Quantization in Acquisition and Computation Networks

by

John Z. Sun

B.S., Cornell University (2007)
S.M., Massachusetts Institute of Technology (2009)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author .....
Department of Electrical Engineering and Computer Science
May 15, 2013

Certified by .....
Vivek K Goyal
Principle Research Scientist, Research Laboratory of Electronics
Thesis Supervisor

Accepted by .....
Professor Leslie A. Kolodziejski
Chair, Committee on Graduate Students

Quantization in Acquisition and Computation Networks

by

John Z. Sun

Submitted to the Department of Electrical Engineering and
Computer Science
on May 15, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

In modern systems, it is often desirable to extract relevant information from large amounts of data collected at different spatial locations. Applications include sensor networks, wearable health-monitoring devices and a variety of other systems for inference. Several existing source coding techniques, such as Slepian-Wolf and Wyner-Ziv coding, achieve asymptotic compression optimality in distributed systems. However, these techniques are rarely used in sensor networks because of decoding complexity and prohibitively long code length. Moreover, the fundamental limits that arise from existing techniques are intractable to describe for a complicated network topology or when the objective of the system is to perform some computation on the data rather than to reproduce the data.

This thesis bridges the technological gap between the needs of real-world systems and the optimistic bounds derived from asymptotic analysis. Specifically, we characterize fundamental trade-offs when the desired computation is incorporated into the compression design and the code length is one. To obtain both performance guarantees and achievable schemes, we use high-resolution quantization theory, which is complementary to the Shannon-theoretic analyses previously used to study distributed systems. We account for varied network topologies, such as those where sensors are allowed to collaborate or the communication links are heterogeneous. In these settings, a small amount of intersensor communication can provide a significant improvement in compression performance. As a result, this work suggests new compression principles and network design for modern distributed systems.

Although the ideas in the thesis are motivated by current and future sensor network implementations, the framework applies to a wide range of signal processing questions. We draw connections between the fidelity criteria studied in the thesis and distortion measures used in perceptual coding. As a consequence, we determine the optimal quantizer for expected relative error (ERE), a measure that is widely useful but is often neglected in the source coding community. We further demonstrate that applying the ERE criterion to psychophysical models can explain the Weber–

Fechner law, a longstanding hypothesis of how humans perceive the external world. Our results are consistent with the hypothesis that human perception is Bayesian optimal for information acquisition conditioned on limited cognitive resources, thereby supporting the notion that the brain is efficient at acquisition and adaptation.

Thesis Supervisor: Vivek K Goyal
Title: Principle Research Scientist, Research Laboratory of Electronics

Acknowledgments

"Spoon feeding in the long run teaches us nothing but the shape of the spoon."
– E.M. Forster

"We must find time to stop and thank the people who make a difference in our lives."
– John F. Kennedy

"It takes more than good memory to have good memories."
– Fortune cookie from Royal East

I first thank Vivek Goyal for his guidance and support during my six years at MIT. He is an inspirational researcher, wise advisor and valiant captain of the good ship STIR. I will keep the lessons he provided for the rest of my career and am fortunate to have him as a lifelong friend.

I also acknowledge my committee, Bob Gray, Sanjay Sarma and Greg Wornell for making this thesis possible. My respect for their wisdom and expertise inspires me to be a better researcher and writer.

I am proud to be a member of the Signal Transformation and Information Representation (STIR) Group. Its past and present members, Adam, Andrea, Dan, Dongeek, Ha, Jonathan, Joong, Kirmani, Ulugbek, Vahid and Vinith, are promising researchers who have bright futures. They are also great people and I will cherish our curling outings, hikes, March Madness brackets and dinners. Eric, Kathryn and Gabrielle are wonderful administrators and I am grateful for their diligence. I offer special thanks to Lav Varshney for always looking out for me at MIT. He is an incredible mentor, coauthor and academic, and I am excited to follow his career.

There are numerous collaborators and colleagues that deserve attention, including Kush Varshney of IBM Research, Jesse Weinstein–Gould and Wynn Sullivan of the Oklahoma City Thunder, and Szymon Jakubczak of CSAIL. I also appreciate conversations with Hye Won, Mina and Vincent during information theory reading groups and talking to other members of the sixth floor of Building 36. I especially thank Da Wang for being a friend and a sounding board for crazy ideas.

Beyond being a special place to do research, MIT is also a wonderful place to live and make friends. I greatly enjoyed serving the EECS Graduate Student Association (GSA), Resources for Easing Friction and Stress (REFS), the Tech newspaper and Sidney-Pacific House Government. The people I worked with in these organizations are amazing individuals who have my respect. I am proud to be part of an enthusiastic poker club, which provided stimulating entertainment on even the darkest days of graduate school. I also thank my old roommates from Cornell, Jon, Zinger and Mitch, for constantly demonstrating that the good old days never ended.

Finally and most importantly, I thank my family. I dedicate this thesis to my parents for their unwavering support and for being my role models. I am grateful for my sister Jenny's love and her contagious smile. Lastly, I am blessed to have Grace in my life. She is the best thing that will ever happen to me.

Contents

1 Introduction **13**

1.1 A Motivating Application 15

1.2 Outline and Contributions 16

1.3 Bibliographical Notes 20

2 Background **23**

2.1 Quantization 24

2.2 High-Resolution Scalar Quantization 27

2.2.1 Fixed-Rate Quantization 29

2.2.2 Entropy-Constrained Quantization 31

2.3 Quantization for Computation 32

2.4 Vector Quantization 37

2.5 Rate–Distortion Theory 39

2.5.1 Multiterminal Shannon Theory 40

2.5.2 Rate–Distortion Theory for Small Distortion 43

2.5.3 Other Related Problems 44

3 DFSQ Simplified **47**

3.1 Motivation 47

3.2 Univariate Functional Quantization 48

3.2.1 Remarks 50

3.2.2 Asymptotically Optimal Quantizer Sequences 51

3.2.3 Negligible Suboptimality of Simple Decoder 52

3.3 Multivariate Functional Quantization 54
3.3.1 Remarks 56
3.3.2 Asymptotically Optimal Quantizer Sequences 57
3.3.3 Vector-Valued Functions 58
3.4 Examples 59
3.4.1 Examples for Univariate Functional Quantization 59
3.4.2 Examples for Multivariate Functional Quantization 60
3.5 Conclusions 63
3.A Proof of Theorem 3.1 64
3.B Proof of Theorem 3.2 67
3.C Quantization for Weighted Distortion 70
3.D Generalizing Theorem 3.3 75
3.E Sources over Finite Support 75

4 Performance Loss of DFSQ **79**
4.1 Motivation 80
4.2 Rate–Distortion Behavior of Distributed Systems for W-MSE 81
4.3 Rate Loss of DFSQ 84
4.3.1 Remarks 86
4.4 Future Work 87

5 Chatting Networks **89**
5.1 Motivation 89
5.2 Preliminaries 91
5.2.1 Prior Work 91
5.2.2 Chatting in DFSQ 92
5.2.3 Don't-Care Intervals in DFSQ 93
5.3 Problem Model 94
5.4 Performance and Design of Chatting Networks 97
5.4.1 Don't-Care Intervals 98
5.4.2 Fixed-Rate Quantization Design 99

5.4.3 Entropy-Constrained Quantization Design 100
5.4.4 Conditions on Chatting Graph 100
5.5 Rate Allocation in Chatting Networks 102
5.6 Maximum Computation 104
5.6.1 Problem Model 104
5.6.2 Quantizer Design 105
5.6.3 Generalizing the Chatting Messages 110
5.6.4 Optimizing a Chatting Network 111
5.7 Conclusions 113
5.A Rate Allocation for Distributed Networks 114

6 Quantization for Relative Error **117**
6.1 Motivation 118
6.2 Optimal Quantization Design 119
6.3 Numerical Results 122
6.4 Generalizing ERE 123
6.5 Applications of ERE 125

7 Understanding Psychophysical Scales **127**
7.1 Motivation 128
7.2 A Bayes-Optimal Model for Limited Perception 131
7.3 A Bayes-Optimal Model for Limited Perception with Coding 134
7.4 Examples 136
7.4.1 Power-Law Stimulus Distributions 137
7.4.2 Natural Sounds 138
7.5 Discussion 141
7.5.1 Key Assumptions 142
7.5.2 Applicability of Theory 143
7.5.3 Future Directions 143
7.A Mathematical Methods for Analog Model 144

