to our setup, extra bandwidth is made available for the exclusive use by a digital receiver, but the analog transmission on the existing band is not allowed to be altered. They proposed *systematic* source-channel codes based on Wyner-Ziv coding and show that it is optimal for the case of a white Gaussian source and channel under the above stated restriction.

The main result in the next chapter is a complete characterisation of the trade-offs between the achievable distortion pairs (at the the analog and digital receivers) given a power constraint at the transmitter which is otherwise unconstrained. An interesting consequence of this result is the existence of an operating point where the digital receiver obtains the classical point-to-point optimal quality (i.e., the best quality achieved when it is the only intended receiver) and the analog receiver seamlessly co-exists with the digital receiver attaining the best possible simultaneously achievable quality.

Further, we show that a constructive scheme consisting of the cascade of source coding with side information (based on Wyner-Ziv (WZ) coding [57]) and channel coding with side information (based on Gel'fand-Pinsker (GP) coding [17]) systems can be used to come up with a hybrid coded-uncoded strategy that can achieve the entire quality of reproduction trade-off region associated with the problem.

In chapter 7, we address our second question which relates to another form of heterogeneity of receivers, namely, the quality of the channel experienced by the receivers. The objective is to devise an encoding strategy which utilises a common power and channel resource to simultaneously satisfy a heterogeneous set of receivers who are experiencing different channels from the transmitter.

For the case of an i.i.d. Gaussian source communicated over a scalar additive white Gaussian noise broadcast channel operating at the same symbol rate (in other words, when the memoryless source and channel have the same bandwidth), the observation of Goblick mentioned above gives a rather simple all-analog solution. The solution composed of merely transmitting individual source samples scaled so as to meet the average encoder power constraint, followed by optimal linear minimum mean-squared estimation of the source samples from the channel observations at individual receivers, results in the best delivered quality (in a mean squared sense) at both the receivers. This illustrates an interesting feature of analog methods – their ability to enable simultaneous enjoyment of the power and bandwidth resource by each of the broadcast receivers.

On the other hand, the obvious digital approach to the above problem is to use the classical separation method of scalable source coding [16] followed by broadcast channel coding [12]. In this approach, the coarse source layer is communicated as a common message intended for all users and the refinement layer is communicated only to some of the users. Assuming a scalar Gaussian broadcast channel [14, pg. 427] with two receivers, the one with the lower noise variance called *strong* and other *weak*, while the common portion of the information, being limited by the weak receiver, is sub-optimal for the strong receiver, the refinement portion is completely unusable by the weak receiver and in fact acts as interference to it. Thus unlike the analog methods, the digital approach necessarily involves a “splitting” of the total system resource.

While the above discussion illustrates the power of analog methods, as pointed out earlier, real-world sources are characterised by a high degree of memory, and thus they are far from the i.i.d. model. As we mentioned earlier, parallel source models describe these sources more effectively than an i.i.d. model. For this case, as we saw in (5.1), analog transmission is sub-optimal, with the sub-optimality growing with increased source memory.

This motivates the main question posed in chapter 7: what is an efficient way to broadcast Gaussian sources with memory over memoryless Gaussian broadcast channels? Our solution is driven by aiming to extract the best of both the analog and the digital worlds. We do this by invoking a hybrid uncoded-coded strategy, where the coded system uses a combination of the tools of successive refinement source coding [16], source coding with side-information (SCSI) or Wyner-Ziv (WZ) coding [57], super-position broadcast channel coding [12], and channel coding with side-information (CCSI) or Gel'fand-Pinsker (GP) [17] coding. Our hybrid approach is driven by the intuition that in a source-channel broadcast to a heterogeneous set of receivers, like in our setting, analog methods have the inherent ability to enable simultaneous enjoyment of the total encoder power resource by each of the broadcast receivers. In contrast, digital methods necessarily involve a splitting of the total system power resource. We would like to point out that this is an open problem in general and we present an achievable strategy which constitutes the best known solution in the literature.

A special case of our problem is the question of communicating a memoryless Gaussian source over a memoryless Gaussian broadcast channel when the source and channel bandwidths are mismatched. When the channel bandwidth is less than the source bandwidth, it is called a *bandwidth contraction* problem, and when the reverse is true, we call it a *bandwidth expansion* problem. Note that in the television broadcast application we mentioned earlier, typically the bandwidth contracts rather than expands. These problems have been addressed by a few researchers [32, 38, 40]. Our achievable solution when specialised to these settings improve the methods proposed in the first two. We would also like to point out that for the special case of the bandwidth expansion, the same achievable solution was independently arrived at in [38].

Chapter 6

The Digital Upgrade Problem

6.1 Problem formulation

6.1.1 Notation

Upper case letters (e.g., X) denote random variables/sets while bold-face letters (e.g., $\mathbf{x}$) denote vectors. For $l > 0$, let $I_l = \{1, 2, \ldots l\}$. The l-fold Cartesian product of a set E with itself is denoted as E^l. An l-sequence (a vector of size l) is denoted in relation to its elements as

$$\mathbf{x}^{(l)} = (x_1, x_2, \ldots x_l) = x_{I_l}$$

Further, $\mathbf{x}^{(m)(l)}$ jointly denotes the block of m l-sequences $\{\mathbf{x}_1^{(l)}, \mathbf{x}_2^{(l)}, \ldots, \mathbf{x}_m^{(l)}\}$.

6.1.2 Problem Set-up

As shown in Figure 6.1, we are given a zero mean parallel Gaussian source $\mathbf{S}_{I_m}$ with m (independent) components. $K_{\mathbf{S}} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ denotes the correlation matrix for the source. The coloured Gaussian broadcast channel comprises m sub-channels, with the digital receiver associated with a zero mean additive parallel Gaussian noise $\mathbf{Z}_{d_{I_m}}$, and the analog receiver with noise $\mathbf{Z}_{a_{I_m}}$. Further, let the covariance matrices $K_{\mathbf{Z}_i}$ of the noises be diagonal matrices with diagonal elements $N_{i1}, N_{i2}, \ldots, N_{im}$ for $i \in \{d, a\}$. The source and channels are independent over realisations. Let $\delta : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ denote the mean-squared error distortion measure. Thus, $\delta(a, b) = (a - b)^2$. The distortion measure on a pair of l-sequences $(\mathbf{a}^(l), \mathbf{b}^l) \in \mathbb{R}^l \times \mathbb{R}^l$ is defined by the average per-symbol distortion

$$\delta(\mathbf{a}^{(l)}, \mathbf{b}^{(l)}) \stackrel{\text{def}}{=} \frac{1}{l} \sum_{j=1}^{l} \delta(a_j, b_j) = \frac{1}{l} ||\mathbf{a}^{(l)} - \mathbf{b}^{(l)}||^2. \tag{6.1}$$

The distortion measure on a pair of blocks of m l-sequences $(\mathbf{c}^{(m)(l)}, \mathbf{d}^{(m)(l)} \in \mathbb{R}^{ml} \times \mathbb{R}^{ml}$ is defined by the sum of the distortions on l-sequences

$$\delta(\mathbf{c}^{(m)(l)}, \mathbf{d}^{(m)(l)}) \stackrel{\text{def}}{=} \sum_{j=1}^{m} \delta(\mathbf{c}_j^{(l)}, \mathbf{d}_j^{(l)}). \tag{6.2}$$

A broadcast source-channel code of block-length l, parametrised by (P, Δ_d, Δ_a), is defined by a triple (F, G_d, G_a) where $F : \mathbb{R}^{ml} \to \mathbb{R}^{ml}$ denotes the encoding function which maps a block of l realisations of the source into a channel input block of length l, i.e., $\mathbf{X}_{I_m}^{(l)} = F(\mathbf{S}_{I_m}^{(l)}))$. Two decoding functions are denoted by $G_i : \mathbb{R}^{ml} \to \mathbb{R}^{ml}$, $i \in \{d, a\}$. The analog decoding function G_a is restricted to be a block-size-m

memoryless mapping. In other words, the reconstruction $\hat{\mathbf{S}}_{a_{I_m}}$ of the analog receiver for a realisation of the source is the corresponding channel output $\mathbf{Y}_{a_{I_m}}$ pre-multiplied by an $m \times m$ diagonal matrix $\mathbf{H}$.

$$\hat{\mathbf{S}}_{a_{I_m}} = \mathbf{H}\mathbf{Y}_{a_{I_m}}.$$

This definition is motivated by the case of a stationary source and channel setting. In the limit of large m, the analog receiver as defined here, corresponds to an LTI (linear, time-invariant) filter with the diagonal elements of $\mathbf{H}$ representing the Fourier coefficients of the filter. G_d, on the other hand, has no such restrictions and can be any function of the block-length l channel output at the digital receiver.

The average system power available per symbol is P, i.e.,

$$\mathbb{E}\left[\frac{1}{l}||F(\mathbf{S}^{(m)(l)})||^2\right] \leq P, \tag{6.3}$$

where $\mathbb{E}$ denotes the expectation operator. We define

$$\Delta_{ij} = \mathbb{E}\left[\delta\left(\mathbf{S}_j^{(l)}, \hat{\mathbf{S}}_{i_j}^{(l)}\right)\right], \ i \in \{d, a\}, \ j = 1, 2, \ldots, m,$$

where $\hat{\mathbf{S}}_{i_j}^{(l)}$ denotes the reconstruction of the jth source component at the ith receiver. For a given power P, a distortion tuple $(\mathbf{D}_{d\,I_m}, \mathbf{D}_{a\,I_m})$ is said to be achievable if for any $\epsilon > 0$, there is a large enough $l(\epsilon)$ such that there exists a broadcast source-channel code of block-length $l(\epsilon)$ for which $\Delta_{ij} \leq D_{ij} + \epsilon, \ i \in \{d, a\}, \ j = 1, 2, \ldots, m$. A pair of distortions (D_d, D_a) is said to be achievable if there is an achievable $(\mathbf{D}_{d\,I_m}, \mathbf{D}_{a\,I_m})$ such that $D_i = \sum_{j=1}^{m} D_{ij}, \ i \in \{d, a\}$. In this chapter, we characterise the optimal trade-off between D_d and D_a, for a given power constraint P.

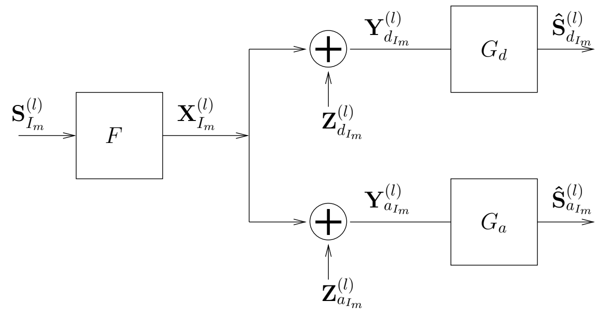


Figure 6.1: Problem set-up. Source $\mathbf{S}_{I_m}$ is communicated over a coloured Gaussian broadcast channel with noises $\mathbf{Z}_{d_{I_m}}$ and $\mathbf{Z}_{a_{I_m}}$ (digital and analog receiver respectively) using block-length l encoding/decoding functions.

6.2 An illustrative example: a 2-colour Gaussian source

Before we present the optimal trade-off, we present an insightful example of a parallel Gaussian source with $m = 2$ components transmitted over a non-degraded Gaussian broadcast channel (with noise variances (N_{i1}, N_{i2}) for $i \in \{d, a\}$ respectively), in order to highlight the key features underlying our achievable strategy. We target a single operating point where we are interested in determining the highest quality deliverable to the analog receiver under the constraint that we provide the point-to-point optimal quality to the digital receiver.

We first provide a quick review of the separation principle [14] based optimal point-to-point solution for the digital receiver. It is well-known that the optimal solution for the point-to-point source-channel problem can be obtained by the separation principle of first optimally source coding and then transmitting the resulting bit stream using an optimal channel code. The lowest distortion D_d attainable by the digital receiver is given by “reverse water-filling” over the source components [14, pg. 348]. We have,

$$D_d = \sum_{k=1}^{m} D_{dk} \tag{6.4}$$

$$\text{where } D_{dk} = \begin{cases} \mu, & \text{if } \mu < \sigma_k^2, \\ \sigma_k^2, & \text{if } \mu \geq \sigma_k^2, \end{cases} \tag{6.5}$$

and μ is chosen such that the total rate

$$\sum_{k=1}^{m} \frac{1}{2} \log(\sigma_k^2 / D_{dk}) = C_d. \tag{6.6}$$

C_d equals the capacity of the digital receiver channel. C_d is in turn given by “water-filling” over the sub-channels [14, pg. 250]

$$C_d = \sum_{k=1}^{m} \frac{1}{2} \log(1 + P_k / N_{dk}), \tag{6.7}$$

$$\text{where } P_k = \begin{cases} (\lambda - N_{dk}), & \text{if } \lambda > N_{dk}, \\ 0, & \text{if } \lambda \leq N_{dk}, \end{cases} \tag{6.8}$$

and we choose λ so that

$$\sum_{k=1}^{m} P_k = P. \tag{6.9}$$

If the point-to-point separation based digital coding approach is used for providing optimal fidelity to the digital receiver, the analog receiver might not be able to get much useful information. In this section we show that one can provide useful information to the analog receiver without compromising the digital receiver performance and achieve a distortion D_a at the analog receiver smaller than $\sigma_1^2 + \sigma_2^2$.We will later show that D_a is indeed the smallest analog distortion possible when the digital receiver achieves its optimal fidelity. We use the following sequence of steps to illustrate our proposed construction.

(a): Let the point-to-point optimal reverse water-filling solution for the digital receiver produce a total distortion D_d from a distortion allocation $D_{dk},\ k = 1, 2$ according to (6.5). We have

$$\frac{1}{2}\log\left(\frac{\sigma_1^2}{D_{d1}}\frac{\sigma_2^2}{D_{d2}}\right) = \frac{1}{2}\log\left(\frac{P_1 + N_{d1}}{N_{d1}}\frac{P_2 + N_{d2}}{N_{d2}}\right).$$

Without loss of generality, assume that $\frac{\sigma_1^2}{D_{d1}} \geq \left(\frac{P_1+N_{d1}}{N_{d1}}\right)$. In this situation, we can source code S_1 using successive refinement [16]. Successive refinement source coding produces a coarse description bit stream which can be decoded to produce an average distortion $D'_{d1} > D_{d1}$, and a refinement bit stream which can be used to refine the results of decoding the coarse description to a higher quality reproduction with an average distortion D_1. The result of Equitz and Cover [16] is that for memoryless Gaussian sources, this can be done without any loss in performance in the sense that the rate of the coarse description bit stream is $1/2\log(\sigma_1^2/D'_{d1})$ and the rate of the refinement bitstream is $1/2\log(\sigma_1^2/D_{d1}) - 1/2\log(\sigma_1^2/D'_{d1}) = 1/2\log(D'_{d1}/D_{d1})$. We will choose D'_{d1} such that the bit-rate of the coarse description is equal to the rate at which the first sub-channel operates so that the coarse description may be sent over the first sub-channel, i.e.,

$$\frac{1}{2}\log\left(\frac{\sigma_1^2}{D'_{d1}}\right) = \frac{1}{2}\log\left(1 + \frac{P_1}{N_{d1}}\right).$$

Then,

$$\frac{1}{2}\log\left(\frac{D'_{d1}}{D_{d1}}\frac{\sigma_2^2}{D_{d2}}\right) = \frac{1}{2}\log\left(1+\frac{P_2}{N_{d2}}\right).$$

Hence, the second sub-channel may carry refinement bits for S_1 as well as a description of S_2 at distortion D_{d2}. Further, these two bit streams may be transmitted using superposition channel coding [12] on the second sub-channel. The refinement bit stream for S_1 can be sent using power P'_2, and the rest of the power $P_2 - P'_2$ may be used to send the bit stream for S_2. At the receiver, the channel codeword corresponding to the transmission of the refinement bit stream will be treated as interference while channel decoding the bit stream for S_2. The decoded bit stream for S_2 can then be used to cancel the interference caused by its transmission while channel decoding the refinement bit stream for S_1. Thus P'_2 is chosen to satisfy

$$\frac{1}{2}\log\left(\frac{D'_{d1}}{D_{d1}}\right) = \frac{1}{2}\log\left(1+\frac{P'_2}{N_{d2}}\right),$$

which gives

$$\frac{1}{2}\log\left(\frac{\sigma_2^2}{D_{d2}}\right) = \frac{1}{2}\log\left(1+\frac{P_2-P'_2}{P'_2+N_{d2}}\right).$$

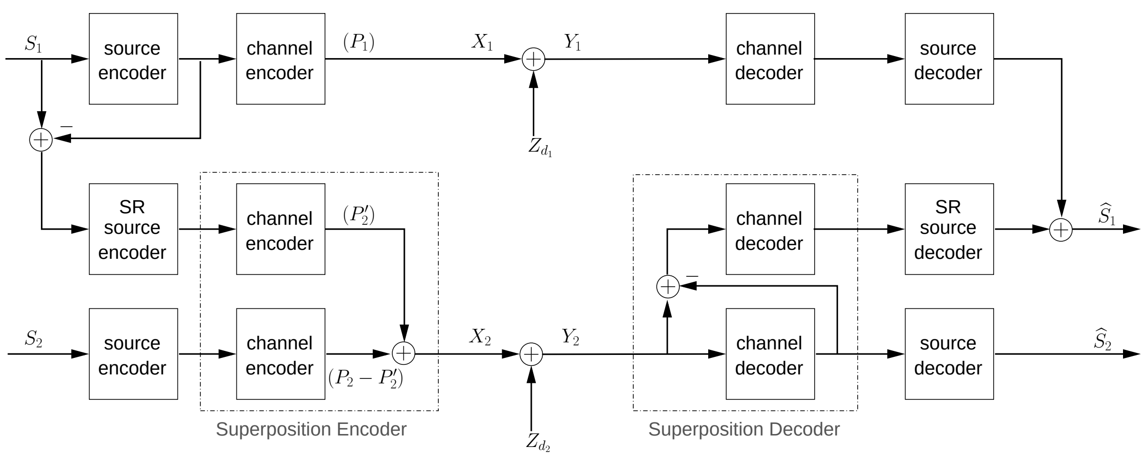
Figure 6.2 (a) illustrates this situation.

(b): Alternatively, we can treat the coarse description of S_1 from the first sub-channel as side information available at the decoder. In this case, we have the classical Wyner-Ziv set-up [57] where a source is encoded in the presence of side-information known only at the decoder. Thus, instead of sending the refinement bit stream, we can send a Wyner-Ziv bit-stream (of rate equal to that of the refinement bit stream) for S_1 on the second sub-channel. This can be done without losing optimality for the digital receiver [57]. This is illustrated in Figure 6.2 (b).

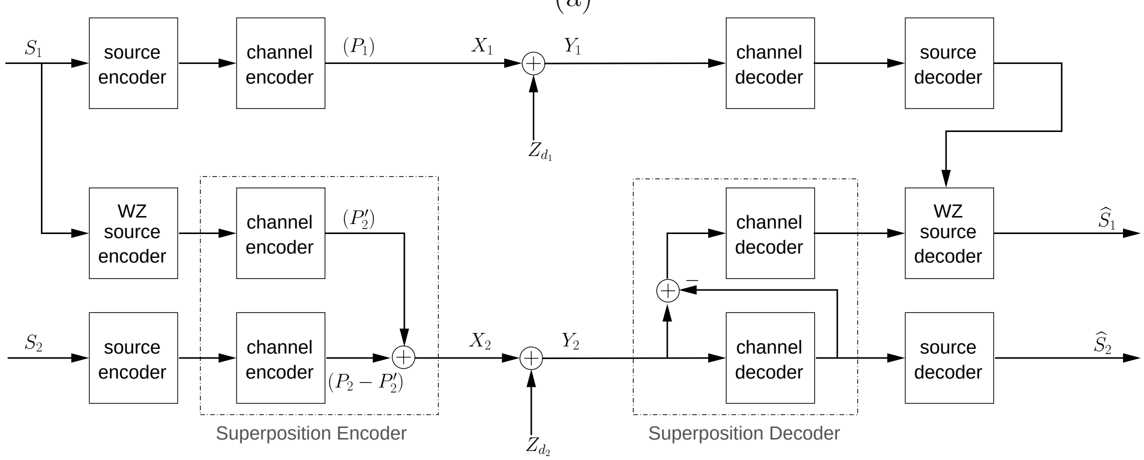
(c): Further, instead of sending the coarse description bit-stream for S_1 over the first sub-channel, we can also send S_1 uncoded (scaled by $\sqrt{P_1/\sigma_1^2}$) over the first sub-channel. The Wyner-Ziv bit-stream for S_1 on the second sub-channel does not depend on the exact realisation of the coarse description of S_1 and can use the corrupted version of S_1 from the first sub-channel as side information at the decoder. The optimality of this scheme follows from the no-rate-loss property of jointly Gaussian sources under Wyner-Ziv coding [57]. Thus the analog receiver can now form an estimate of S_1 from its output Y_1 of the first sub-channel. Figure 6.2 (c) depicts this setting.

(d): It is also possible to provide the analog receiver with an estimate of the second component without losing optimality for the digital receiver. To see this, note that we may use channel coding with side information as in Gel'fand-Pinsker (GP) [17] or dirty paper coding (DPC) [11] instead of superposition coding to achieve the same rates. Here the channel codeword for the bit stream of S_2 can be treated as non-causal side information available at the encoder when channel encoding the Wyner-Ziv bit stream. Then the digital receiver does not need to decode the bit stream for S_2 in order to decode the Wyner-Ziv bit stream for S_1. This is illustrated in Figure 6.2 (d).

(e): This further allows us to send S_2 uncoded (scaled by $\sqrt{(P_2 - P_2')/\sigma_2^2}$) as opposed to using the channel codeword for the bit stream of S_2 without impacting the



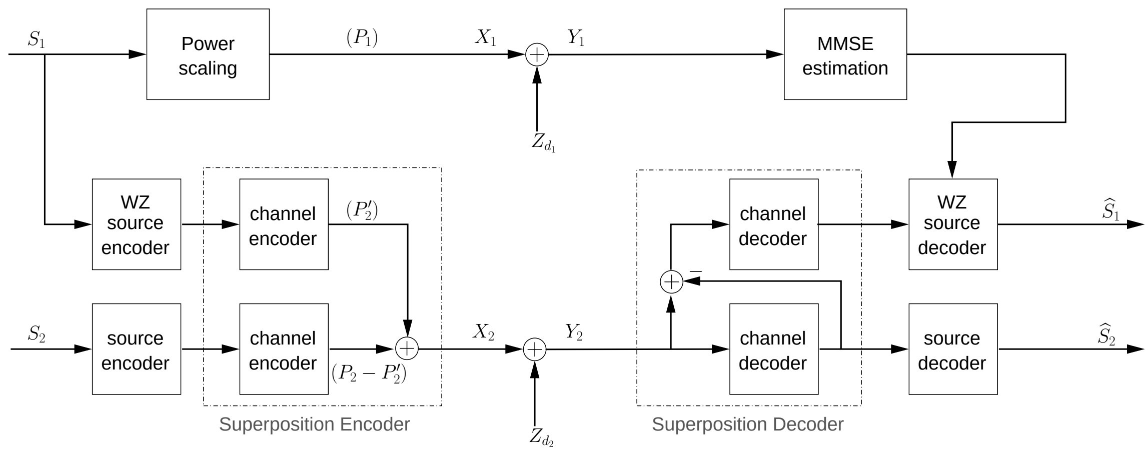
(a)



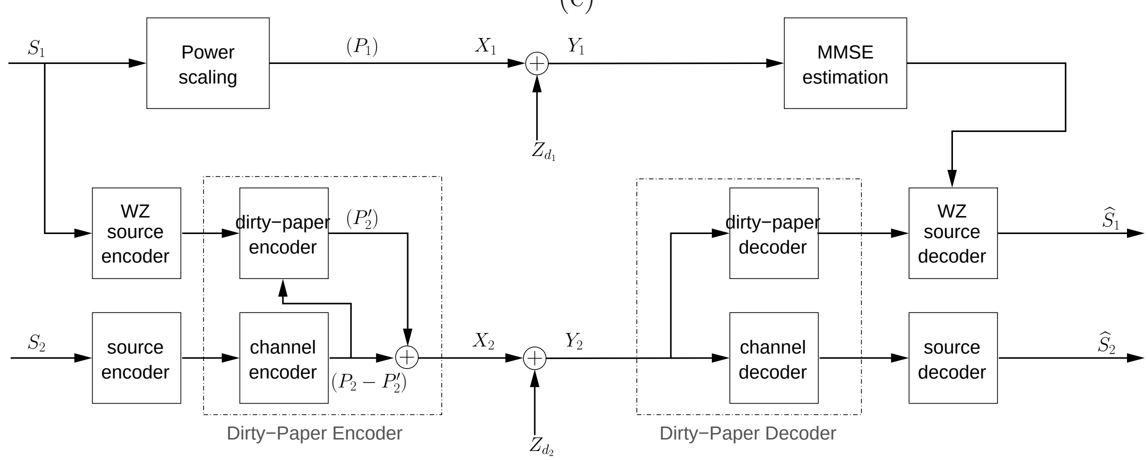
(b)

Figure 6.2: Illustration of hybrid analog-digital coding scheme for the case of digital receiver with point-to-point optimal quality: (a) Separation scheme showing successive refinement (SR) on source component S_1 and super-position coding on second sub-channel. (b) Separation scheme with a coarse description bit-stream on first sub-channel and Wyner-Ziv (W-Z) bit-stream on second sub-channel for S_1. (continued)

optimality for the digital receiver. The analog receiver can now form an estimate of S_2 from Y_2. This is shown in Figure 6.2 (e).



(c)



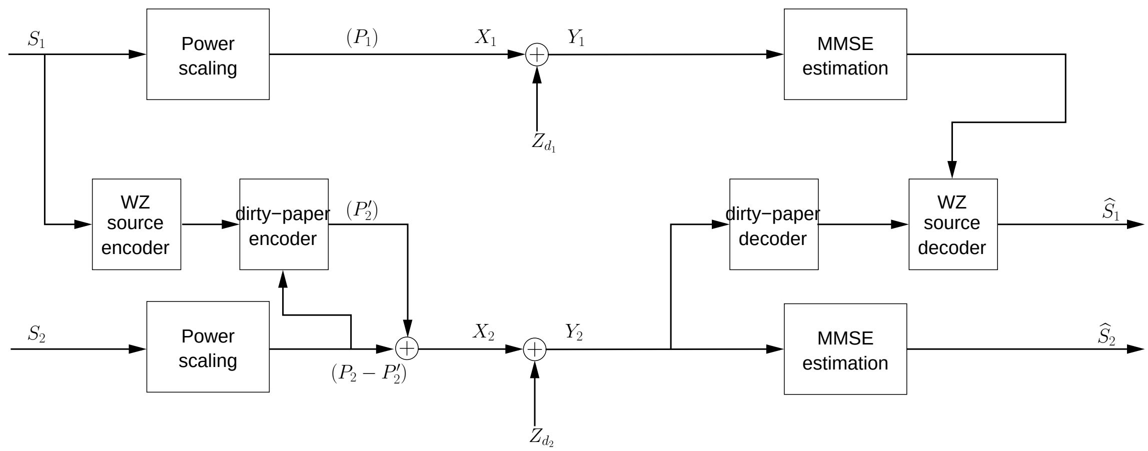
(d)

Figure 6.2: Illustration of hybrid analog-digital coding scheme for the case of digital receiver with point-to-point optimal quality (continued): (c) Hybrid scheme with an uncoded coarse description for S_1 on first sub-channel. (d) Hybrid scheme with dirty-paper coding (DPC) on second sub-channel for source bits for S_2 and W-Z bits for S_1. (continued)

To summarise, the total distortion observed at the analog receiver is given by

$$D_a = \frac{\sigma_1^2}{1 + \frac{P_1}{N_{a1}}} + \frac{\sigma_2^2}{1 + \frac{P_2 - P_2'}{P_2' + N_{a2}}}$$

which is strictly less than $(\sigma_1^2 + \sigma_2^2)$, the distortion for the analog receiver in the



(e)

Figure 6.2: Illustration of hybrid analog-digital coding scheme for the case of digital receiver with point-to-point optimal quality (continued): (e) Maximally analog hybrid scheme with uncoded description for S_2 on second sub-channel.

separation approach. Figure 6.2 summarises our proposed approach.