8 Conclusion **147**

List of Figures

1-1 Model of a distributed computation network 15

2-1 Companding model for design of nonuniform scalar quantizers 26
2-2 Point density function in the companding model 27
2-3 Source coding model of quantization 29
2-4 Distributed quantization model for DFSQ 33
2-5 The multiterminal source coding problem 41

3-1 Convergence of centroid and midpoint reconstructions 53
3-2 Empirical and theoretical performance for examples 61

5-1 Sample chatting graph 95
5-2 Motivating chatting example 96
5-3 Cost allocation in chatting networks 104
5-4 Maximum computation network 106
5-5 Performance of maximum computation network with chatting 110
5-6 Optimization of chatting bit 112
5-7 Distortion improvement using chatting 113

6-1 Codebooks optimized for relative error 123
6-2 Performance of scalar quantizers optimized for ERE 124
6-3 Codebooks optimized for gERE 125

7-1 Quantization model for human perception 132
7-2 Analog model for human perception 134

7-3 Quantization with coding model for human perception 135
7-4 Psychophysical scales for bounded power-law densities 139
7-5 Predicted psychophysical scales for acoustic datasets 140

Chapter 1

Introduction

Despite previous theoretical breakthroughs in understanding information compression for distributed networks, hardware limitations and deployment feasibility have historically inhibited the growth of practical systems such as sensor networks and cloud computing. However, recent innovations in implementation have now made distributed systems ubiquitous in many environments and applications. In fact, the rapid proliferation of such systems and the resulting data deluge [7] has opened a gap between theory and practice and introduced a difficult open problem—how can we efficiently eliminate information redundancy in large asynchronous networks where nodes have different objectives and appetites for data?

In this thesis, we consider a more modest objective of understanding how impactful information can be extracted out of data when sensing nodes know the computational aims of the network. With the insight that information communication is a precious commodity, we provide precise conditions under which incorporating system goals into acquisition and compression blocks can lead to more information per bit communicated. Further gains can be realized when sensors collaborate. We believe the insights and intuitions gained in this study can influence the design of efficient source coding algorithms, network architectures and sensor implementations. These components can then aid in constructing smarter systems that addresses the changing nature of data and information in a network.

There has been substantial effort to study distributed coding using information

theoretic concepts, taking advantage of long code blocks and powerful decoders to approach fundamental limits of compression. However, techniques inspired by this theory are infeasible for many applications. In particular, strong dependencies between source variables imply low information content per variable, but exploiting this is difficult under strict latency requirements. Other reasons why Shannon-style analysis may be a poor proxy for studying real-world systems include high complexity, need for large memory storage, unrealistic synchronization assumptions, and sensitivity to the accuracy of the assumed probabilistic models near the fundamental limit.

The existence of methods for which there is a very optimistic theory has strongly discouraged the development of alternative approaches. Thus, the failure of these methods in practice has left a glaring technological gap for understanding compression performance under application-specific constraints. In fact, most sensor network systems use simple uniform scalar quantization and either compression that does not exploit intersensor correlation or no compression at all. Rather than require long blocks, the complementary asymptotic of *high-resolution quantization theory* is more useful for these scenarios; most of this theory is focused on the scalar case, where the block length is one. The principal previous work in applying high-resolution quantization theory to the acquisition and computation network of Figure 1-1 is the *distributed functional scalar quantization* (DFSQ) framework [120]. The key message from DFSQ is that the design of optimal encoders for systems that perform nonlinear computations can be drastically different from what traditional source coding theory suggests. In recent years, ideas from DFSQ have been applied to compressed sensing [172], compression for media [173], and channel state feedback in wireless networks [151].

We will briefly summarize the contributions of this thesis in Section 1.2 after a short thought experiment that illustrates the usefulness of the overall work in Section 1.1. The publications developed for this thesis are listed in Section 1.3.

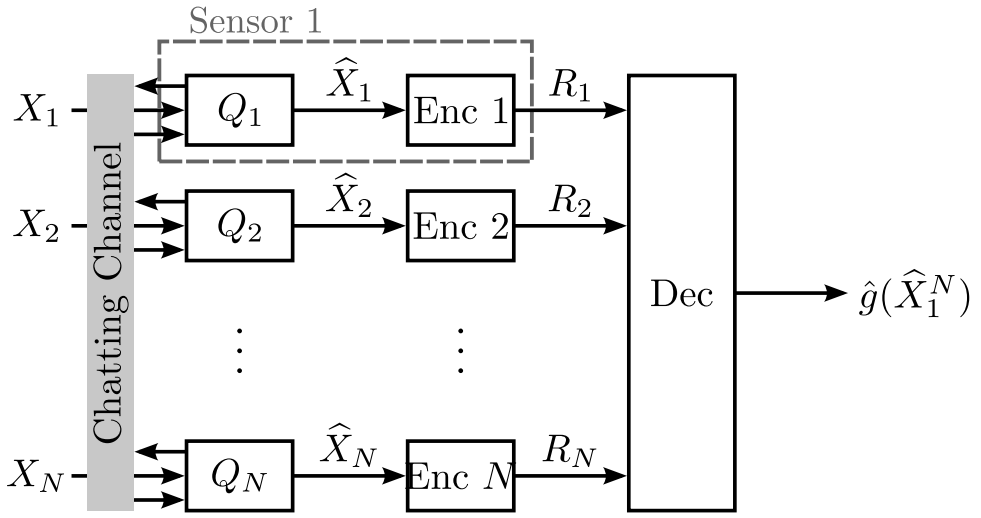


Figure 1-1: A distributed computation network, where N sensors (comprising quantizer and encoder) observe realizations of correlated sources. Each observation X_n is encoded and communicated with rate R_n to a fusion center. Simultaneously, each sensor can interact with a subset of other sensors using a noiseless but rate-limited chatting channel to improve compression. The decoder at the fusion center computes an estimate of the function $g(X_1^N)$ from the received data using a reconstruction function $\hat{g}(\widehat{X}_1^N)$ but cannot observe messages communicated on the chatting channel.

1.1 A Motivating Application

Communication through sensor networks is a major topic of interest. Unlike previous network architectures such as the Internet, sensor networks are designed to be disposable, cheap and application-specific. This not only creates innovations in hardware, but also new design constraints and desired properties. In particular, factors such as the goals at the fusion center, energy constraints, and computational abilities of individual nodes lead to domain-specific knowledge that can be exploited.

We provide a simple and admittedly contrived example that inspires some of the goals of the thesis. Consider sensors placed throughout a nuclear power plant that collect temperature readings and communicate to a fusion center that sounds an alarm if any temperature reading exceeds a threshold which may vary over time and be adaptive. Having good spatial coverage is important but too many sensors lead to communication interference, unneeded power consumption and high cost. Hence, a balance must be struck between the number of sensors, the complexity of the

communication topology, and the need for fast and reliable temperature information to indicate the health of the plant.

In this setting, stale information is not useful so the large-blocklength analysis is not applicable. Given constraints on the communication rates of the network, how should data be conveyed to the central fusion center? How can we fold in the objective of the network, which is to sound an alarm whenever *any* sensor reading is abnormally large? Generally, system designers will opt for simple design and will simply convey temperature data with good signal fidelity. However, this does not exploit the correlations among sensor data and, more importantly, the computation that governs the fusion center's objectives.

Using the ideas in this thesis, we will see that we can precisely analyze the effect of compression on the computation performed at the fusion center and how it scales with the communication rate. We can also determine the effect of correlations and intersensor communication over side channels, e.g. using cognitive radio [203]. The main insight is that incorporating the fusion center's goals into the compression at each sensor can yield substantial savings over just coding for source reproduction. As a result, the sensor network in this power plant can enjoy better estimates of sensor readings or allow for more sensors using the same amount of communication resources.

1.2 Outline and Contributions

In this thesis, we provide theoretical and empirical results for quantization in distributed systems described by the topology in Figure 1-1. The thesis has three major themes:

1. **Theoretical foundations** – provide fundamental understanding of the performance and design trade-offs of distributed functional scalar quantization for a variety of source models and computations.

2. **Intersensor collaboration** – demonstrate that communication between sensors can lead to dramatic improvement in compression performance when the block length of acquisition is restricted to be small.

3. **Coding for perceptually relevant distortion measures** – use the underlying theory developed in this thesis to understand perceptual coding and human perception.

We will focus our efforts on quantizer design. Concurring with Neuhoff [127], we feel that "high-resolution quantization theory has surpassed rate–distortion theory in its relevance to practical code design." High-resolution quantization theory gives successful architectural prescriptions and quantitative performance estimates. It thus provides a foundation for our invention and analysis toward the goal.

The outline of the thesis will be as follows:

Chapter 2 – Background

In this section, we provide a comprehensive survey of source coding results directly related to the thesis, with strong emphasis placed on the two asymptotic theories of high resolution and large blocklength. Of particular interest are results in DFSQ, vector quantization for nondifference distortion measures, multiterminal source coding and the Shannon lower bound for rate–distortion theory. We will also formally define ideas such as quantization and source coding, and describe the notation that will be used in the thesis.

Chapter 3 – DFSQ Simplified

The main analysis technique of the thesis is DFSQ; previous investigations have developed rigorous analyses to understand performance limits when sources have joint densities supported on a unit cube and the decoder uses a complicated joint centroid estimator. The aims of this chapter are to loosen the requirements on the source model and introduce a "simple" decoder that is the desired computation performed on the quantized measurements. Remarkably, the simpler decoder achieves the same

asymptotic performance as the optimal decoder even though it does not use the joint density of the source. Moreover, the communication and computation aspects of the decoding are decoupled, which is more useful in real-world applications. Our results demonstrate that similar analytical forms for performance arise in these new systems, demonstrating the generality of DFSQ theory. The new results are based on careful application of Taylor's theorem and lead to new insights on the rate of convergence of different quantization decoders. Performance results are given using Monte Carlo simulation of real quantizers, illustrating that the high-resolution approximation may be close to the performance of quantizers at low rates under mild conditions.

Chapter 4 – Performance Loss from DFSQ

High-resolution theory provides asymptotic achievability bounds on the best scalar quantizers. It is natural to wonder what the gap is between these results and the Shannon-style results that provide the ultimate performance bounds of compression, especially where the goal is to optimize fidelity of a function of source realizations. In this chapter, we develop bounds that characterize the achievable rate gap between DFSQ and Shannon-theoretic coding for distributed networks. We use existing results in Shannon theory that look at the behavior of the rate–distortion function for very small distortion, effectively a high-resolution large-blocklength asymptotic. In order to understand the performance gap, we introduce some new performance bounds for this regime.

Chapter 5 – Chatting Networks

A celebrated result in distributed source coding argues that intersensor collaboration, or "chatting" cannot improve compression performance by very much. However, this remarkable intuition relies on complicated decoders that exploit coding over long blocks; it does not hold for scalar quantizers. In this chapter, we show intersensor communication can improve the overall compression ability of the network. The key result is that chatting can dramatically decrease loss in reproduction fidelity even when intersensor communication is at very low rate. We also solve the rate

allocation problem when communication links have heterogeneous costs and provide a detailed example to demonstrate the theoretical and practical gains from chatting. This example for the maximum computation gives insight on the gap between chatting and distributed networks, and how to optimize the intersensor communication.

There are great opportunities to leverage chatting in many systems where communication between nearby nodes may be much cheaper than long-range communication. Thus, rather than insisting that the encodings be conditionally independent given the measurements, low-rate communication among neighboring nodes can exploit dependencies efficiently between measurements, and account for the goals at the fusion center.

Chapter 6 – Quantization for Expected Relative Error

In a separate problem formulation, we study quantization for expected relative error (ERE), a distortion measure that holds great engineering and scientific significance but has not gained much attention in the source coding community. We provide fundamental results on quantizer design and performance for ERE and draw connections to standards in the speech coding literature. Moreover, we introduce new applications for this theory, including coding for wireless channels and for human perception. The ERE measure is a type of nondifference distortion measures; this class has been most prominently used in the context of perceptual coding for media applications, e.g. speech. We discuss the relationship between these frameworks and functional scalar quantization in both the univariate and multivariate settings.

Chapter 7 – Understanding Psychophysical Scales

The results from optimizing coding for expected relative error are similar to observations in psychophysical experiments on human perception. This chapter aims to strengthen the applicability of statistical signal processing in the neuroscienfic literature by explaining a celebrated phenomenon called the Weber–Fechner law as a Bayesian optimization under communication limitations in neural channels. We study two models based on the analyses developed earlier in the thesis to understand the

optimality of human perception and provide falsifiable hypotheses for a wide range of stimuli. Testing the theory on real-world auditory data provided some evidence that our model is consistent with human perception.

Chapter 8 – Conclusion

We conclude with a survey of the main ideas in the thesis and a prognosis of the future directions of DFSQ theory. In particular, the notion of chatting has promise in influencing novel multimodal sensing methodologies and network architectures.

1.3 Bibliographical Notes

Parts of Chapter 3 appear in the paper:

- J. Z. Sun, V. Misra and V. K. Goyal, "Distributed Functional Scalar Quantization Simplified," accepted to *IEEE Transactions on Signal Processing*, 2013.

Parts of Chapter 4 appear in the paper:

- J. Z. Sun and V. K. Goyal, "Rate Loss in Distributed Functional Source Coding," accepted to *IEEE International Symposium on Information Theory*, 2013.

Parts of Chapter 5 appear in the papers:

- J. Z. Sun and V. K. Goyal, "Chatting in Distributed Quantization Networks," in *Proceedings of the Fiftieth Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2012.
- J. Z. Sun and V. K. Goyal, "Intersensor Communication in Distributed Quantization Networks," in review, 2013.

Parts of Chapter 6 appear in the paper:

- J. Z. Sun and V. K. Goyal, Scalar Quantization for Relative Error, in *Proceedings of the IEEE Data Compression Conference*, Mar. 2011. (Capocelli Prize for Best Student Paper)

Parts of Chapter 7 appear in the papers:

- J. Z. Sun, G. I. Wang, V. K. Goyal and L. R. Varshney, "A Framework for Bayesian Optimality of Psychophysical Laws," *Journal of Mathematical Psy-*

chology, vol. 56, no. 6, pp. 495–501, Dec. 2012.

- L. R. Varshney and J. Z. Sun, "Why do We Perceive Logarithmically?" *Significance*, vol. 10, no. 1, pp. 28–31, Feb. 2013.

Chapter 2

Background

The main aim of the thesis is to study efficient techniques for lossy compression in a network. Abstractly, lossy compression is the design of an output (reproduction) space that represents an input (source) space with good fidelity under resource limitations. When the limitations include transmission or storage, the source may be transformed into a binary representation; this is commonly referred to as *source coding*. When the input space is continuous-valued, a fundamental element of a lossy compression scheme is a *quantizer*.

Quantization has two complementary asymptotic theories [127]. The more well-known theory addresses performance when long blocks of data are compressed together, with length increasing without bound while holding the per-entry resource allocation fixed. This yields the rate–distortion theory initiated by Shannon [162, 164] and developed comprehensively by Gallager [47] and Berger [11]. Increasing block length yields space-filling gains and other advantages [105], so rate–distortion theory provides ultimate bounds which are not achievable with finite blocks. The "other" asymptotic theory, pioneered by Bennett [10], makes approximations that introduce errors that become negligible as the the size of the output space increases without bound while holding the block length fixed. The goal of this high-resolution theory is to allow quantizer analysis and design to be done with elementary calculus. This theory is especially useful in the design of scalar quantizers. Rates as low as 4 bits per sample are generally sufficient for high-resolution theory to yield good

designs [59, 127], and it has been argued that high-resolution theory has been more impactful than rate–distortion theory on practical source coding.

In this chapter, we formalize the model for lossy compression and introduce the notation that will be used in the thesis. We also summarize existing work in quantization theory and information theory, focusing on results that pertain to networks or computations. Except for specific situations that will be highlighted or when the context is clear, capital letters are used to indicate scalar random variables, e.g. X. We denote an N-dimensional vector as x_1^N and use x_i to denote the ith scalar entry of the vector. When talking about a function, e.g. $f(x)$, we drop the arguments for convenience unless necessary in the discussion. We employ the notation $\|f\|_p$ for $p \geq 1$ to denote the $\mathcal{L}_p$ norm $(\int_{-\infty}^{\infty} |f(x)|^p \, dx)^{1/p}$. A set of consecutive integers $\{a, \ldots, b\}$ is denoted $[a : b]$. Finally, we express asymptotic growth rates with $\mathcal{O}$-notation, as commonly used in mathematics and engineering [23].