6.3 Main result and discussion

6.3.1 Main result

Theorem: Given an average system power P per symbol and a target analog receiver distortion D_a, the smallest distortion D_d achievable by the digital receiver is given by:

$$D_d = \min \sum_{j=1}^{m} D_{dj},$$

where the minimisation is over variables D_{dj}, D_{aj}, and P_j which are non-negative for $j \in I_m$, and are subject to the following additional constraints:

$$P \geq \sum_{j=1}^{m} P_j; \tag{6.10}$$

$$\sum_{j=1}^{m} \log \frac{\sigma_j^2}{D_{dj}} \leq \sum_{j=1}^{m} \log \left(1 + \frac{P_j}{N_{dj}}\right); \tag{6.11}$$

$$D_a \geq \sum_{j=1}^{m} D_{aj}; \tag{6.12}$$

$$\text{and } \forall j \in I_m,\ P_j \geq N_{aj} \left(\frac{\sigma_j^2}{D_{aj}} - 1\right). \tag{6.13}$$

Interpretation: We note that the optimisation problem for D_d subject only to the constraints (6.10) and (6.11) corresponds to the well-known point-to-point problem with only the digital receiver. The presence of the analog receiver is enforced by constraints (6.12) and (6.13). (6.13) imposes extra constraints on the allowable range of values that can be taken by P_j. In particular, a solution for D_d for which (6.13) are inactive, has the interpretation of offering the point-to-point optimal quality for the digital receiver while resulting in an analog distortion D_a that is better than the source variance.

It is also useful to note that the minimum feasible analog distortion D_a is given by

$$D_a = \min \sum_{j=1}^{m} D_{aj}, \tag{6.14}$$

where now the minimisation is only over non-negative variables D_{aj} and P_j for $j \in I_m$ subject to the constraints (6.10) and (6.13). For sufficiently large values of P, the optimal solution is given by

$$D_a = \frac{1}{P + \sum_{j=1}^{m} N_{aj}} \left(\sum_{j=1}^{m} \sqrt{\sigma_j^2 N_{aj}} \right)^2 \tag{6.15}$$

which is analogous to (58) in Berger and Tufts [6][1].

It is useful to note that the theorem also covers the case of bandwidth expansion if we allow some of the source components to have zero variances. We need to interpret $\frac{0}{0}$ as 1 when solving the optimisation problem, or equivalently, not impose the constraint (6.13) on those source components with zero variance and set the $D_{aj} = 0$ for those components.

---

[1] Here we explain how (5.1) was obtained. The analog distortion for large values of P is given above by (6.15). The digital distortion is obtained by the "inverse water-filling solution," where the total bit rate available is given by the capacity of the channel which can be obtained by the "water-filling solution" for the channel [14]. For large values of P, these solutions will allocate non-zero bitrates to all the source components and non-zero powers to all the sub-channels. Then, the overall distortion D_d is given by

$$\sum_j \log \frac{\sigma_j^2}{D_d/m} = \sum_j \log \frac{\mu}{N_{d_j}}$$

such that $\sum_j \mu - N_{d_j} = P$. That is,

$$D_d = \frac{m^2}{P + \sum_{j=1}^{m} N_{d_j}} \left(\prod_{j=1}^{m} \sigma_j^2 N_{d_j} \right)^{1/m}.$$

Thus, when $N_{d_j} = N_{a_j} = N_j$, say,

$$\frac{D_a}{D_d} = \left(\frac{\sum_{j=1}^{m} \sigma_j \sqrt{N_j}/m}{\left(\prod_{j=1}^{m} \sigma_j \sqrt{N_j} \right)^{1/m}} \right)^2,$$

which when specialised to the white channel where all N_j's are equal gives (5.1).

6.3.2 Comments on computing the distortion trade-off

The optimisation problem in the theorem can be easily verified to be a convex problem in D_{dj}, D_{aj}, P_j. When the channels have identical noise distributions, i.e., $N_{dj} = N_{aj} = N_j,\ j \in I_m$, the optimisation problem can be written as a geometric program [10] by re-parameterising the problem using the variables $p_j \overset{\text{def}}{=} P_j + N_j,\ j \in I_m$ in place of $P_j,\ j \in I_m$.

$$
\begin{aligned}
D_d = \min \ & \sum_{k=1}^{m} D_{d_k} \\
\text{subj.} \ & \left(\prod_{k=1}^{m} \sigma_k^2 N_k\right)\left(\prod_{k=1}^{m} p_k^{-1} D_{d_k}{}^{-1}\right) \le 1, \\
& \sigma_k^2 N_k p_k^{-1} D_{a_k}{}^{-1} \le 1,\ k = 1, 2, \ldots, m, \\
& \sum_{k=1}^{m} D_a^{-1} D_{a_k} \le 1, \\
& \sum_{k=1}^{m} \left(P + \sum_{j=1}^{m} N_j\right)^{-1} p_k \le 1, \\
& \sigma_k^{-2} D_{a_k} \le 1,\ k = 1, 2, \ldots, m, \\
& \sigma_k^{-2} D_{d_k} \le 1,\ k = 1, 2, \ldots, m, \\
\text{and} \ & N_k p_k^{-1} \le 1,\ k = 1, 2, \ldots, m.
\end{aligned}
$$

Fast optimisation routines (e.g. MOSEK [1]) can be used to solve this problem efficiently. As an illustration, the (D_d, D_a) distortion trade-off region for the source with component variances (0.8, 0.9, 0.2, 0.5, 0.3), sent over channels with identical noise variances (0.1, 0.2, 0.4, 0.6, 0.2), respectively, under a total power constraint $P = 1$ is shown in Fig. 6.3. The point A is obtained by using the separation approach for the digital receiver. At this point the analog receiver gets no useful information and its distortion is the sum of the variances of the source components. Without compromising the digital receiver's performance, we can improve the analog receiver's distortion to point B. The optimal performance for analog receiver is obtained at point C where both the digital and analog receivers have the same distortion.

6.4 A Wyner-Ziv Gel'fand-Pinsker system

We now present the main achievability result of this work along the lines of the hybrid strategy discussed in Section 6.2. Since the achievability proof involves relatively well-known arguments from [57] and [17], we only present a sketch of the main steps. In the rest of this section we will sketch a proof of the fact that given any choice of non-negative numbers $D_{dj}, D_{aj}, P_j,\ j \in I_m$ satisfying the conditions (6.10)-(6.13), the digital and analog receivers can simultaneously achieve distortions $\sum_{j=1}^{m} D_{dj}$ and $\sum_{j=1}^{m} D_{aj}$ respectively. This will prove the achievability of our main result.

As mentioned in Section 6.1.1, we are given a zero mean parallel Gaussian source $\mathbf{S}_{I_m}$ with m components. The covariance matrix of the source is
$K_{\mathbf{S}} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$. The coloured Gaussian broadcast channel consists of m sub-channels, with the digital receiver associated with a zero mean additive parallel Gaussian noise $\mathbf{Z}_{d_{I_m}}$ and the analog receiver with $\mathbf{Z}_{a_{I_m}}$. Further, we have, $K_{\mathbf{Z}_i} =$

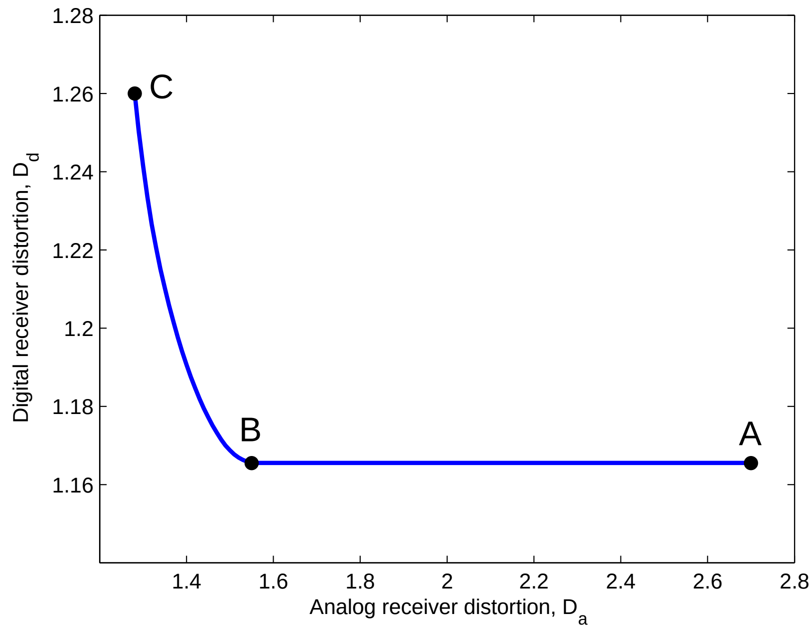


Figure 6.3: The optimal trade-off between analog receiver distortion D_a and the digital receiver distortion D_d for vector source with component variances (0.8, 0.9, 0.2, 0.5, 0.3), sent over channels with identical noise variances (0.1, 0.2, 0.4, 0.6, 0.2), respectively, under a total power constraint $P = 1$. The point A is obtained by using the separation approach for the digital receiver. At this point the analog receiver gets no useful information and its distortion is the sum of the variances of the source components. Without compromising the digital receiver's performance, we can improve the analog receiver's distortion to point B. The optimal performance for analog receiver is obtained at point C where both the digital and analog receivers have the same distortion at this point.

$\text{diag}(N_{i1}, N_{i2}, \ldots, N_{im})$ denote the noise covariance matrices for $i \in \{d, a\}$.

The block-length l output of the encoder to the j-th sub-channel, $\mathbf{X}_j^{(l)}$ is the sum of two parts: a digital (coded) part $\mathbf{X_c}_j^{(l)}$ which is useful only for the digital receiver and an analog (uncoded) part $\mathbf{X_u}_j^{(l)} = \sqrt{\alpha_j P_j / \sigma_j^2}\mathbf{S}_j^{(l)}$, where α_j is defined by

$$\alpha_j = \left(1 - \frac{D_{aj}}{\sigma_j^2}\right)\left(1 + \frac{N_{aj}}{P_j}\right). \tag{6.16}$$

It is easy to show that $0 \leq \alpha_j \leq 1$ for valid choices of D_{aj} and P_j. The average power of the coded part $\mathbf{X_c}_j^{(l)}$ will be chosen to be $(1-\alpha_j)P_j$ so that the average power of $\mathbf{X}_j^{(l)}$ is P_j. Then, by (6.10), the encoder output satisfies the average power constraint.

The analog receiver treats the coded part as noise and outputs the linear least-squared error (LLSE) estimate

$$\hat{\mathbf{S}}_{a_j}^{(l)} = \frac{\alpha_j P_j}{P_j + N_{aj}} \mathbf{Y}_{a_j}^{(l)}.$$

This defines the analog receiver's input-output map G_a. It is easy to show that the resulting distortion on the j-th source component at the analog receiver is D_{aj} by re-writing (6.16) as

$$1 + \frac{\alpha_j P_j}{(1-\alpha_j)P_j + N_{aj}} = \frac{\sigma_j^2}{D_{aj}}. \tag{6.17}$$

Thus the total distortion at the analog receiver is $\sum_{j=1}^{m} D_{aj}$. Similarly, the digital receiver forms an intermediate estimate $\hat{\mathbf{S}}_{dj}^{a(l)}$ of the j-th source component by treating the coded part as noise

$$\hat{\mathbf{S}}_{d_j}^{a(l)} = \frac{\alpha_j P_j}{P_j + N_{dj}} \mathbf{Y}_{d_j}^{(l)}$$

resulting in an intermediate distortion

$$D_{dj}^{a} = \frac{\sigma_j^2((1-\alpha_j)P_j + N_{dj})}{(P_j + N_{dj})}. \tag{6.18}$$

To generate the coded-part, the encoder treats the intermediate estimates $\hat{\mathbf{S}}_{d_{I_m}}^{a(l)}$ as side-information on $\mathbf{S}_{I_m}^{(l)}$ which will be available at the digital receiver. Also, it treats the uncoded-parts $\mathbf{X_u}_{I_m}^{(l)}$ as additive channel-side information ("dirt" in the sense of [11]) which affects its transmission. This channel-side information is available at the encoder. This sets up a point-to-point source-channel coding problem from the encoder to the digital receiver with source side-information only at the decoder and channel side-information only at the encoder. The coded-part is realised by a separation-based strategy consisting of a Wyner-Ziv source code followed by a Gel'fand-Pinsker/Costa channel code for this source-channel coding problem. We note that this separation strategy was proved to be optimal in a point-to-point setting in [31]. However, we do not make use of this fact here.

The Gel'fand-Pinsker/Costa channel is a parallel additive Gaussian noise channel with noises $\mathbf{Z}_{d_{I_m}}$ and additive side-information ("dirt") $\sqrt{\alpha_j P_j/\sigma_j^2}\mathbf{S}_j^{(l)}$, $j \in I_m$ which is available at the encoder. There are average power constraints on each sub-channel: $(1-\alpha_j)P_j$ for the j-th sub-channel. The j-th sub-channel can support a rate of $C_j = \frac{1}{2}\log\left(1 + \frac{(1-\alpha_j)P_j}{N_{dj}}\right)$, the same as the capacity of an AWGN with noise variance N_{dj}. In other words, for sufficiently large l, there is an encoder mapping $F_{cj} : \{0,1\}^{lC_j} \times \mathbb{R}^l \to \mathbb{R}^l$ satisfying the power constraint such that there is a decoder map $G_{cj} : \mathbb{R}^l \to \{0,1\}^{lC_j}$ and the probability that $G_{cj}\left(F_{cj}\left(m, \sqrt{\alpha_j P_j/\sigma_j^2}\mathbf{S}_j^{(l)}\right) + \mathbf{Z}_{d_j} + \sqrt{\alpha_j P_j/\sigma_j^2}\mathbf{S}_j^{(l)}\right)$ $\neq m$ is less than ϵ, for any $m \in \{0,1\}^{lC_j}$, for a given $\epsilon > 0$. This fact was shown by Costa in [11]. The reader is referred to section II of [11] for a proof[2]. Hence the total

[2]Note, however, that unlike in Costa's setting, the input to the Costa channel encoder here is dependent on the channel-side information. This can be easily handled by a dithering argument. The encoder adds (modulo-2) a random binary dither sequence to the output bits from the Wyner-Ziv encoder before sending it to the Costa encoder and the digital decoder removes the dither by adding (modulo-2) the same sequence after performing the Costa decoding. If the dither process is independent of the source and is i.i.d. with 1 and 0 occurring with equal probabilities, the input to the Costa encoder is independent of the channel-side information. This also ensures that the output

rate supported by the channel code is

$$C = \sum_{j=1}^{m} C_j = \sum_{j=1}^{m} \frac{1}{2} \log \left(1 + \frac{(1-\alpha_j)P_j}{N_{dj}} \right). \tag{6.19}$$

The Wyner-Ziv source coding problem has a parallel Gaussian source $\mathbf{S}_{I_m}^{(l)}$ with side-information at the decoder $\mathbf{S}_{d_{I_m}}^{a(l)}$ which are within MSE distortions $D_{d1}^a, D_{d2}^a, \ldots,$ D_{dm}^a respectively. The rate at which the j-th source component may be encoded to achieve a distortion of D_{dj} is given by $R_j = \frac{1}{2} \log \frac{D_{dj}^a}{D_{dj}}$. More formally, for sufficiently large l, there is an encoder mapping $F_{sj} : \mathbb{R}^l \to \{0,1\}^{lR_j}$ and a decoder mapping $G_{sj} : \{0,1\}^{lR_j} \times \mathbb{R}^l \to \mathbb{R}^l$ such that

$$\mathbb{E}\left[\delta \left(\mathbf{S}_j^{(l)}, G_{sj} \left(F_{sj} \left(\mathbf{S}_j^{(l)} \right), \mathbf{S}_{dj}^{a(l)} \right) \right) \right] \le D_{dj} + \epsilon,$$

for a given $\epsilon > 0$. This fact can be directly inferred from an extension to the continuous alphabet case of the proof of Wyner and Ziv [57, Sec. IV]. A rigorous way of doing this extension can be found in [33] (also see [22, 56])[3]. Hence, the total rate of the source code is

$$R = \sum_{j=1}^{m} R_j = \sum_{j=1}^{m} \frac{1}{2} \log \left(\frac{D_{dj}^a}{D_{dj}} \right). \tag{6.20}$$

The separation approach taken for the coded part is feasible if the Wyner-Ziv source code rate is less than or equal to the Gel'fand-Pinsker/Costa channel code

---

of the Costa encoder is independent of the uncoded part of the transmission and thus the power constraint is satisfied. We can even avoid the need for the encoder and the digital decoder to share any common randomness by a joint-design which is not pursued here.