2.1 Quantization

We consider a type of quantization that is a deterministic mapping from the real space to a finite set of known points called the *codebook* $\mathcal{C}$. For example, we can compress a vector of continuous-valued observations or signal samples X_1^L to a set of K points $\mathcal{C} \triangleq \{c_k\}_{k=1}^K \subset \mathbb{R}^L$. This mapping, denoted $Q_K^{(L)}$ satisfies $Q_K^{(L)}(x) = c_k$ if $x \in P_k$, where the cells $\{P_k\}_{k=1}^L$ form a *partition* $\mathcal{P}$ of $\mathbb{R}^L$. The case with $L = 1$ is known as *scalar* quantization; otherwise it is called *vector* quantization. In this thesis, we will focus mainly on the design of scalar quantizers and drop the notation L when appropriate.

A scalar quantizer is called *regular* if the partition cells are intervals containing the corresponding codewords; the generalization to vector quantization is that partition cells containing the corresponding codewords are convex. In general, this does not need to be the case and *irregular* quantizers can oftentimes yield improved performance in a network setting at the cost of being difficult to analyze and design. We will mostly consider regular scalar quantizers and assume the codebook entries

are indexed from smallest to largest. Denoting $P_k = (p_{k-1}, p_k]$ for each k implies $p_0 < c_1 \leq p_1 < c_2 \leq \cdots < c_K \leq p_K$, with $p_0 = -\infty$ and $p_K = \infty$. We define the *granular* region as $(p_1, p_{K-1}]$ and its complement $(-\infty, p_1] \cup (p_{K-1}, \infty)$ as the *overload* region.

Performance analysis of quantizers is generally studied in a Bayesian framework, where the L-dimensional input source follows a probabilistic model that is assumed to be memoryless in time,[1] i.e. when a joint density $f_{X_1^L}$ exists and is known. We define a distortion function as $d : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$, corresponding to the goodness of a reconstruction in the space of $\mathcal{Y}$ of a signal in the space of $\mathcal{X}$. The goal of compression is to have a low expected per-scalar distortion $D = L^{-1} \operatorname{E}[d(X_1^L, Q_K^{(L)}(X_1^L))]$. Uniform or linear quantization, where partition cells in the granular region have equal size and shape, is commonly used in practice and has interesting asymptotic properties [6, 10, 61, 72, 131]. However, nonuniform quantization can provide significant improvement in distortion. There is a wide body of literature that describes how to design the codebooks and partitions; comprehensive surveys of these works have been compiled [54, 60, 66]. For a given K, exact optimality in quantization is difficult to achieve but iterative methods that exploit nearest-neighbor encoding and centroid reconstruction conditions, i.e. Lloyd–Max algorithms, are the best known tools for quantizer design [21, 98, 104, 117]. More recently, generalized approximate message passing can be used to design optimal irregular quantizers [82].

In this thesis, we consider a class of scalar quantizers that are easy to specify and have desirable asymptotic properties (also see [52]). As illustrated in Figure 2-1, this class uses the companding method to generate nonuniform quantizers from a uniform one, where the scalar source is transformed using a nondecreasing and smooth *compressor* function $c : \mathbb{R} \to (0, 1)$, then quantized using a uniform quantizer with K equidistant levels on $(0, 1)$, and finally passed through the *expander* function c^{-1}. Compressor functions are defined such that $\lim_{x \to -\infty} c(x) = 0$ and $\lim_{x \to \infty} c(x) = 1$. It is then convenient to define a *point density function* as $\lambda(x) = c'(x)$. Because of

[1] The memoryless assumption is implicit for quantizers that do not have memory since correlation across time cannot be exploited. Later, we will discuss briefly how to design vector quantizers to account for memory.

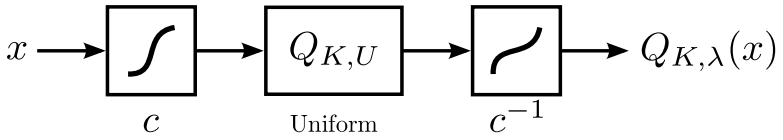


Figure 2-1: A block diagram for companding as a constructive method for nonuniform scalar quantization. Companders are defined by the codebook size K and the point density function λ, which is the derivative of the compressor function c. The notation $Q_{K,U}$ is used to describe the canonical uniform quantizer with K partitions in $(0, 1)$.

the limiting conditions on c, there is a one-to-one correspondence between λ and c, and hence a quantizer of the form shown in Figure 2-1 can be uniquely specified using a point density function and codebook size. We denote such a quantizer $Q_{K,\lambda}$. By virtue of this definition, the integral of the point density function over any quantizer cell is $1/K$:

$$\int_{p_k}^{p_{k+1}} \lambda(x)\, dx = \frac{1}{K}, \qquad k \in [1 : K]. \tag{2.1}$$

In practice, scalar quantization is rarely performed by an explicit companding operation. A slight modification that avoids repeated computation of c^{-1} derives partition boundaries from the compressor function by applying the compressor c and comparing to threshold values (multiples of $1/K$) to determine the partition cell P_k, but then obtains c_k from a precomputed table. This modification allows the choice of $\mathcal{C}$ to be an additional degree of freedom in the quantizer, and different codeword designs will have different statistical properties. For example, the *centroid* reconstruction

$$c_k = \mathrm{E}[X | X \in P_k] \tag{2.2}$$

yields the best mean squared error (MSE) and decorrelates the output random variable and the quantization error, i.e. $\mathrm{E}[Q_{K,\lambda}(X)(X - Q_{K,\lambda}(X))] = 0$. A much simpler choice for non-extremal codewords,

$$c_k = \frac{p_{k-1} + p_k}{2}, \qquad k \in [2 : K - 1], \tag{2.3}$$

is known as *midpoint* reconstruction.[2] Here, the extremal reconstruction values are

[2]Note that midpoint reconstruction is the centroid for a uniform source

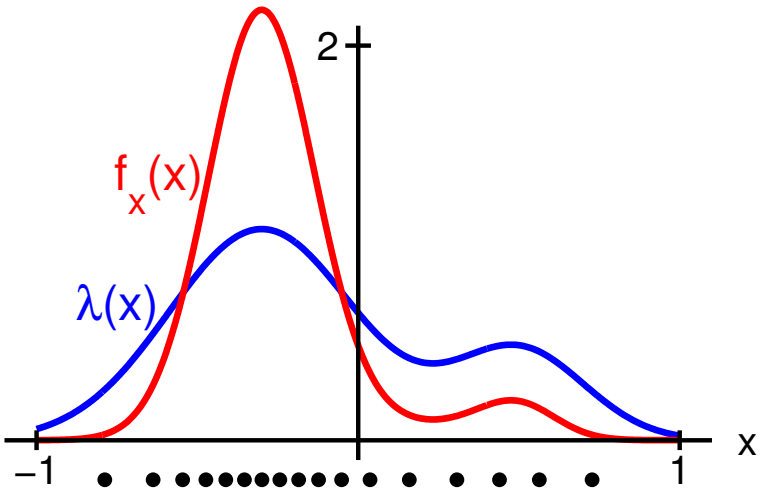


Figure 2-2: A sample source density f_X and point density function λ, with corresponding codebook designed via (2.3). This choice of λ follows (2.7), which turns out to be the optimal design of a quantizer without additional coding.

fixed to be $c_1 = p_1$ and $c_K = p_{K-1}$. Midpoint reconstruction is suboptimal in MSE relative to centroid reconstruction, but it has the simplicity of depending only on λ and K—not on the source density. Particular quantizer-source pairs may also have the nice property of the input and quantization error becoming uncorrelated as K becomes large [113]. A simple model employed in the signal processing community that accounts for the decaying correlation is the additive noise model, where the quantization error is modeled as uniform noise (see, e.g. [137, 192]). However, this is generally not true without the use of subtractive dithering [3, 67].

2.2 High-Resolution Scalar Quantization

It is difficult to express the performance of a quantizer for a particular choice of K. Fortunately, the distortion can be well-approximated by a simple expression as K becomes large in the companding model, which can then be used to optimize λ. This asymptotic decoupling of K and λ forms the basis of high-resolution quantization.

We first present results for the ubiquitous MSE distortion on a memoryless scalar

source X, which is defined as

$$D_{\text{mse}}(K, \lambda) = \mathrm{E}[|X - Q_{K,\lambda}(X)|^2], \tag{2.4}$$

where the expectation is with respect to the source density f_X. Under the additional assumption that f_X is continuous (or simply measurable) with tails that decay sufficiently fast,

$$D_{\text{mse}}(K, \lambda) \simeq \frac{1}{12K^2} \mathrm{E}[\lambda^{-2}(X)], \tag{2.5}$$

where $\simeq$ indicates that the ratio of the two expressions approaches one as K increases [10, 141]. More rigorously, the $\simeq$ notation says

$$\lim_{K\to\infty} K^2 D_{\text{mse}}(K, \lambda) = \frac{1}{12} \mathrm{E}[\lambda^{-2}(X)].$$

The key insight is that the MSE of a scalar quantizer can be approximated by a simple relationship between the source distribution, point density and codebook size, and this relation becomes more precise with increasing K. Moreover, quantizers designed according to the companding model are *asymptotically optimal*, meaning that the quantizer optimized over λ has distortion that approaches the performance of the best Q_K found by any means [17, 19, 99]:

$$\inf_{Q_K} \mathrm{E}\left[|X - Q_K(X)|^2\right] \simeq \inf_{\lambda} \frac{1}{12K^2} \mathrm{E}[\lambda^{-2}(X)].$$

Experimentally, the distortion resulting from using companding quantizers specified by small K can be predicted by the asymptotic analysis, as demonstrated later in the thesis.

Although the previous discussion was on MSE, similar performance results have been rigorized for difference distortion measures, a class where $d(x, y) = \rho(x - y)$ under certain mild conditions on ρ [56, 198]. For example, the rth-power absolute error (RAE) distortion measure

$$D_{\text{rae}} = \mathrm{E}[|x - Q(x)|^r]$$

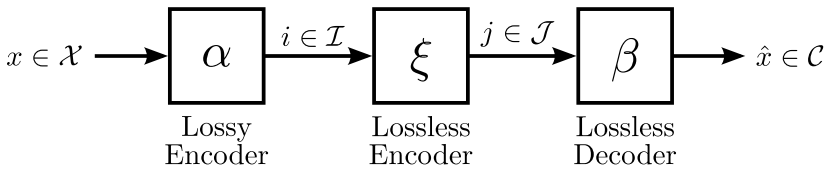


Figure 2-3: A block diagram for source coding as a means for communicating lossy compression representations. The scalar case is shown, but the generalization to vector sources is natural. The lossy encoder is a quantization mapping with output corresponding to the index of the codeword output. The lossless encoder then maps the index to a string of bits that may not have the same length. The source coding model allows for comparison between the distortion and communication rate in different schemes.

has an attractive optimal asymptotic distortion form satisfying

$$D_{\text{rae}}(K, \lambda) \simeq \frac{1}{(1+r)2^r K^r} \, \mathrm{E}[\lambda^{-r}(X)],$$

extending the MSE measure nicely. Other distortion measures are discussed later in the chapter.

Since the dependence on K and λ is separated in the limit, calculus techniques can be used to optimize companders using the distortion equation [104, 141]. However, when the quantized values are to be communicated or stored, it is natural to map each codeword to a string of bits and consider the trade-off between distortion and communication rate R, defined to be the expected number of bits per sample. More formally, source coding employs the structure in Figure 2-3, comprising a codebook $\mathcal{C}$ with corresponding index set $\mathcal{I}$, partition $\mathcal{P}$, binary codebook $\mathcal{J}$, lossy encoder $\alpha : \mathbb{R}^L \to \mathcal{I}$, lossless encoder $\xi : \mathcal{I} \to \mathcal{J}$, and lossless decoder $\beta : \mathcal{J} \to \mathcal{C}$. In scalar quantization ($L = 1$), there are two methods in designing $\mathcal{J}$ that are of interest: fixed-rate and entropy-constrained quantization. We will discuss both in detail below.

2.2.1 Fixed-Rate Quantization

In the simpler case, $\mathcal{J}$ is simply the binary expansion of the index set and the communication rate is $R = \lceil \log_2(K) \rceil$, corresponding to the length of the binary representation of the maximum index. We call this *fixed-rate* or *codebook-constrained* quanti-

zation. Assuming a K where the ceiling operation can be ignored, the distortion-rate trade-off can be written as

$$D_{\text{mse,fr}}(R, \lambda) \simeq \frac{1}{12} \, \mathrm{E}\left[\lambda^{-2}(X)\right] 2^{-2R}. \tag{2.6}$$

The best choice of λ has been solved in several ways, but the most elegant proof uses Hölder's inequality [63], which says, for measurable functions v and w and constants p and q satisfying $1 \leq p, q \leq \infty$ and $1 = 1/p + 1/q$,

$$\|vw\|_1 \leq \|v\|_p \, \|w\|_q,$$

where equality holds if v and w are linearly dependent. Setting

$$v(x) = (f_X(x)\lambda^{-2}(x))^{1/3} \qquad \text{and} \qquad w(x) = \lambda^{2/3}(x),$$

we have the following:

$$\begin{aligned}
\mathrm{E}[\lambda^{-2}(X)] &= \int f_X(x)\lambda^{-2}(x)\, dx \\
&= \left(\int f_X(x)\lambda^{-2}(x)\, dx\right)\left(\int \lambda(x)\, dx\right)^2 \\
&\geq \left(\int \left(f_X(x)\lambda^{-2}(x)\right)^{1/3} \lambda^{2/3}(x)\, dx\right)^3 \\
&= \left(\int f_X^{1/3}(x)\, dx\right)^3.
\end{aligned}$$

Using the equality condition, the lower bound is met with point density

$$\lambda^*_{\text{mse,fr}}(x) \propto f_X^{1/3}(x), \tag{2.7}$$

and the resulting optimal distortion satisfies

$$D^*_{\text{mse,fr}}(R) \simeq \frac{1}{12} \, \|f_X\|_{1/3} \, 2^{-2R}. \tag{2.8}$$

Recall that the point density must integrate to one, which provides the scaling term in (2.7). An example for point density optimization is shown in Figure 2-2.

2.2.2 Entropy-Constrained Quantization

Exploiting the probabilistic nature of the quantizer output, the codeword indices can be coded to produce bit strings of different lengths based on probabilities of occurrence. This is referred to as *variable-rate* quantization and is clearly more general than fixed-rate quantization. If the decoding latency is allowed to be large, one can employ block entropy coding and the communication rate approaches $H(Q_{K,\lambda}(X))$, which has the following asymptotic form:

$$H(Q_{K,\lambda}(X)) \simeq h(X) + \log K + \mathrm{E}[\log \lambda(X)],$$

where the proof is given in [120, Section II-C].[3] This asymptotic limit, called *entropy-constrained* quantization, yields a distortion–rate trade-off with λ:

$$D_{\mathrm{mse,ec}}(R, \lambda) \simeq \frac{1}{12} \mathrm{E}\left[\lambda^{-2}(X)\right] 2^{2h(X)} 2^{2\,\mathrm{E}[\log \lambda(X)]} 2^{-2R}. \tag{2.9}$$

The constant term dependent on λ can be bounded below using Jensen's inequality [63]:

$$\begin{aligned} \mathrm{E}\left[\lambda^{-2}(X)\right] 2^{2\,\mathrm{E}[\log \lambda(X)]} &= 2^{\log \mathrm{E}\left[\lambda^{-2}(X)\right]} 2^{2\,\mathrm{E}[\log \lambda(X)]} \\ &\geq 2^{\mathrm{E}\left[\log \lambda^{-2}(X)\right]} 2^{2\,\mathrm{E}[\log \lambda(X)]} \\ &= 1, \end{aligned}$$

with equality if $\lambda(X)$ is a constant. Hence, the resulting optimal quantizer has a constant point density $\lambda^*_{\mathrm{mse,ec}}(x)$ on the support of the source density, which remarkably

[3] Like in most works on source coding, the log function is base 2. When considering the differential entropy, it is customary to express the base as e, which we indicate using ln. We restrict to expressing performance in bits rather than nats and will make the proper conversions when necessary.

corresponds to uniform quantization. The optimal asymptotic distortion satisfies

$$D^*_{\text{mse,ec}}(R) \simeq \frac{1}{12} 2^{-2(R-h(X))}. \tag{2.10}$$

Note that block entropy coding suggests that the sources are transmitted in blocks even though the quantization is scalar. Hence, (2.10) is an asymptotic result and serves as a lower bound on practical entropy coders with finite block lengths that match the complexity restrictions of a system.