[3]The results in the literature assume that the side information at the decoder is corrupted by Gaussian noise whereas in our problem we assumed that the side information is within a given MSE distortion from the source. However, this difference is not difficult to handle. See [38, Appendix IV] for a proof.

rate, i.e., $R \leq C$. It is easy to see that the condition (6.13) in fact implies $R = C$ from (6.18), (6.19), and (6.20). Thus the digital receiver will be able to reconstruct the j-th source components within a distortion D_{dj} and thereby achieve a distortion of $\sum_{j=1}^{m} D_{dj}$.

6.5 Converse

Having presented the achievable strategy in Section 6.4, we now prove the optimality of the proposed scheme. The structural constraint on the analog receiver implies that in estimating any source component, it makes use of information received only over the corresponding sub-channel. To formalise this, for $j = 1, 2, \ldots, m$, let $D_{aj} \geq 0$ be the distortion achieved by the analog receiver on the j-th component. We have,

$$
\begin{aligned}
\frac{l}{2} \log \frac{\sigma_j^2}{D_{aj}} &\leq I(\mathbf{S}_j^{(l)}; \hat{\mathbf{S}}_{aj}^{(l)}) \\
&\leq I(\mathbf{Y}_{aj}^{(l)}; \mathbf{X}_{I_m}^{(l)}) \\
&= h(\mathbf{Y}_{aj}^{(l)}) - h(\mathbf{Z}_{aj}^{(l)}) \\
&\leq \frac{l}{2} \log \frac{P_j + N_{aj}}{N_{aj}},
\end{aligned}
$$

where $P_j \stackrel{\text{def}}{=} (\text{Var}(\mathbf{X}_j^{(l)}))/l$. The first inequality is Shannon's lower bound [14] for Gaussian sources and squared-error distortion. The second inequality is the data processing inequality, and the last follows from the fact that Gaussian distribution maximises the differential entropy for a given variance.

Again we have the following Shannon's lower bound

$$
\begin{aligned}
I(\mathbf{S}_{I_m}^{(l)};\hat{\mathbf{S}}_{d_{I_m}}^{(l)}) &= h(\mathbf{S}_{I_m}^{(l)}) - h(\mathbf{S}_{I_m}^{(l)}|\hat{\mathbf{S}}_{d_{I_m}}^{(l)}) \\
&\geq h(\mathbf{S}_{I_m}^{(l)}) - h(\mathbf{S}_{I_m}^{(l)} - \hat{\mathbf{S}}_{d_{I_m}}^{(l)}) \\
&\geq \sum_{j=1}^{m} \frac{l}{2}\log 2\pi e\sigma_j^2 - \sum_{j=1}^{m}\frac{l}{2}\log 2\pi e D_{dj} \\
&= \sum_{j=1}^{m}\frac{l}{2}\log\frac{\sigma_j^2}{D_{dj}}.
\end{aligned}
$$

The second inequality follows from the fact that given the diagonal elements of the covariance matrix of a random vector, independent Gaussian random variables maximise the differential entropy. Further, using the data processing inequality,

$$
\begin{aligned}
\sum_{j=1}^{m}\frac{l}{2}\log\frac{\sigma_j^2}{D_{dj}} &\leq I(\mathbf{S}_{I_m}^{(l)};\hat{\mathbf{S}}_{d_{I_m}}^{(l)}) \\
&\leq I(\mathbf{X}_{I_m}^{(l)};\mathbf{Y}_{d_{I_m}}^{(l)}) \\
&\leq \sum_{j=1}^{m}\frac{l}{2}\log\frac{P_j + N_{dj}}{N_{dj}}.
\end{aligned}
$$

Thus, for a given power P and target analog distortion profile $(D_{a_1}, D_{a_2}, \ldots, D_{a_m})$, the smallest sum distortion D_d achievable by the digital receiver is lower bounded by

$$\begin{aligned} \min \quad & \sum_{j=1}^{m} D_{dj} \\ \text{s.t.} \sum_{j=1}^{m} \log \frac{\sigma_j^2}{D_{dj}} &\leq \sum_{j=1}^{m} \log \frac{P_j + N_{dj}}{N_{dj}}, \\ \frac{\sigma_j^2}{D_{aj}} &\leq \frac{P_j + N_{aj}}{N_{aj}}, \ \forall j \in I_m; \\ \sum_{j=1}^{m} P_j &\leq P, \\ \text{and } D_{dj}, \ P_j &\geq 0, \ \forall j \in I_m. \end{aligned}$$

Now, for a given target analog distortion D_a, we can further minimise the above over all $(D_{a_1}, D_{a_2}, \ldots, D_{a_m})$ s.t. $\forall j \in I_m, \ D_{aj} \geq 0$ and $D_a \geq \sum_{j=1}^{m} D_{aj}$, to get the lower bound on the digital distortion D_d which proves the converse statement in the main theorem.

Chapter 7

Source-Channel Broadcast

7.1 Problem statement

The setup is similar to the one in the previous chapter, see Fig. 6.1. The key difference here is that neither of the decoders is constrained to be linear (or analog). For the most part we will consider only a scalar broadcast channel in the sense that all the sub-channels to a particular user will have the same statistics. In the notation of the last chapter, this means $N_{i_1} = N_{i_2} = \ldots$, for each i. The upshot of this assumption is a simplification which results from recognising that there is an ordering of the users according to the noise variance of their channel. We will call the user with the smaller noise variance the *strong user*, and the user with the larger noise variance the *weak user*[1]. Besides the ease of exposition resulting from this assumption[2], it can be justified by recognising that in practice for many broadcast applications, the SNR gap between groups of users to whom different qualities of service are targeted will be large enough that there is an ordering between the groups of users. And at large SNRs the coloured channel (or inter-symbol interference channel) to the users can be approximated by a white channel.

We will explicitly model the fact that the source and channel bandwidths do not necessarily match. We let our source have K components: $S_k(i)$ is a vector Gaussian source which is independent over its components indexed by $k = 1, 2, \ldots, K$ with statistics $S_k(i) \sim \mathcal{N}(0, \sigma_k^2)$ and i.i.d. over time index $i = 1, 2, \ldots$. Without loss of generality, we will assume that $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_K^2$. We will assume that there are M parallel broadcast sub-channels of the same statistics. When $M = K$, the bandwidth of the source matches the bandwidth of the channel. The weak user observes $Y_{wm}(i) = X_m(i) + Z_{wm}(i),\ m = 1, 2, \ldots, M,\ i = 1, 2, \ldots$, and the strong user observes $Y_{sm}(i) = X_m(i) + Z_{sm}(i)$, where $X_m(i)$ is the input to the m-th sub-channel. The noise processes are $Z_{wm}(i)$ and $Z_{sm}(i)$ i.i.d. (over m and i) Gaussian

[1] It is known that the performance of such a broadcast channel is identical to that of a *degraded* broadcast channel where the weak user receives the signal received by the strong user but further corrupted by an additive Gaussian noise independent of the additive noise corrupting the strong user's channel and which has a variance equal to the difference between the variances of the additive noises of the weak user and the strong user in the original channel. This fact is often expressed by saying that the original broadcast channel is *stochastically degraded*. A proof of the above stated equivalence for the case of channel coding appears in [14, pg. 422]. The same idea can be used to show an equivalence for the problem of interest here, see e.g. [38, Appendix 1]. However, we do not need to make use of this equivalence in our discussion here.

[2] Note that the ideas presented in this chapter also apply to the case of parallel vector broadcast channels, but we do not explore them here.

with variances N_w and N_s respectively, where $N_s < N_w$. The source-channel encoder $f^n : \mathbb{R}^{Kn} \to \mathbb{R}^{Mn}$ maps an n-length block of the source to an n-length block of the channel input. There is an average power constraint on the encoder so that $\left(\sum_{i=1}^{n}\sum_{m=1}^{M} X_m^2(i)\right)/(nM) \leq P$. The source-channel decoders $g_s^n : \mathbb{R}^{Mn} \to \mathbb{R}^{Kn}$ and $g_w^n : \mathbb{R}^{Mn} \to \mathbb{R}^{Kn}$ at the strong and weak user, respectively, reconstruct n-length blocks,

$$
\begin{aligned}
&\left\{\widehat{S}_{sk}(i),\ i = 1, \ldots, n, k = 1, \ldots, K\right\} \\
&\quad = g_s^n\left(\{Y_{sm}(i),\ i = 1, 2, \ldots, n,\ m = 1, 2, \ldots, M\}\right), \text{and} \\
&\left\{\widehat{S_{wk}}(i),\ i = 1, \ldots, n, k = 1, \ldots, K\right\} \\
&\quad = g_w^n\left(\{Y_{wm}(i),\ i = 1, 2, \ldots, n,\ m = 1, 2, \ldots, M\}\right)
\end{aligned}
$$

of the source from n-length blocks of the channel outputs. Distortions are measured as the average of the mean-squared error distortion over all source components[3]

$$
D_j^n = \frac{1}{nK}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(S_k(i) - \widehat{S}_{jk}(i)\right)^2,\ j \in \{s, w\}.
$$

A pair of distortions (D_s, D_w) will be said to be *achievable* if for any $\epsilon > 0$, for sufficiently large n, there is (f^n, g_s^n, g_w^n) such that $D_j^n \leq D_j + \epsilon,\ \ j \in \{s, w\}$. The problem is to characterise the region of all the achievable distortions (D_s, D_w) for a given transmit power P. This remains open. We present below an inner bound to it (or an achievable region).

[3]Note that the power constraint and the distortions are defined slightly differently from the previous chapter. Here we average over the number of sub-channels and components respectively. This is to allow easier comparison with previous works.

7.2 An achievable solution: the extreme points

In this section we present all the key ideas involved in our achievable strategy. We consider two extreme cases – when the weak user achieves its point-to-point optimal quality, and when the strong user achieves its point-to-point optimal quality – to illustrate two complimentary strategies which together constitute the general achievable solution.

7.2.1 Weak-user-optimal case

From Shannon's separation theorem [14, pg. 216] we know that the optimal solution for the point-to-point source-channel problem can be obtained by the separation principle of first optimally source coding and then transmitting the resulting bit stream using an optimal channel code. Thus the lowest distortion D_w^* attainable by the weak user is given by "reverse water-filling" over the source components [14, pg. 348]

$$D_w^* = \frac{1}{K}\sum_{k=1}^{K} D_k,\ \text{where } D_k = \begin{cases} \mu, & \text{if } \mu < \sigma_k^2, \\ \sigma_k^2, & \text{if } \mu \geq \sigma_k^2, \end{cases} \tag{7.1}$$

where μ is chosen such that the total rate $(1/2)\sum_{k=1}^{K}\log(\sigma_k^2/D_k)$ equals the capacity C_w of the weak user's channel. C_w is in turn given by $\sum_{m=1}^{M}\log(1+P_m/N_w)$ where we choose $P_1 = P_2 = \ldots = P_M = P$.

If this separation strategy is followed for the broadcast case as well, the strong user also recovers the source at a distortion D_w^*. However, without compromising the quality of reproduction for the weak user, better quality can be delivered to the strong user. Before presenting our solution in full generality, it is useful to consider the special case of $K = M = 2$ and $\sigma_1^2 > \sigma_2^2$. Let us suppose that the optimal point-to-point reverse water-filling solution for the weak user allocates distortions D_1 and D_2 for the source components S_1 and S_2, respectively. Also, let us denote the powers allocated to the sub-channels X_1 and X_2 by P_1 and P_2 respectively. $P_1 = P_2 = P$. Then

$$\frac{1}{2}\log\left(\frac{\sigma_1^2}{D_1}\frac{\sigma_2^2}{D_2}\right) = \frac{1}{2}\log\left(\frac{P_1 + N_w}{N_w}\frac{P_2 + N_w}{N_w}\right).$$

We can source code S_1 using a successive refinement strategy thereby producing two bit streams: a coarse description at distortion D_1' and a refinement stream which refines from D_1' to D_1. Since Gaussian sources are successively refinable [16], this can be done without loss of optimality for the weak user. We choose D_1' such that the bitrate of the refinement stream is equal to the rate at which the first sub-channel operates. i.e.,

$$\frac{1}{2}\log\left(\frac{D_1'}{D_1}\right) = \frac{1}{2}\log\left(1 + \frac{P_1}{N_w}\right).$$

Combining the two equations above gives

$$\frac{1}{2}\log\left(\frac{\sigma_1^2}{D_1'}\frac{\sigma_2^2}{D_2}\right) = \frac{1}{2}\log\left(1 + \frac{P_2}{N_w}\right).$$

In other words, without loss of optimality, we may send the coarse description for S_1 and the bit stream for S_2 over the second sub-channel and the refinement bitstream for S_1 over the the first sub-channel. This further suggests that instead of sending the refinement bitstream over the first sub-channel, we may send uncoded the quantisation error resulting from the coarse quantisation of S_1 appropriately scaled to satisfy the power constraint. The input to the first sub-channel will be $\sqrt{(P_1/D_1')}\left(S_1(i) - \widehat{S}_1'(i)\right)$, where $\widehat{S}_1'(i)$ is the i-th sample of the coarsely quantised version of S_1. It is easy to see that this satisfies the power constraint on the first sub-channel and also results in no loss of optimality for the weak user. The second fact is analogous to the optimality of uncoded transmission for the point-to-point Gaussian source-channel problem. The strong user can achieve a lower distortion on S_1 because of the uncoded transmission of the quantisation error. The strong user estimates the refinement component as $(P_1/(P_1 + N_s))\sqrt{D_1'/P_1}Y_1$ and adds it to the coarse description to form its reproduction of S_1. We note that the resulting distortion for the strong user on S_1 is $D_1'/(1 + P_1/N_s)$.