In general, the optimal entropy-constrained quantizer (at a finite rate) for a distribution with unbounded support can have an infinite number of codewords [69]. The compander model used in this thesis cannot generate all such quantizers. A common alternative is to allow the codomain of c to be $\mathbb{R}$ rather than $(0, 1)$, resulting in a point density that cannot be normalized [56, 66]. To avoid parallel developments for normalized and unnormalized point densities, we restrict our attention to quantizers that have a finite number of codewords K at any finite rate R. This may preclude exact optimality, but it does not change the asymptotic behavior as K and R increase without bound.[4] Specifically, the contribution to overall distortion from the overload region is made negligible as K and R increase, so the distinction between having finitely- or infinitely-many codewords becomes unimportant.

2.3 Quantization for Computation

The above discussion briefly mentions alternatives to MSE. Extending the classes of distortion measures is particularly pertinent in media compression, where nondifference distortion measures can be more perceptually meaningful [78, 129]. Many works on the design of efficient quantizers use high resolution; we discuss these results in more detail in Section 2.4. Similar needs for larger classes of distortion measures exist in the detection and estimation literature, where high-resolution has been used to de-

[4] The restriction of finite codewords places the practical quantizers designed using this thesis in the class of variable-rate rather than true entropy-constrained quantizers. We will label the quantizers as entropy-constrained to preserve the meaning of the asymptotic nature of the result.

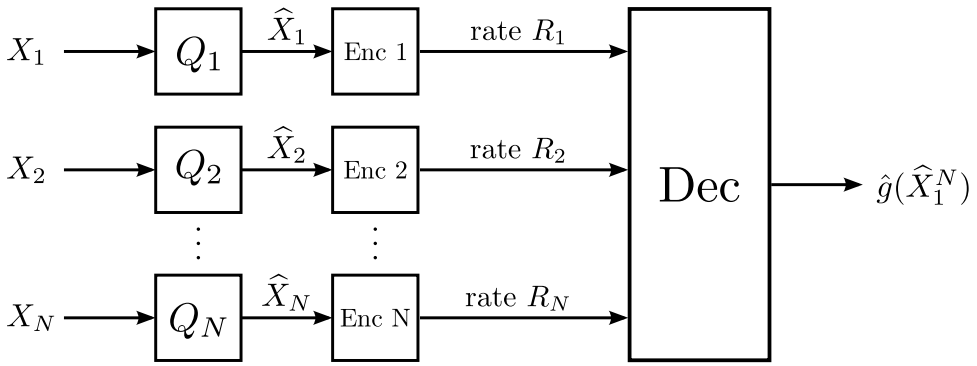


Figure 2-4: A distributed computation network, where each of N spatially-separated sources generate a scalar X_n. The scalars are encoded and communicated over rate-limited links to a central decoder, where Encoder n is allowed transmission rate R_n. The decoder computes an estimate of the function $g(X_1^n) = g(X_1, \ldots, X_n)$ from the received data using $\hat{g}(\widehat{X}_1^n)$.

termine parameters of interest from noisy data [9,68,147]. More recently, companding quantizers have been used to solve distributed estimation problems [111].

This thesis is predominately interested in exploring lossy compression in a network setting. In this application, it is customary to think about sensors or agents that collect local information and try to influence a fusion center or central decision-maker. Generally, the fusion center cares about a function of the sensors' observations that may be nonlinear. Although this problem can be framed as multidimensional companding using a nondifference distortion measure, it is illuminating as its own framework. In a paper that provides the foundation for the thesis, Misra, Goyal and Varshney devised the *distributed functional scalar quantization* (DFSQ) model [120], which will be described in this section.

In a DFSQ model, sensors employ scalar quantization to acquire and compress information, and communicate directly to the fusion center, which performs a reconstruction using $\hat{g}$ on the quantized data to approximate a desired computation g, Here, nonuniform quantizers optimized for g often perform much better than uniform quantization or nonuniform quantizers that do not account for g. In [120], distortion performance and quantizer design are discussed for the distributed setting shown in Figure 2-4, where g is a scalar-valued function. For DFSQ, the cost of interest is

functional MSE (fMSE):

$$D_{\text{fmse}}(K_1^N, \lambda_1^N) = \mathrm{E}\left[\left|g(X_1^N) - \hat{g}(Q_{K_1^N,\lambda_1^N}(X_1^N))\right|^2\right], \tag{2.11}$$

where $\hat{g}$ is chosen to be the joint centroid (JC) reconstruction

$$\hat{g}_{\text{jc}}(x_1^N) = \mathrm{E}\left[g(X_1^N) \,\middle|\, Q_{K_1^N,\lambda_1^N}(X_1^N) = Q_{K_1^N,\lambda_1^N}(x_1^N)\right], \tag{2.12}$$

and $Q_{K_1^N,\lambda_1^N}$ is scalar quantization performed on a vector such that

$$Q_{K_1^N,\lambda_1^N}(x_1^N) = \left(Q_{\lambda_1,K_1}(x_1), \ldots, Q_{\lambda_N,K_N}(x_N)\right).$$

Note the complexity of computing $\hat{g}_{\text{jc}}$: it requires integrating over an N-dimensional partition cell with knowledge of the joint source density $f_{X_1^N}$.

Before understanding how a quantizer affects fMSE, it is convenient to define how a computation affects distortion locally at each sensor:

Definition 2.1. The *univariate functional sensitivity profile* of a function g is defined as

$$\gamma(x) = |g'(x)|.$$

The *nth functional sensitivity profile* of a multivariate function g is defined as

$$\gamma_n(x) = \left(\mathrm{E}\left[|g_n(X_1^N)|^2 \,\middle|\, X_n = x\right]\right)^{1/2}, \tag{2.13}$$

where $g_n(x)$ is the partial derivative of g with respect to its nth argument evaluated at the point x.

Given the functional sensitivity profile, the main result of [120] says

$$D_{\text{fmse}}(K_1^N, \lambda_1^N) \simeq \sum_{n=1}^{N} \frac{1}{12K_n^2}\, \mathrm{E}\left[\left(\frac{\gamma_n(X_n)}{\lambda_n(X_n)}\right)^2\right], \tag{2.14}$$

provided the following conditions are satisfied:

MF1. The function g is Lipschitz continuous and twice differentiable in every argument except possibly on a set of Jordan measure zero.

MF2. The source pdf $f_{X_1^N}$ is continuous, bounded, and supported on $[0,1]^N$.

MF3. The function g and point densities λ_n allow $\mathrm{E}[(\gamma_n(X_n)/\lambda_n(X_n))^2]$ to be defined and finite for all n.

Following the same recipes to optimize over λ_1^N as for MSE in Section 2.2, the relationship between distortion and communication rate is found. In both the fixed-rate and entropy-constrained settings, the functional sensitivity profile acts to shift quantization points to where they can reduce the distortion in the computation. For fixed rate, the minimum high-resolution distortion is asymptotically achieved by

$$\lambda^*_{n,\mathrm{fmse,fr}}(x) \propto \left(\gamma_n(x) f_{X_n}(x)\right)^{1/3}, \tag{2.15}$$

where f_{X_n} is the marginal distribution of X_n. This leads to a distortion that satisfies

$$D^*_{\mathrm{fmse,fr}}(R_1^N) \simeq \sum_{n=1}^{N} \frac{1}{12} \|\gamma_n^2 f_{X_n}\|_{1/3} 2^{-2R_n}. \tag{2.16}$$

If rate allocation (see Section 2.4) can be employed and we define $R_{\mathrm{tot}} = \sum_n R_n$, then

$$D^*_{\mathrm{fmse,fr}}(R_{\mathrm{tot}}) \simeq \frac{N}{12} \left(\prod_{n=1}^{N} \|\gamma_n^2 f_{X_n}\|_{1/3} \right)^{1/N} 2^{-2R_{\mathrm{tot}}/N}. \tag{2.17}$$

In the entropy-constrained setting, the optimizing point density is asymptotically

$$\lambda^*_{n,\mathrm{fmse,ec}}(x) \propto \gamma_n(x), \tag{2.18}$$

leading to fMSE that satisfies

$$D^*_{\mathrm{fmse,ec}}(R_1^N) \simeq \sum_{n=1}^{N} \frac{1}{12} 2^{2h(X_n) + 2\,\mathrm{E}[\log \gamma_n(X_n)]} 2^{-2R_n}. \tag{2.19}$$

Similarly to fixed-rate, using rate allocation provides a cleaner form:

$$D^*_{\text{fmse,ec}}(R_{\text{tot}}) \simeq \frac{N}{12}\left(\prod_{n=1}^{N} 2^{2h(X_n)+2\,\mathrm{E}[\log \gamma_n(X_n)]}\right)^{1/N} 2^{-2R_{\text{tot}}/N}. \tag{2.20}$$

Notice unnormalized point densities are not required here since the sources are assumed to have bounded support.