The performance of the strong user can be further improved. Without losing optimality for the weak user, we may send the coarse description of S_1 and the bit stream for S_2 using superposition coding over the second sub-channel. In particular, we send the coarse quantisation bit stream of S_1 using power $P_2 - P_2'$ (defined below) and the bit stream for S_2 using power P_2' such that the decoder can first decode the former bit stream assuming the latter as interference. The decoder then cancels the interference from the bit stream for S_1 and decodes the bit stream for S_2. Thus P_2' is given by

$$\frac{1}{2}\log\left(\frac{\sigma_1^2}{D_1'}\right) = \frac{1}{2}\log\left(1 + \frac{P_2 - P_2'}{P_2' + N_w}\right).$$

This also gives the relation

$$\frac{1}{2}\log\left(\frac{\sigma_2^2}{D_2}\right) = \frac{1}{2}\log\left(1 + \frac{P_2'}{N_w}\right)$$

which indicates why decoding of the bit stream for S_2 after interference cancellation succeeds. This scheme suggests that we may send S_2 uncoded using power P_2' instead of sending its quantised bits. Since, after cancelling the interference from the S_1 bit stream, the channel to the weak user is an AWGN channel, the optimality of this scheme follows from the optimality of uncoded transmission for point-to-point Gaussian source-channel coding. The strong user can now reconstruct S_2 at a lower distortion, $D_2(1 + P_2'/N_w)/(1 + P_2'/N_s)$. The scheme is summarised in Fig. 7.1. The overall distortion achieved by the strong user is

$$D_s = \frac{1}{2}\left(\frac{1 + \frac{P_1}{N_w}}{1 + \frac{P_1}{N_s}}D_1 + \frac{1 + \frac{P_2'}{N_w}}{1 + \frac{P_2'}{N_s}}D_2\right).$$

The extension to $K = M > 2$ is straightforward. Let us assume without loss of generality that under the point-to-point optimal inverse water-filling solution for the weak user, the first L source components satisfy $(1/2)\log(\sigma_k^2/D_k) \geq (1/2)\log(1 + P_k/N_w)$. For these L components we define D_k' such that $(1/2)\log(D_k'/D_k) = (1/2)\log(1 + P_k/N_w)$. The k-th such component ($k \leq L$) is source coded to a distortion of D_k' and the resulting error is sent uncoded (scaled by $\sqrt{P_k/D_k}$) over the k-th sub-channel. For sub-channels $m > L$, we define P_m' as $(1/2)\log(\sigma_m^2/D_m) = (1/2)\log(1+$

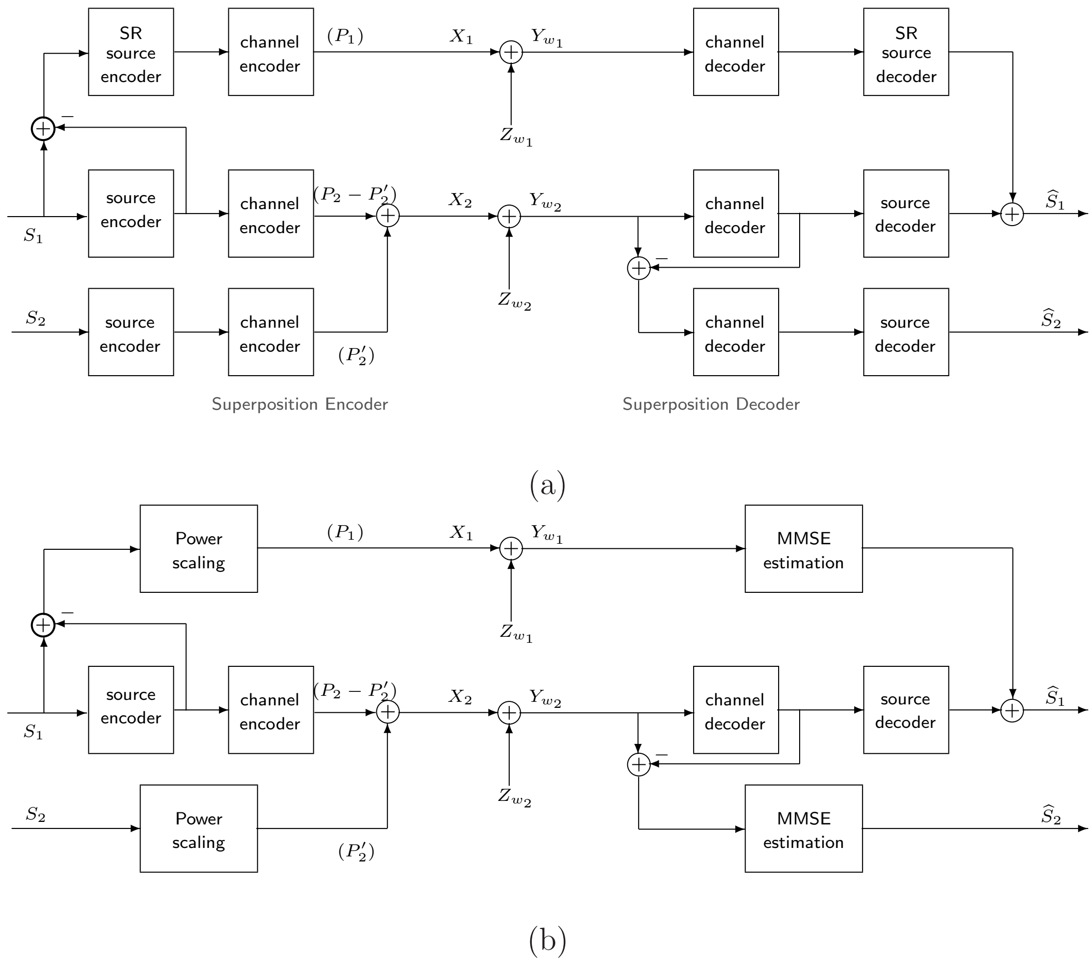


Figure 7.1: Weak-user-optimal case: (a) separation scheme showing successive refinement (SR) and superposition coding, and (b) the hybrid digital-analog scheme.

P'_m/N_w). The m-th such component is sent uncoded over the m-th sub-channel scaled by $\sqrt{P'_m/\sigma^2_m}$. The rest of the power $(P_m - P'_m)$ for these sub-channels $m > L$ are used to send the source coded bits from the first L components. On these sub-channels, the decoders first decode these bits, cancel the interference caused by them, and then estimate the source components. The first L source components are estimated directly from the corresponding sub-channel outputs. Thus, without compromising the quality of reproduction for the weak user, the strong user achieves a lower distortion.

$$D_s = \frac{1}{K}\left(\sum_{k=1}^{L}\frac{1+\frac{P_k}{N_w}}{1+\frac{P_k}{N_s}}D_k + \sum_{k=L+1}^{K}\frac{1+\frac{P_k'}{N_w}}{1+\frac{P_k'}{N_s}}D_k\right) < \frac{1}{K}\sum_{k=1}^{K}D_k.$$

This scheme directly extends to the $K \neq M$ case. When $K < M$ (bandwidth expansion), the extra sub-channels will be used to send additional coded bits. For the case of $K > M$ (bandwidth contraction), only at most M source components can be sent (wholly or partially) uncoded. In summary, we have the following

Theorem 7.2.1 *For the source-channel problem in Section 7.1* (D_s, D_w^*) *is achievable, where* D_w^* *is given by (7.1) and* D_s *is as defined below.*

$$D_s = \frac{1}{K}\left(\sum_{k=1}^{L}\frac{1+\frac{P_k}{N_w}}{1+\frac{P_k}{N_s}}D_k + \sum_{k=L+1}^{K'}\frac{1+\frac{P_k'}{N_w}}{1+\frac{P_k'}{N_s}}D_k + \sum_{k=K'+1}^{K}D_k\right),$$

where D_k*'s are given by (7.1),* $P_k = P$*,*

$$L = \min\left\{\left|\left\{k : \frac{\sigma_k^2}{D_k} \geq 1 + \frac{P_k}{N_w}\right\}\right|, M\right\},$$

$$K' = \min\left\{\left|\left\{k : \mu \leq \sigma_k^2\right\}\right|, K\right\},$$

and the P_k'*'s are defined by*

$$\frac{\sigma_k^2}{D_k} = 1 + \frac{P_k'}{N_w},\ k = L+1, \ldots, K'.$$

7.2.2 Strong-user-optimal case

If the point-to-point separation approach is used for providing optimal fidelity to the strong user, since the rate of transmission is greater than the channel capacity of the weak user, the weak user will not be able to get any useful information. However, in this section we show that one can provide useful information to the weak user without compromising the strong user's performance. The scheme is the same as the one described in section 6.2. We repeat the arguments in brief here. Let the point-to-point optimal reverse water-filling solution for the strong user produce a total distortion D_s^* from a distortion allocation $D_k,\ k = 1, 2, \dots, K$ according to (7.1), where μ is now chosen so that the total rate equals the capacity $C_s = \sum_{m=1}^{M} \log(1 + P_m/N_s),\ P_m = P$ of the strong user's channel. It is again helpful to consider the special case of $K = M = 2$ and $\sigma_1^2 > \sigma_2^2$ before the general case. With $P_1 = P_2 = P$, we have

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D_1} \frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(\frac{P_1 + N_s}{N_s} \frac{P_2 + N_s}{N_s} \right).$$

We source code S_1 using successive refinement such that now the bitrate of the coarse description at distortion D_1'' is equal to the rate at which the first sub-channel operates.

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D_1''} \right) = \frac{1}{2} \log \left(1 + \frac{P_1}{N_s} \right).$$

Note that this is different from the previous section where we set the bitrate of the refinement bit stream equal to the rate of the first sub-channel. Thus

$$\frac{1}{2} \log \left(\frac{D_1''}{D_1} \frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(1 + \frac{P_2}{N_s} \right).$$

Instead of sending the coarse description over the first sub-channel and the refinement bit stream on the second, without losing optimality for the strong user, we may send S_1 uncoded (scaled by $\sqrt{P_1/\sigma_1^2}$) over the first sub-channel and on the second sub-channel a Wyner-Ziv bit stream (of rate equal to that of the refinement bit stream) which assumes that the corrupted version of S_1 from the first sub-channel will be available as side information at the decoder. The optimality of this scheme follows from the no rate-loss property of jointly Gaussian sources under Wyner-Ziv coding.

Thus the weak user can now form an estimate of S_1 from its output Y_1 of the first sub-channel. It is also possible to provide the weak user with an estimate of the second component without losing optimality for the strong user. Let us define P_2'' such that the Wyner-Ziv bit stream can be sent using this power, superposition coded with the bit stream for S_2 which uses the rest of the power $P_2 - P_2''$. The decoding order is first the bit stream for S_2, followed by the Wyner-Ziv bit stream. Thus

$$\frac{1}{2}\log\left(\frac{D_1''}{D_1}\right) = \frac{1}{2}\log\left(1 + \frac{P_2''}{N_s}\right),$$

and

$$\frac{1}{2}\log\left(\frac{\sigma_2^2}{D_2}\right) = \frac{1}{2}\log\left(1 + \frac{P_2 - P_2''}{P_2'' + N_s}\right).$$

We can use dirty-paper coding of Gel'fand and Pinsker [17] and Costa [11] instead of superposition coding to achieve the same rates. Here the channel codeword for the bit stream of S_2 is treated as non-causal side information available at the encoder when channel encoding the Wyner-Ziv bit stream. Then the decoder does not need to decode the bit stream for S_2 in order to decode the Wyner-Ziv bit stream. This allows

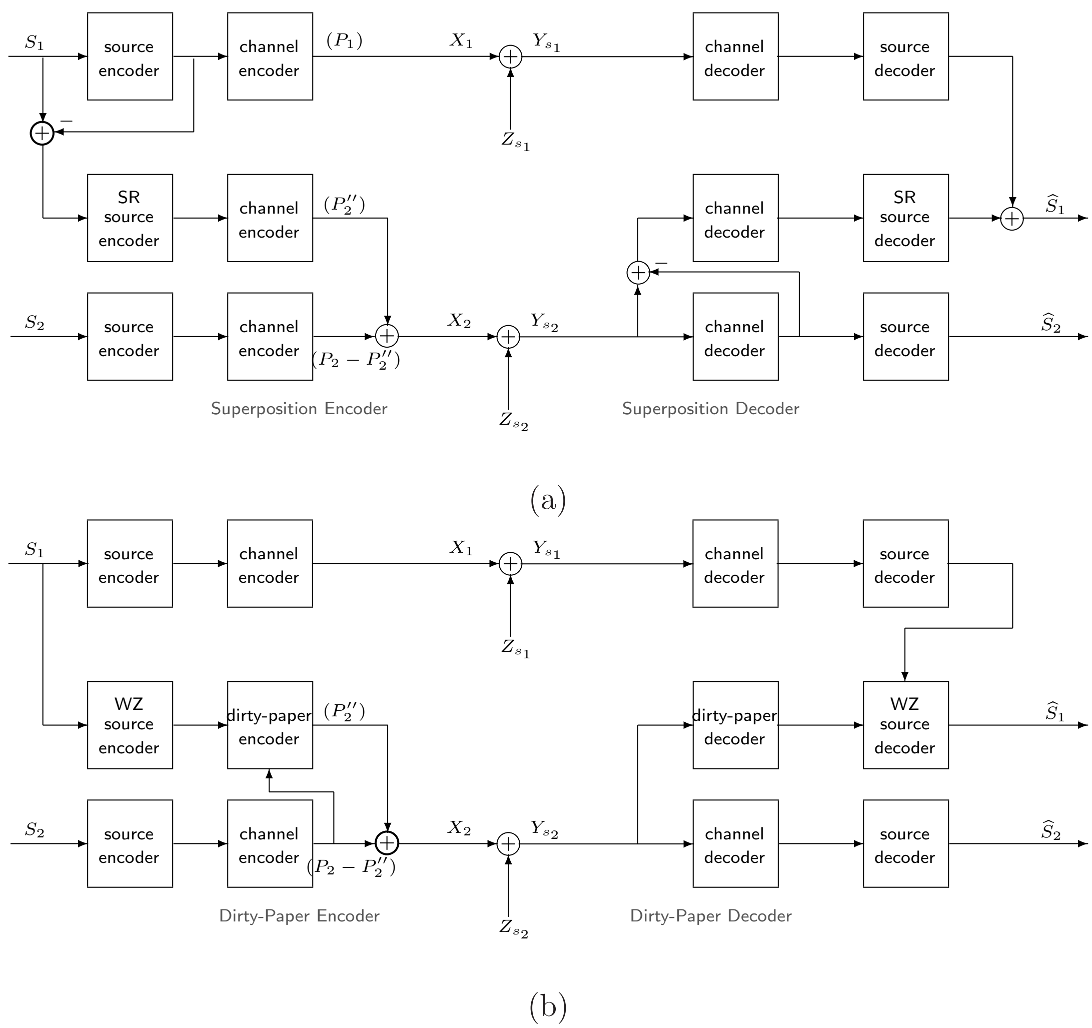
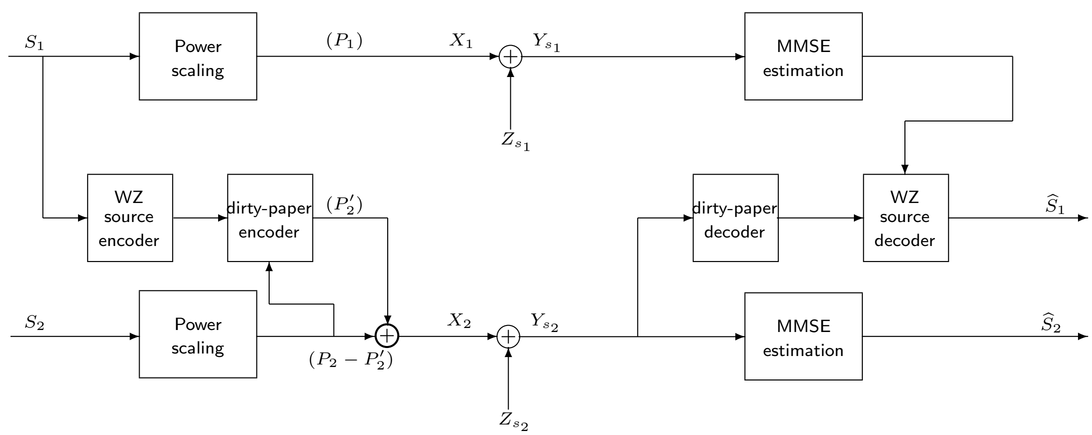


Figure 7.2: Strong-user-optimal case: (a) separation scheme showing successive refinement (SR) and superposition coding, (b) separation scheme with Wyner-Ziv (W-Z) code and dirty-paper coding (DPC). (continued)

us to send S_2 uncoded (scaled by $\sqrt{(P_2 - P_2'')/\sigma_2^2}$) without impacting the optimality for the strong user. The weak user can now form an estimate of S_2 from Y_2.

Again, we can easily extend the above intuition to $K = M > 2$. Let L be the number of source components and sub-channels such that $(1/2)\log(\sigma_k^2/D_k) > (1/2)\log(1 + P_k/N_s)$. Since σ_k^2 are monotonically decreasing, these will be the first



(c)

Figure 7.2: Strong-user-optimal case: (c) the hybrid digital-analog scheme.

L components. For these components, we define D_k'' such that $(1/2)\log(\sigma_k^2/D_k'') = (1/2)\log(1 + P_k/N_s)$. For the rest of the sub-channels ($m > L$) we define P_k'' by $(1/2)\log(\sigma_m^2/D_m) = (1/2)\log(1+(P_m - P_m'')/(P_m'' + N_s))$. The first L source components are sent uncoded scaled by $\sqrt{P_k/\sigma_k^2}$ on their corresponding sub-channels, and rest of the source components are sent uncoded scaled by $\sqrt{(P_k - P_k'')/\sigma_k^2}$ on their corresponding sub-channels. The first L source components are Wyner-Ziv source coded at rates of $(1/2)\log(D_k''/D_k)$ assuming the availability at the decoder (strong user) of the noise corrupted versions sent over the corresponding sub-channels. These source coded bits are sent using dirty-paper coding over the rest of the sub-channels $m > L$. The resulting distortion for the weak user is

$$D_w = \frac{1}{K}\left(\sum_{k=1}^{L}\frac{\sigma_k^2}{1+\frac{P_k}{N_w}} + \sum_{k=L+1}^{K}\frac{\sigma_k^2}{1+\frac{P_k - P_k''}{N_w}}\right)$$

which is strictly less than $\frac{1}{K}\sum_{k=1}^{L}\sigma_k^2$, the distortion for the weak user in the separation approach under which no information is decodable by this user.

This scheme also directly extends to the bandwidth expansion ($K < M$) and bandwidth contraction ($K > M$) scenarios. To summarise, we can state the following

Theorem 7.2.2 *The distortion pair* (D_s^*, D_w) *is achievable for the source-channel coding problem in section 7.1, where the point-to-point optimal distortion* D_s^* *for the strong-user and* D_w *are as follows.*

$$D_s^* = \frac{1}{K}\sum_{k=1}^{K} D_k, \text{ where } D_k = \begin{cases} \mu, & \text{if } \mu < \sigma_k^2, \\ \sigma_k^2, & \text{if } \mu \geq \sigma_k^2, \end{cases} \tag{7.2}$$

where μ *is chosen such that the total rate* $(1/2)\sum_{k=1}^{K}\log(\sigma_k^2/D_k)$ *equals the capacity* C_s *of the weak user's channel.* $C_s = \sum_{m=1}^{M}\log(1 + P_m/N_s)$ *where* $P_1 = P_2 = \ldots = P_M = P$.

Let

$$L = \min\left\{\left|\left\{k : \frac{\sigma_k^2}{D_k} \geq 1 + \frac{P_k}{N_s}\right\}\right|, M\right\},$$

$$K' = \min\left\{\left|\left\{k : \mu \leq \sigma_k^2\right\}\right|, K\right\},$$

and the P_k''*'s be defined by*

$$\frac{\sigma_m^2}{D_m} = 1 + \frac{P_m - P_m''}{P_m'' + N_s}.$$

Then,

$$D_w = \frac{1}{K}\left(\sum_{k=1}^{L}\frac{\sigma_k^2}{1 + \frac{P_k}{N_w}} + \sum_{k=L+1}^{K'}\frac{\sigma_k^2}{1 + \frac{P_k - P_k''}{N_w}} + \sum_{k=K'}^{K}\sigma_k^2.\right)$$

7.3 An achievable trade-off

We may also trade-off the quality of reproductions at the two users without being optimal for either. Clearly, time sharing between the two achievable extreme points is a possibility. It is often possible to do better. A natural strategy suggested by the above discussion is to combine the schemes for the weak- and strong-user-optimal cases. For the $K = M = 2$ case with $\sigma_1^2 > \sigma_2^2$, we can quantise S_1 to a distortion level D_1', say. The corresponding bits are sent to the weaker user at a power level $P_2 - P_2'$ assuming that the rest of the power P_2' in this sub-channel will interfere. i.e.,

$$\frac{1}{2}\log\left(\frac{\sigma_1^2}{D_1'}\right) = \frac{1}{2}\log\left(1 + \frac{P_2 - P_2'}{P_2' + N_w}\right).$$

Both decoders first decode this bitstream and cancel the interference caused by it. The quantisation error on S_1 is sent over the first sub-channel at a power level P_1 uncoded. The weak user estimates the quantisation error on S_1 from its output of the first sub-channel. S_2 is sent uncoded over the second sub-channel at a power level of $P_2' - P_2''$. This is treated as side information known at the transmitter for sending Wyner-Ziv bits at the rate $(1/2)\log(1 + P_2''/N_s)$ on the quantisation error of S_1 to the strong user using dirty-paper coding. Both users estimate S_2 from their outputs of the second sub-channel after cancelling the interference from the quantisation bitstream of S_1. The strong user decodes the dirty-paper coded bitstream from the second sub-channel. These Wyner-Ziv bits are decoded to produce $\widehat{S}_1$ using as side-information the output of the first sub-channel which is a noise corrupted version of S_1. By choosing different values for the power allocations P_1, P_2, P_2', P_2'' such that $P_1 + P_2 = 2P$ and $P_2 \geq P_2' \geq P_2''$, we will get different achievable points. Further flexibility in the form of allowing uncoded transmission of S_1 over the second sub-channel with potentially S_2 being sent source coded and Wyner-Ziv coded might, in some cases, lead to operating points which are better.

In general, we have the following

Theorem 7.3.1 *For any choice of non-negative* $P_1, P_2, \ldots, P_M$, $D_1', D_2', \ldots, D_K'$, $D_1'', D_2'', \ldots, D_M''$, *integers* L, K', *and reals* $\lambda_{L+1}, \lambda_{L+2}, \ldots, \lambda_M$, $\gamma_{L+1}, \gamma_{L+2}, \ldots, \gamma_{K'}$ *in* $[0,1]$ *satisfying the conditions (7.3)-(7.8), the following* (D_s, D_w) *is achievable*

$$D_s = \frac{1}{K}\left(\sum_{k=1}^{L} D_k'' + \sum_{k=L+1}^{K'} \frac{\sigma_k^2}{\frac{(1-\lambda_k)P_k+N_s}{(1-\lambda_k)(1-\gamma_k)P_k+N_s}} + \sum_{k=K'+1}^{K} D_k''\right),$$

$$D_w = \frac{1}{K}\left(\sum_{k=1}^{L} \frac{D_k'}{\frac{P_k+N_w}{N_w}} + \sum_{k=L+1}^{K'} \frac{\sigma_k^2}{\frac{(1-\lambda_k)P_k+N_w}{(1-\lambda_k)(1-\gamma_k)P_k+N_w}} + \sum_{k=K'+1}^{K} D_k'\right).$$

The conditions are

$$\sum_{m=1}^{M} P_m \leq P, \tag{7.3}$$

$$D_k'' \leq D_k' \leq \sigma_k^2,\ k = 1, 2, \ldots, K, \tag{7.4}$$

$$L \in \{0, 1, \ldots, \min(K, M)\}, \tag{7.5}$$

$$K' \in \{L, L+1, \ldots, \min(K, M)\}, \tag{7.6}$$

$$\sum_{k=1}^{L} \log \frac{\sigma_k^2}{D_k'} + \sum_{k=K'+1}^{K} \log \frac{\sigma_k^2}{D_k'} \leq \sum_{m=L+1}^{M} \log \frac{P_m + N_w}{(1-\lambda_m)P_m + N_w}, \quad \textit{and} \tag{7.7}$$

$$\sum_{k=1}^{L} \log \frac{\frac{D_k'}{1+\frac{P_k}{N_s}}}{D_k''} + \sum_{k=K'+1}^{K} \log \frac{D_k'}{D_k''} \leq \sum_{m=L+1}^{K'} \log \frac{(1-\lambda_m)(1-\gamma_m)P_m + N_s}{N_s} + \sum_{m=K'+1}^{M} \log \frac{(1-\lambda_m)P_m + N_s}{N_s}. \tag{7.8}$$

The proof is relegated to appendix B.

7.4 Specialisation to memoryless sources and channels with bandwidth mismatch

A special case of the problem is when the source is also white, but has a bandwidth different from the channel. If we define the degree of mismatch by $\alpha = M/K$, we have

Theorem 7.4.1 *For the special case of the problem in section 7.1 with* $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_M^2 = \sigma_S^2$*, the following* (D_s, D_w) *trade-off is achievable:*

- *For* $\alpha < 1$ *(bandwidth contraction)*

$$\left\{ \left(D_s^{BC}(\lambda,\gamma), D_w^{BC}(\lambda,\gamma) \right) : 0 \leq \lambda \leq 1, 0 \leq \gamma \leq 1 \right\},$$

where

$$D_w^{BC}(\lambda,\gamma) = \frac{\alpha\sigma_S^2}{\frac{(1-\lambda)P+N_w}{(1-\lambda)(1-\gamma)P+N_w}} + \frac{(1-\alpha)\sigma_S^2}{\left(\frac{P+N_w}{(1-\lambda)P+N_w}\right)^{\frac{\alpha}{1-\alpha}}}, \text{ and}$$

$$D_s^{BC}(\lambda,\gamma) = \frac{\alpha\sigma_S^2}{\frac{(1-\lambda)P+N_s}{(1-\lambda)(1-\gamma)P+N_s}} + \frac{(1-\alpha)\sigma_S^2}{\left(\frac{P+N_w}{(1-\lambda)P+N_w}\frac{(1-\lambda)(1-\gamma)P+N_s}{N_s}\right)^{\frac{\alpha}{1-\alpha}}}.$$

- *For $\alpha > 1$ (bandwidth expansion)*

$$\left\{\left(D_s^{BE}(\lambda), D_w^{BE}(\lambda)\right) : 0 \leq \lambda \leq 1\right\},$$

where

$$D_w^{BE}(\lambda) = \frac{\sigma_S^2}{\left(\frac{P+N_w}{(1-\lambda)P+N_w}\right)^{\alpha-1}\left(\frac{P+N_w}{N_w}\right)}, \text{ and}$$

$$D_s^{BE}(\lambda) = \frac{\sigma_S^2}{\left(\frac{P+N_w}{(1-\lambda)P+N_w}\right)^{\alpha-1}\left(\frac{P+N_s}{N_s}\right)\left(\frac{(1-\lambda)P+N_s}{N_s}\right)^{\alpha-1}}.$$

We prove this as a special case of Theorem 7.3.1 in appendix A.

As pointed out earlier, many researchers have investigated this special case. The trade-off for the bandwidth expansion case above also appears in [38], but the bandwidth contraction case is new. The following remarks on the extreme points of these trade-offs are in order.

- At the weak-user-optimal points, our achievable schemes under bandwidth contraction and bandwidth extraction reduce to the schemes proposed by Mittal and Phamdo [32].

- At the strong-user-optimal point under bandwidth expansion, our scheme reduces to the systematic lossy source-channel codes of Shamai, Verdù, and Zamir [40].