In the distributed setting where sources are correlated, there is another type of coding that can yield a better distortion–rate trade-off than the methods we have considered. If coding length is allowed to be long, then the correlation at the quantizer outputs can be exploited via Slepian–Wolf coding (see Section 2.5). It is customary to consider the sum-rate constraint in this setting, which yields an optimal trade-off that asymptotically satisfies

$$D^*_{\text{fmse,sw}}(R_{\text{tot}}) \simeq \frac{N}{12}\left(2^{2h(X_1^N)}\prod_{n=1}^{N} 2^{2\,\mathrm{E}[\log \gamma_n(X_n)]}\right)^{1/N} 2^{-2R_{\text{tot}}/N}. \tag{2.21}$$

The main usefulness of DFSQ is asymptotically precise characterization of the performance of scalar quantizers in distributed networks for varying coding complexities. In addition, DFSQ theory yields tractable quantizer design using the companding model for a wide range of source joint densities and computations. Despite the similar setups, the necessary conditions in [120], [102] and [111] are not equivalent, and cumulatively provide a compelling story on using high-resolution analysis to studying lossy compression in distributed settings.

Beyond the foundational analysis on asymptotic distortion and quantizer design, [120] also provide insight on how irregular quantizers can fit into this model. It turns out under reasonable conditions, optimal quantizers for DFSQ tend to be regular. Additionally, they introduce concepts such as don't-care intervals and chatting, which will be described in greater detail in Section 5.2.

2.4 Vector Quantization

As a bridge between high-resolution scalar quantization and rate–distortion theories, we summarize some key results in vector quantization. Since the focus of this thesis is two extremal cases, i.e. fixed L with increasing K and increasing L and fixed K/L, our coverage will be brief. We reemphasize that excellent and comprehensive reviews of vector quantization are readily available [54, 60, 66].

The pioneer of high-resolution vector quantization is undoubtedly Zador, who argued that the best L-dimensional fixed-rate quantizer has distortion–rate performance governed by

$$D_{\text{fr}}^{(L)}(R) \simeq b_{\text{fr},L} \|f\|_{L/(L+2)} 2^{-2R},$$

and the best L-dimensional entropy-constrained quantizer has distortion–rate performance satisfying

$$D_{\text{ec}}^{(L)}(R) \simeq b_{\text{vr},L} 2^{2h_L(X)} 2^{-2R},$$

where h_L is the dimension-normalized differential entropy of the source and $b_{\text{fr},L}$ and $b_{\text{ec},L}$ are constants that depend on the source density [204,205]. These arguments were later formalized in the fixed-rate [17] and entropy-constrained settings [65]. Work by Gersho on block quantization popularized the use of lattice vector quantization as a way to realize optimality using high-resolution analysis [53].

Although many structures have been proposed, the two dominant methods for vector quantizers are predictive coding and transform coding, both of which exploit redundancy in real-world signals with linear processing followed by scalar quantization [66, Section II-B]. Predictive coding exploits the auto-regressive nature of many speech and image signals using sliding-block codes and have influenced many modern standards [54,79]. Meanwhile, transform coding uses orthogonal transforms to decorrelate a block of scalars [58,71], and is a staple in modern compression algorithms [58]. There are preliminary investigations into DFSQ theory applied to transform coding, but the results are not rigorous [119].

The model of interest in the work is a distributed system comprising N com-

panding scalar quantizers, which is a strict subclass of N-dimensional companding vector quantizers. Although the concept of point densities extend fully to vector quantizers, companding does not and there are several works that discuss design of compressor function in the multidimensional setting [123, 167]. Beyond companding models, network quantization has been studied using iterative descent methods for a variety of network topologies [43, 153, 154]. Other relevant and current work on network quantization include distributed transform coding to exploit correlations and network topologies [51, 88, 165] and achievable systems that have good scaling properties through reuse of quantization indices [109].

The fMSE distortion measure which dominates discussion of this thesis is strongly related to a class of nondifference distortion measures called *locally quadratic*, which are predominantly studied in context of predictive coders or lattice quantizers. The earliest work in this field were for speech coding and were known as Itakura–Saito [74] or input-weighted quadratic distortion measures. These investigations were on iterative methods exploiting the necessary nearest-neighbor and centroid conditions, and provide algorithmic results on construction of optimal quantizers [18, 64, 98]. Later, locally quadratic distortion measures were studied using high-resolution theory both in fixed-rate [50, 97] and entropy-constrained [102] settings for vector quantizers.

Finally, we conclude this section with some discussion on rate or bit allocation, which was briefly mentioned in Section 2.3. Rate allocation is the problem of optimizing the total resource constraint over a block of observations, e.g. transform coefficients in transform coding or sensors in a distributed network. When the system allows for this flexibility across measurements, rate allocation optimizes coding over heterogeneous sources and lead to better performance. Solutions to rate allocation employ Lagrangian optimization or integer programming on high-resolution approximations to distortion, as in (2.5) [54, 71, 161]; we delay formalizing the mathematical framework until Chapter 5.

2.5 Rate–Distortion Theory

While our main aims are to understand companding quantizers for fMSE, it is useful to study how the compression performance compares to the fundamental limits when complexity and block length are not constrained. We briefly summarize some key results in rate–distortion theory that are important in the thesis. Several classic texts cover rate–distortion theory with varying degrees of rigor [11, 27, 28, 34, 47].

We begin by carefully defining the rate–distortion function for a memoryless source $X \in \mathcal{X}$ given a distortion measure d, where we recall the average per-letter distortion between two vectors x_1^L and $\hat{x}_1^L$ is

$$d(x_1^L, \hat{x}_1^L) = \frac{1}{L} \sum_{i=1}^{L} d(x_i, \hat{x}_i).$$

Then, we can define a lossy source code as follows:

Definition 2.2. A $(2^{LR}, L)$ *lossy source code* consists of:

- a codebook $\mathcal{C} = \{\hat{x}_1^L(1), \hat{x}_1^L(2), \ldots, \hat{x}_1^L(2^{\lfloor LR \rfloor})\}$;
- an encoder that assigns to each sequence $x_1^L \in \mathcal{X}^L$ an index $m(x_1^L) \in [1 : 2^{\lfloor LR \rfloor}]$;
- a decoder that assigns to each index $m \in [1 : 2^{\lfloor LR \rfloor}]$ an reproduction codeword $\hat{x}_1^L \in \mathcal{C}$.

Definition 2.3. The distortion associated with the code is defined as $\mathrm{E}[d(X_1^L, \widehat{X}_1^L)]$. A rate–distortion pair (R, D) is *achievable* if there exists a sequence of $(2^{LR}, L)$ codes with

$$\limsup_{L \to \infty} \mathrm{E}[d(X_1^L, \widehat{X}_1^L)] \leq D.$$

Definition 2.4. The *Shannon rate–distortion function* $r(D)$ is the infimum of rates R such that (R, D) is achievable. Similarly, the *Shannon distortion–rate function* $\delta(R)$ is the infimum of distortions D such that (R, D) is achievable.[5]

[5] In information theory texts, the rate–distortion and distortion–rate functions are simply denoted $R(D)$ and $D(R)$, and the use of $r(D)$ and $\delta(R)$ have been used to indicate the operational rate–distortion and operational distortion-rate functions [66]. Because this thesis is dominated by performance results for achievable companding quantizers, we try to prevent overloading the meaning of D and R, which led to this choice of notation.