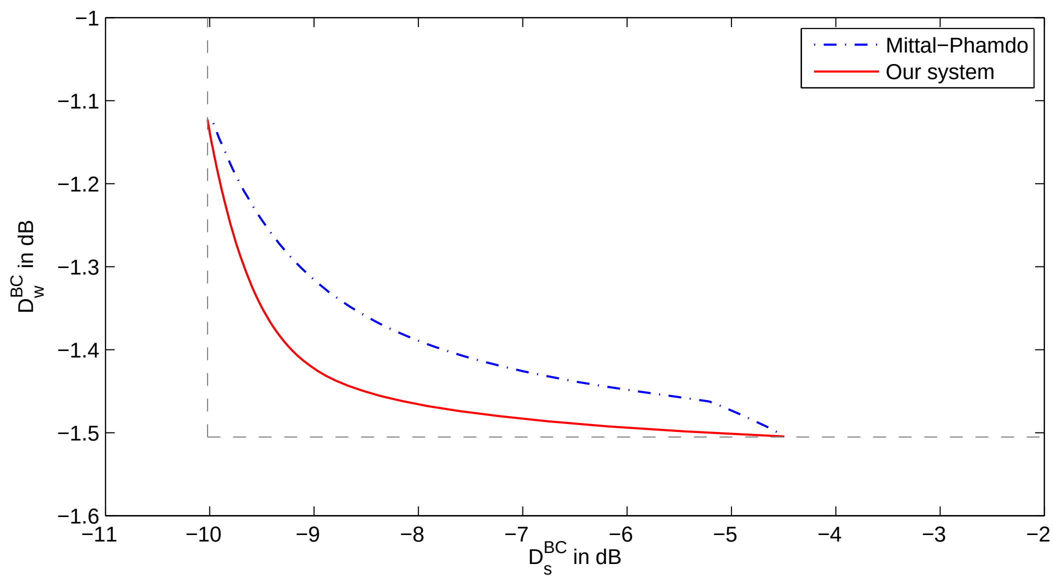
- At the strong-user-optimal points, the achievable scheme is strictly better than the solution offered by Mittal and Phamdo in [32]. The gap can be computed explicitly to be

$$\text{(bandwidth contraction)} \qquad \frac{\alpha\sigma_S^2 N_s}{N_w + P}\left(1 - \frac{1}{\left(1+\frac{P}{N_s}\right)^{\alpha}}\right), \quad \text{and}$$

$$\text{(bandwidth expansion)} \qquad \frac{\sigma_S^2}{\left(1+\frac{P}{N_s}\right)^{\alpha}\left(1+\frac{N_w}{P}\right)}.$$

  As pointed out above at the weak-user-optimal point, the schemes coincide. For the boundary points in between, an explicit computation is cumbersome, but numerical computation over a wide range of settings suggest that the achievable scheme strictly out performs the schemes of Mittal and Phamdo. Fig. 7.3 shows a comparison of the trade-offs achieved by our scheme with those of Mittal and Phamdo [32] for a few examples.

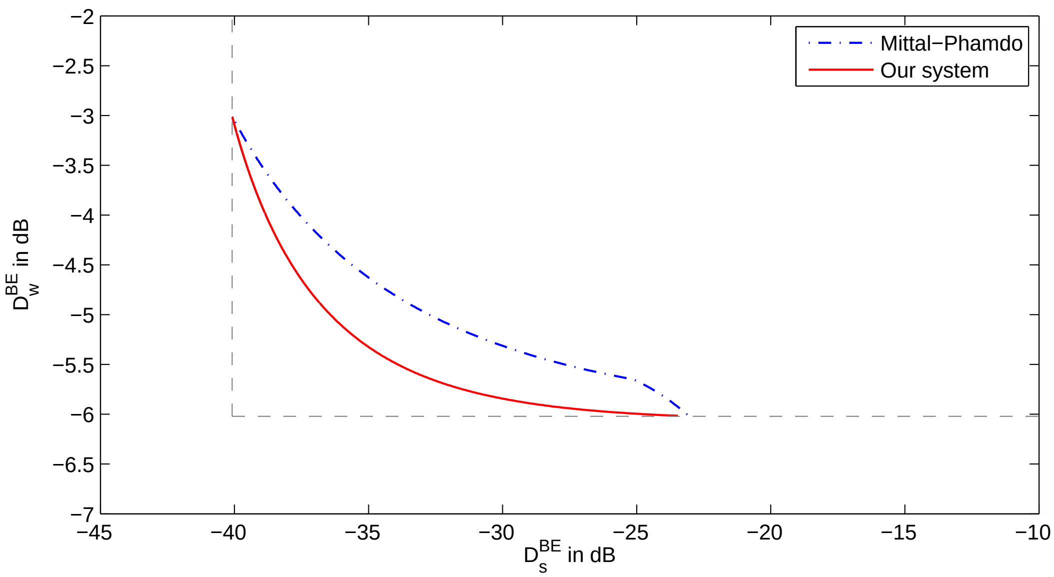
Outer bounds Only one non-trivial outer bound is available in the literature for this problem. It is due to Reznic, Feder, and Zamir [38] who developed it for the case of memoryless source and channel under bandwidth expansion. This bound, however, does not match the best available inner bound described above. The same bounding technique can be used to derive outer bounds for the coloured source-



(a)

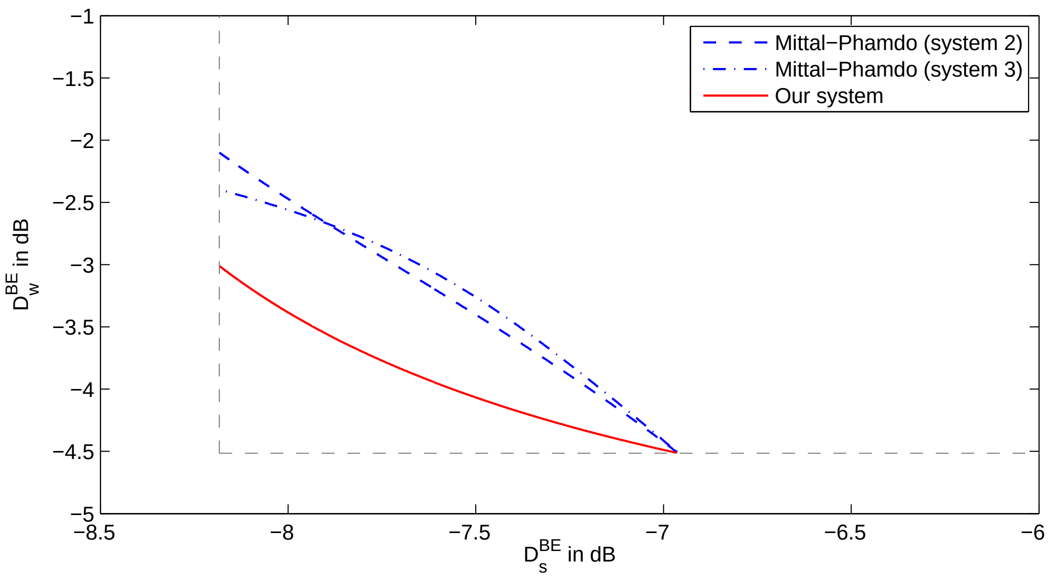
Figure 7.3: Comparison of distortion trade-offs achieved by our scheme with that of Mittal and Phamdo: (a) Bandwidth contraction. $\sigma_S^2 = 1$, $P/N_s = 20$dB, $P/N_w =$ 0dB, $\alpha = 0.5$. The best scheme suggested by Mittal and Phamdo was chosen for comparison [32, Fig. 14]. The dashed lines are drawn at the weak and strong user optimal distortions and thus give the trivial outer bound to the trade-off region. The strong-user-optimal points appear to coincide, but there is a small gap which is not visible at the scale of this plot. The weak user optimal points indeed coincide where both schemes reduce to the same scheme.

channel problem considered here. However, it does not always lead to a non-trivial bound. For instance, for memoryless source and channel with bandwidth contraction, the technique yields the trivial bound (resulting from considering the point-to-point source-channel problems involving either the weak user or the strong users alone).



(b)

Figure 7.3: Comparison of distortion trade-offs achieved by our scheme with that of Mittal and Phamdo: (b) Bandwidth expansion. $\sigma_S^2 = 1$, $P/N_s = 20$dB, $P/N_w = 0$dB, $\alpha = 2.0$. Again the best scheme of Mittal and Phamdo for this setting was chosen for comparison [32, Fig. 12]. The dashed lines are drawn at the weak and strong-user-optimal distortions and thus give the trivial outer bound to the trade-off region. The strong-user-optimal points appear to coincide, but there is a small gap which is not visible at the scale of this plot. The weak-user-optimal points indeed coincide where both schemes reduce to the same scheme.



(c)

Figure 7.3: Comparison of distortion trade-offs achieved by our scheme with those of Mittal and Phamdo: (b) Bandwidth expansion. $\sigma_S^2 = 1$, $P/N_s = 4$dB, $P/N_w = 0$dB, $\alpha = 1.5$. Two schemes of Mittal and Phamdo (systems 2 and 3) together give the best performance of all the new schemes proposed in [32, Fig. 13]. The dashed lines are drawn at the weak and strong user optimal distortions and thus give the trivial outer bound to the trade-off region. The gap between the strong-user-optimal points is visible in this plot. The weak-user-optimal points coincide.

Appendix A

Proof of Theorem 7.4.1

The proof for the bandwidth expansion case ($\alpha = M/K > 1$) follows from the following choice of parameters in theorem 7.3.1 : $L = K' = K$. Thus only the first term in the expressions for D_s and D_w in theorem 7.3.1 is present. We also choose $D'_k = D', D''_k = D'',\ k = 1, \ldots, K$ and $P_m = P,\ m = 1, \ldots, M$. And $\lambda_m = \lambda,\ \gamma_m = 0\ m = K+1, \ldots, M$. The primed quantities are chosen to satisfy the conditions (7.7) and (7.8). i.e.,

$$K \log \frac{\sigma_S^2}{D'} = (M-K) \log \frac{P + N_w}{(1-\lambda)P + N_w},$$

$$K \log \frac{\frac{D'}{1+\frac{P}{N_s}}}{D''} = (M-K) \log \frac{(1-\lambda)P + N_s}{N_s}.$$

Substituting these in the expression for the achievable (D_s, D_w) gives the result.

The choice of parameters for the bandwidth contraction case is: $L = 0,\ K' = M$, and thus the first term in the expressions for D_s and D_w in theorem 7.3.1 is absent here. We let $D'_k = D'$, $D''_k = D''$, $k = M+1, \ldots, K$, and $P_m = P$, $m = 1, \ldots, M$.

Also, $\lambda_m = \lambda$, $\gamma_m = \gamma$, $m = 1, \ldots, M$, and

$$(K-M)\log\frac{\sigma_S^2}{D'} = M\log\frac{P+N_w}{(1-\lambda)P+N_w},$$
$$(K-M)\log\frac{D'}{D''} = M\log\frac{(1-\lambda)(1-\gamma)P+N_s}{N_s}.$$

These choices give the achievability result for bandwidth contraction.

Appendix B

Proof of Theorem 7.3.1

The main ideas involved have already been described in section 7.2. We sketch the main steps of the proof which use the following results: successive refinement source coding [16], source coding with side-information or Wyner-Ziv (WZ) coding [57], super-position broadcast channel coding [12], and channel coding with side-information or Gel'fand-Pinsker (GP) coding [17].

The m-th sub-channel is allocated a power of P_m such that it satisfies the power constraint by (7.3). The coding will be performed as usual on block-length n sequences of sufficient length that the source codes invoked below have distortions close to optimal and the channel codes have low probabilities of errors. The sketch below does not attempt to make this formal.

Source components 1 through L An n-length block of the k-th such source component is source-coded (quantised) using an optimal source-code at distortion

D'_k. This codeword will be made available to both the strong and the weak receivers. The rate required to do this is

$$\sum_{k=1}^{L} \frac{1}{2} \log \frac{\sigma_k^2}{D'_k}.$$

Let the decoder reconstruction of the i-th sample be $\hat{S}'_k(i)$.

The quantisation error of the k-th source component is transmitted over the k-th sub-channel using power P_k. In other words, the input to the k-th sub-channel is

$$X_k(i) = \sqrt{\frac{P_k}{D'_k}}(S_k(i) - \hat{S}'_k(i)),\ i = 1, 2, \ldots, n.$$

To produce its reconstruction, the weak-user adds to $\hat{S}'_k(i)$ the LLSE estimate of $S_k(i) - \hat{S}'_k(i)$ from $Y_{wk}(i)$ at an MSE estimation error of $\frac{D'_k}{1+\frac{P_k}{N_w}}$. This gives the first term in the expression for D_w in the theorem. The strong user also performs the same to get an intermediate reconstruction of the source component at MSE distortion of $\frac{D'_k}{1+\frac{P_k}{N_s}}$. We would like to enhance this to a distortion of D''_k in the first term for the expression for D_s in the theorem. Using an extension of Wyner and Ziv's result [57, Sec. IV] to the continuous alphabet case [33], this can be done even if the encoder does not have access to the strong user's reconstruction (which is the case here). The encoder sends codewords to the strong user at a rate

$$\sum_{k=1}^{L} \frac{1}{2} \log \frac{\frac{D'_k}{1+\frac{P_k}{N_s}}}{D''_k}.$$

Source components $L + 1$ **through** K' The k-th such source component is sent uncoded on the k-th sub-channel using power $\gamma_k(1 - \lambda_k)P_k$. The rest of the power spent on this sub-channel is utilised to send the coded parts as will be discussed below. However, we need to note the fact that both decoders will be able to decode a coded part sent using power $(1-\gamma_k)P_k$ on this sub-channel and subtract it off before estimating S_k. The rest of the coded part which is sent at power $(1-\gamma_k)(1-\lambda_k)P_k$ will act as interference. Hence the MSE distortion for the k-th source component will be

$$D_{j\,k} = \frac{\sigma_k^2}{\frac{(1-\lambda_k)P_k+N_k}{(1-\lambda_k)(1-\gamma_k)P_k+N_j}},\ j \in \{s, w\}.$$

This gives the second term in the expressions for D_s and D_w in the theorem.

Source components $K'+1$ through K These source components will have no uncoded component unlike the above two cases. Thus they are source coded using an optimal successive refinement code [16] which works at a coarse description distortion D'_k, and fine description distortion of D''_k. These give the last term in the expressions for D_w and D_s respectively in the theorem. The bitrate for the coarse description which will be made available to both the users is

$$\sum_{k=K'+1}^{K} \frac{1}{2} \log \frac{\sigma_k^2}{D'_k},$$

and the refinement layer which will be made available only to the strong user has a bitrate of

$$\sum_{k=K'+1}^{K} \frac{1}{2} \log \frac{D'_k}{D''_k}.$$

Thus over all we need to send codewords at a rate of

$$\sum_{k=1}^{L} \frac{1}{2} \log \frac{\sigma_k^2}{D_k'} + \sum_{k=K'+1}^{K} \frac{1}{2} \log \frac{\sigma_k^2}{D_k'} \tag{B.1}$$

to both the users, and in addition enhancement codewords at the rate of

$$\sum_{k=1}^{L} \frac{1}{2} \log \frac{\frac{D_k'}{1+\frac{P_k}{N_s}}}{D_k''} + \sum_{k=K'+1}^{K} \frac{1}{2} \log \frac{D_k'}{D_k''} \tag{B.2}$$

to the strong user. In the next two steps we show how this is accomplished.

Sub-channels $L+1$ through K' As discussed above, on the m-th sub-channel a power of $\gamma_m(1-\lambda_m)P_m$ is used for uncoded transmission. The rest of the power is allocated as follows: $\lambda_m P_m$ is used for sending bits to both the receivers. The rest of the power $(1-\gamma_m)(1-\lambda_m)P_k$ is used to send bits which will be decoded only the the strong receiver. When decoding the common bits, both receivers treat the rest of the power as interference. Hence the bitrate of the common part is limited by the weaker user resulting in a rate of

$$\sum_{m=L+1}^{K'} \frac{1}{2} \log \frac{P_m + N_w}{(1-\lambda_m)P_m + N_w}.$$

As mentioned earlier, upon decoding, both users will subtract the codeword corresponding to the decoded common bits from their received signal. To send additional bits to the strong user, we use the concept of dirty-paper coding [11] [17]. The uncoded transmission can be thought of as Gaussian side-information (or "dirt") which is known at the encoder. From Costa [11, Sec. II], we know that a power allocation of $(1-\lambda_m)(1-\gamma_m)P_m$ can support a rate of

$$\sum_{m=L+1}^{K'} \frac{1}{2}\log\frac{(1-\lambda_m)(1-\gamma_m)P_m+N_s}{N_s}$$

to the strong user.

Sub-channels $K'+1$ through M In these sub-channels no uncoded transmission is performed. We use the idea of superposition coding [12] to deliver a common bitstream to both the users and in addition a refinement bitstream to only the stronger user. With a power allocation of $\lambda_m P_m$ to the common bitstream and the rest $(1-\lambda_m)P_m$ to the refinement bitstream, we get the following bitrates for the common and refinement bit streams respectively

$$\sum_{m=K'+1}^{M} \frac{1}{2}\log\frac{P_m+N_w}{(1-\lambda_m)P_m+N_w}$$
$$\sum_{m=K'+1}^{M} \frac{1}{2}\log\frac{(1-\lambda_m)P_m+N_s}{N_s}.$$

.

Thus the total rate available for sending a common bitstream to both the users is

$$\sum_{m=L+1}^{M} \frac{1}{2}\log\frac{P_m+N_w}{(1-\lambda_m)P_m+N_w}.$$

Condition (7.8) ensures that this is sufficient to handle the rate of the common bit stream in (B.1). Similarly, the total rate available for sending an enhancement bitstream to the stronger user is

$$\sum_{m=L+1}^{K'} \frac{1}{2} \log \frac{(1-\lambda_m)(1-\gamma_m)P_m + N_s}{N_s} + \sum_{m=K'+1}^{M} \frac{1}{2} \log \frac{(1-\lambda_m)P_m + N_s}{N_s}$$

which is larger than the rate of the enhancement bitstream in (B.2) by condition (7.8). This completes the sketch of the proof. □

Chapter 8

Conclusions and Future Work

In Part I we characterised the optimal performance of the quadratic Gaussian CEO problem. The primary limitation of the CEO model is that conditioned on the quantity of interest to the decoder, the observations of the sensors were modelled to be conditionally independent. In practice, this may not always be true. As we discussed earlier, the optimality of Berger-Tung-Housewright strategies for the the quadratic Gaussian CEO problem has been extended to a slightly more general Gaussian model where the sensor observations satisfy a certain Markov condition. However, this extension does not cover all the possibilities, and it is suspected that Berger-Tung-Housewright strategies may not be sufficient to achieve the optimal performance even for the quadratic Gaussian case under all settings. Other strategies along the lines of Körner and Marton [30] may be needed. The CEO problem itself remains open for non-Gaussian distributions.

We also presented extensions to scenarios where the sensors are prone to failures and also when the sensors may communicate with each other. The results we presented were of a limited nature. While we were able to characterise a corner point of the distortion trade-off region under different sensor failure scenarios, a complete characterisation of this region remains open. Also, we showed that the sum rate performance of the Gaussian CEO problem cannot be improved by allowing the sensors to co-operate by communicating over rate constrained links. However, a characterisation of the rate region remains open even in the quadratic Gaussian setting. In section 4.3, we presented a simple three-user Gaussian problem which is still open and is perhaps one of the simpler examples which needs to be resolved before the question of the rate region under an interactive model can be solved.

We discussed source-channel broadcast of Gaussian sources over Gaussian broadcast channels in Part II. While we could find the optimal trade-offs in quality when serving an analog receiver and a digital receiver (with potentially different channels from the source) simultaneously, only an achievable trade-off is available for two digital receivers with different channels. We believe that the limitation is primarily due to the lack of good outer bounds on the region of distortions that can be supported simultaneously. Even in simple cases like when the source and channel are memoryless, but with mismatching bandwidths, a tight result is not available. In fact, the best available outer bound for the case where the source bandwidth is larger than the channel bandwidth is the trivial outer bound which considers the receivers separately.

It is also not clear if codes with more structure can be used to obtain better trade-offs. These could be subjects of further investigation.

Bibliography

[1] MOSEK ApS optimization software. Website. `http://www.mosek.com`.

[2] IF Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *Communications Magazine, IEEE*, 40(8):102–114, 2002.

[3] S. Battista, F. Casalino, and C. Lande. MPEG-4: A Multimedia Standard for the Third Millennium. 1. *IEEE Multimedia*, 6(4):74–83, October-December 1999.

[4] T. Berger. *Rate Distortion Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[5] T. Berger. *Multiterminal Source Coding*, pages 171–231. Springer-Verlag, 1977.

[6] T. Berger and D. W. Tufts. Optimum Pulse Amplitude Modulation Part I: Transmitter-Receiver Design and Bounds from Information Theory. *IEEE Transactions on Information Theory*, 13(2):196–208, April 1967.

[7] T. Berger and R. W. Yeung. Multiterminal Source Encoding with One Distortion Criterion. *IEEE Transactions on Information Theory*, 35(2):228–236, Mar 1989.

[8] T. Berger, Z. Zhang, and H. Viswanathan. The CEO Problem. *IEEE Transactions on Information Theory*, 42(3):887–902, May 1996.

[9] R. S. Blum, S. A. Kassam, and V. Poor. Distributed detection with multiple sensors I. Advanced topics. *Proceedings of the IEEE*, 85(1):64–79, 1997.

[10] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, U.K., 2004.

[11] M. H. M Costa. Writing on Dirty Paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.

[12] T. M. Cover. Broadcast Channels. *IEEE Transactions on Information Theory*, 18(1):2–14, January 1972.

[13] T. M. Cover. A Proof of the Data Compression Theorem of Slepian and Wolf for Ergodic Sources. *IEEE Transactions on Information Theory*, 21:226–228, Mar 1975.

[14] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.

[15] S. C. Draper and G. W. Wornell. Side information aware coding strategies for sensor networks. *Selected Areas in Communications, IEEE Journal on*, 22(6):966–976, 2004.

[16] W. H. R. Equitz and T. M. Cover. Successive Refinement of Information. *IEEE Transactions on Information Theory*, 37(2):269–275, March 1991.

[17] S. I. Gel'fand and M. S. Pinsker. Coding for Channel with Random Parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[18] T. J. Goblick, Jr. Theoretrical Limitations on the Transmission of Data from Analog Sources. *IEEE Transactions on Information Theory*, 11(4):558–567, October 1965.

[19] S. Hadim and N. Mohamed. Middleware for Wireless Sensor Networks: A Survey. *Communication System Software and Middleware, 2006. Comsware 2006. First International Conference on*, pages 1–7, 2006.

[20] B. G. Haskell, A. Puri, and A. N. Netravali. *Digital Video: An Introduction to MPEG-2*. Kluwer Academic Publishers, 1996.

[21] Jason Hill, Robert Szewczyk, Alec Woo, Seth Hollar, David Culler, and Kristofer Pister. System architecture directions for networked sensors. *SIGOPS Oper. Syst. Rev.*, 34(5):93–104, 2000.

[22] K. B. Housewright. *Source Coding for Multiterminal Systems*. PhD thesis, University of California, Los Angeles, 1977.

[23] International Organization for Standardization. *Digital Compression and Cod-*

ing of Continuous-tone Still Images: Requirements and Guidelines. ISO/IEC JTC 1/SC 29 10918-1, 1 edition, 1994.

[24] P. Ishwar, R. Puri, K. Ramchandran, and SS Pradhan. On Rate-Constrained Distributed Estimation in Unreliable Sensor Networks. *Selected Areas in Communications, IEEE Journal on*, 23(4):765–775, 2005.

[25] *Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: Video (MPEG-2), 2nd edition*, 2nd edition, 2000.

[26] *Information Technology – Generic Coding of Moving Pictures and Associated Audio Information – Part 3: Audio (MP3), 2nd edition*, 2nd edition, 1998.

[27] Q. Jiang and D. Manivannan. Routing protocols for sensor networks. *Consumer Communications and Networking Conference, 2004. CCNC 2004. First IEEE*, pages 93–98, 2004.

[28] A. Kaspi and T. Berger. Rate-distortion for Correlated Sources with Partially Separated Encoders. *IEEE Transactions on Information Theory*, 28(6):828–840, Nov 1982.

[29] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon. Wireless sensor networks for structural health monitoring. *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 427–428, 2006.

[30] J. Körner and K. Marton. How to Encode the Modulo-Two Sum of Binary Sources. *IEEE Transactions on Information Theory*, 25(2):219–221, Mar 1979.

[31] N. Merhav and S. Shamai. On joint source-channel coding for the Wyner-Ziv source and the Gel'fand-Pinsker channel. *IEEE Transactions on Information Theory*, 49(11):2844–2855, November 2003.

[32] U. Mittal and N. Phamdo. Hybrid Digital-Analog (HDA) Joint Source-Channel Codes for Broadcasting and Robust Communications. *IEEE Transactions on Information Theory*, 48(5):1082–1102, May 2002.

[33] Y. Oohama. Gaussian Multiterminal Source Coding. *IEEE Transactions on Information Theory*, 43(6):1912–1923, Nov 1997.

[34] Y. Oohama. The Rate-Distortion Function for the Quadratic Gaussian CEO Problem. *IEEE Transactions on Information Theory*, 44(3):55–67, May 1998.

[35] Y. Oohama. Multiterminal Source Coding for Correlated Meemoryless Gaussian Sources with Several Side Informations at the Decoder. In *IEEE Information Theory Workshop*, page 100, Kruger National Park, South Africa, 1999.

[36] Y. Oohama. Rate-Distortion Theory for Gaussian Multiterminal Source Coding Systems with Several Side Informations at the Decoder. *IEEE Transactions on Information Theory*, 51(7):2577–2593, Jul 2005.

[37] L. Ozarow. On a Source Coding Problem with Two Channels and Three Receivers. *Bell System Technical Journal*, 59(10):1909–1921, December 1980.

[38] Z. Reznic, M. Feder, and R. Zamir. Distortion Bounds for Broadcasting with Bandwidth Expansion. *IEEE Transactions on Information Theory*, 52(8):3778–3788, August 2006.

[39] K. Romer and F. Mattern. The design space of wireless sensor networks. *Wireless Communications, IEEE [see also IEEE Personal Communications]*, 11(6):54–61, 2004.

[40] S. Shamai, S. Verdu, and R. Zamir. Systematic Lossy Source/Channel Coding. *IEEE Transactions on Information Theory*, 44(2):564–579, March 1998.

[41] C. E. Shannon. The Mathematical Theory ofCommunication. *Bell System Technical J*, 27:379–423, 1948.

[42] C. E. Shannon. Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers*, 37:10–21, 1949.

[43] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Natl. Convention Rec. Part 4*, pages 142–163, 1959.

[44] D. Slepian and J. K. Wolf. Noiseless Coding of Correlated Information Sources. *IEEE Transactions on Information Theory*, 19:471–480, July 1973.

[45] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin. Habitat monitoring with sensor networks. *Communications of the ACM*, 47(6):34–40, 2004.

[46] S. Tavildar, P. Viswanath, and A. B. Wagner. The Gaussian Many-Help-One Distributed Source Coding Problem. In *IEEE Information Theory Workshop*, pages 596–600, Chengdu, China, October 2006.

[47] R. R. Tenney and N. R. Sandell. Detection with Distributed Sensors. *Aerospace and Electronic Systems, IEEE Transactions on*, pages 501–510, 1981.

[48] M. D. Trott. Unequal error protection codes: Theory and practice. In *Proc. IEEE Information Theory Workshop*, page 11, Haifa, Israel, June 1996.

[49] J. N. Tsitsiklis. *Decentralized detection*, pages 297–344. JAI, Greenwich, CT, 1993.

[50] S. Y. Tung. *Multiterminal Source Coding*. PhD thesis, School of Electrical Engineering, Cornell University, Ithaca, NY, May 1978.

[51] P. Viswanath. Sum Rate of Gaussian Multiterminal Source Coding. In P. Gupta, G. Kramer, and A. J van Wijngaarden, editors, *Advances in Network Information Theory*, volume 66 of *DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, pages 43–60. American Mathematical Society and DIMACS, 2004.

[52] H. Viswanathan and T. Berger. The Quadratic Gaussian CEO Problem. *IEEE Transactions on Information Theory*, 43(5):1549–1559, Sep 1997.

[53] R. Viswanathan and P. K. Varshney. Distributed detection with multiple sensors I. Fundamentals. *Proceedings of the IEEE*, 85(1):54–63, 1997.

[54] A. B. Wagner and V. Anantharam. An infeasibility result for the multiterminal source-coding problem, 2005. submitted to *IEEE Trans. Inf. Theory*. `http://www.arXiv.org/abs/cs/0511103`.

[55] A. B. Wagner, S. Tavildar, and P. Viswanath. Rate region of the quadratic gaussian two-terminal source-coding problem, 2006. submitted to *IEEE Trans. Inf. Theory*. `http://www.arxiv.org/abs/cs/0510095`.

[56] A. D. Wyner. The Rate-Distortion Function for Source Coding with Side Information at the Decoder-II: General Sources. *Information Control*, 38:60–80, July 1978.

[57] A. D. Wyner and J. Ziv. The Rate-Distortion Function for Source Coding with Side Information at the Decoder. *IEEE Transactions on Information Theory*, 22(1):1–10, January 1976.