Using an achievability argument based of joint typicality encoding and converse employing Fano's inequality, the Shannon rate–distortion function for a memoryless source and a distortion measure d (satisfying $d(x, \hat{x}) = 0$ only if $x = \hat{x}$) is

$$r(D) = \min_{p(\hat{x}|x)\,:\,\mathrm{E}[d(X_1^L, \hat{X}_1^L)] \le D} I(X; \widehat{X}), \tag{2.22}$$

for $D \geq D_{\min} = \min_{\hat{x}(x)} \mathrm{E}[d(X_1^L, \widehat{X}_1^L)]$, where I is the mutual information between the source and its reproduction. For the Hamming error measure, we can recover the optimal coding rate for lossless coding of a discrete memoryless source using the above result; it is simply $r(0) = H(X)$, where H is Shannon entropy. In the finite-alphabet case, $r(D)$ is nonincreasing, convex and continuous in D and the Shannon distortion–rate function $\delta(R)$ is easily determined.

2.5.1 Multiterminal Shannon Theory

Distributed or multiterminal source coding has received considerable attention over the past few decades due to the emergence of networks and because of the mathematical beauty of the results. Because it is studied under the same topology as DFSQ (Figure 2-4), multiterminal source coding provides a benchmark for comparison with this thesis.

We begin by summarizing the Slepian–Wolf theorem, which says the achievable rate region $\mathcal{R}^*$ for distributed lossless source coding of a 2-DMS (X_1, X_2) (Figure 2-5) is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &> H(X_1|X_2), \\ R_2 &> H(X_2|X_1), \\ R_1 + R_2 &> H(X_1, X_2). \end{aligned} \tag{2.23}$$

Achievability proofs typically employ random binning followed by joint-typicality decoding and can easily be extended to N sources [26]. What is remarkable is that, in this setting, distributed coding matches the outer bound for joint encoding, mean-

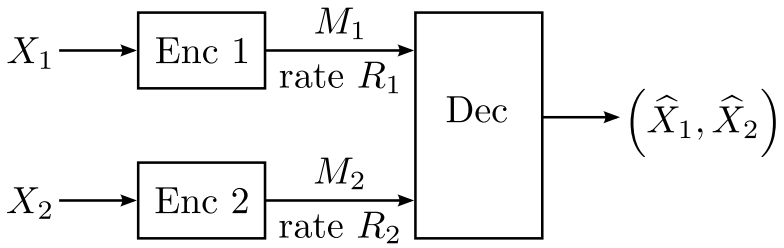


Figure 2-5: The multiterminal source coding problem, where each of several correlated memoryless sources are coded separately into a message M_i with R_i. The decoder then determines the reconstructions by jointly decoding all messages. The figure shown only has two sources, corresponding to the 2-DMS Slepian–Wolf problem. However, the multiterminal source coding problem generalizes to arbitrary number of sources.

ing communication between encoders cannot improve the asymptotic communication rate. However, the ability for encoders to communicate can improve the error exponents associated with the large-blocklength analysis [28].

Practical Slepian-Wolf (SW) coding usually involves designing codes that approach one of the two kinks in the curve. By symmetry, the other kink can also be approached and any point on the line between them is found by time sharing. The kink corresponds to $R_1 + R_2 = H(X_2) + H(X_1|X_2) = H(X_1, X_2)$, which is the problem of compressing X_1 with side information X_2 at the decoder. This is is mathematically equivalent to a "correlation channel" in the channel coding literature. Hence, linear channel codes such as coset codes and syndromes can be used [194, 208]. This approach is often referred to as *asymmetric coding*, as compared to *symmetric coding*, which approaches the line between the two kinks directly. More recently SW coding has been reinvigorated by the application of capacity-achieving channel-coding techniques [150, 197]. Moreover, there has been a flurry of results for universal SW coding [20, 33, 49, 116, 136], many based on the methods in [28]. These new schemes show promise, but their applicability in real technologies are limited by the need for long block lengths and high computational resources at the decoder.

Assume now that the decoder wishes to reconstruct approximations of the sources with bounded error with respect to a distortion measure, corresponding to multiterminal lossy source coding. The rate region is not solved, and the most general results are the Berger–Tung inner and outer bounds [182]. In more restrictive situations, the

rate region is known:

- Under a difference distortion measure when the desired distortion is small [207].
- In the quadratic Gaussian CEO problem, where each node observes noisy versions of the Gaussian source and distortion measure is MSE [134, 135, 149, 188] (the more general CEO problem was introduced in [14]). A similar model was proposed earlier in [62].
- For Gaussian sources and MSE distortion [190] (also see [133]).
- Under the logarithmic loss distortion measure [25].

When the sources are continuous-valued, lossy source coding is usually modeled as quantization followed by lossless source coding, matching the model used in [43]; we shall adopt this model in our proposal as well. Generally speaking, few have studied the effect of side information at the encoder since it does not reduce the Shannon distortion–rate. However, there have been investigations on coded side information at the decoder. In the lossless setting, this problem has been studied in [2] and more recently in [112].

Another influential lossy source coding framework is Wyner–Ziv coding, in when only one source is to be reconstructed while the other serves as side information [195, 196]. The key result, the Wyner–Ziv theorem, says the rate–distortion function of X with side information Y available noncausally at the decoder is

$$r_{\mathrm{WZ}}(D) = \min(I(X;U) - I(Y;U)) = \min I(X;U|Y),$$

where the minimum is over all $p(u|x)$ and functions $\hat{x}(u,y)$ with $|\mathcal{U}| \leq |\mathcal{Y}| + 1$ such that $\mathrm{E}[d(X_1^L, \widehat{X}_1^L)] \leq D$. Visualizing $p(u|x)$ as a "test channel" is a common way to visualize the problem. The decoder then is the minimum-error estimator.

Unlike the Slepian–Wolf case, distributed encoding generally performs worse than joint encoding. However, this is still a positive result because the gap is very small under mild conditions [206] and the side information is not useful when sent only to the encoder. When the side information gives insight on the distortion measure used at the decoder, encoder-side information can aid compression [115]. Wyner–Ziv

coding has received considerable attention for video coding [1, 152, 183].

2.5.2 Rate–Distortion Theory for Small Distortion

In general, the rate–distortion function is difficult to characterize, but the Shannon lower bound (SLB) is a useful bounding tool when d is a difference distortion measure, i.e. $d(x, \hat{x}) = \rho(x - \hat{x})$ [164]. For MSE distortion, the SLB says

$$r(D) \geq h(X) - \frac{1}{2}\log(2\pi e D), \tag{2.24}$$

where h is the differential entropy. In the scalar case, it has been shown that this bound becomes tight as D becomes small [11, 100, 103]. In the multiterminal setting with N encoders, the SLB on sum rate is also shown to be tight for difference distortion measures when D is small, achieved through lattice quantization followed by Slepian–Wolf coding [207].

There are also low-distortion analysis for nondifference distortion measures using the SLB, more specifically *input-weighted locally quadratic* distortion measures [101]:

$$d(x_1^N, \hat{x}_1^N) = (x_1^N - \hat{x}_1^N)^T M(x_1^N)\,(x_1^N - \hat{x}_1^N) + \mathcal{O}(\|x_1^N - \hat{x}_1^N\|^3),$$

where we clarify that the input is N-dimensional but may be coded in blocks of length L. We will focus on a specific subclass called the weighted mean-squared error (W-MSE) criterion, which is related to fMSE:

$$d(x_1^N, \hat{x}_1^N) = \left\| W(x_1^N)(x_1^N - \hat{x}_1^N) \right\|^2, \tag{2.25}$$

where W is the source-dependent weighting matrix. Using a similar approach to the SLB, Linder and Zamir show that, in the joint-encoder case for W-MSE distortion,

$$\lim_{D\to 0} \left(r(D) + \frac{N}{2}\log(2\pi e D/N) \right) = h(X_1^N) + \mathrm{E}[\log |W(X_1^N)|]. \tag{2.26}$$

Recall again that the achievable strategy employs coding of blocks of length L of the

N-dimensional source.

The behavior of the rate–distortion function in the small-distortion regime is particularly relevant to this thesis because it allows comparison with high-resolution scalar quantization. We will use these results to determine the fMSE rate–distortion function, and subsequently the rate loss from using DFSQ theory in Chapter 4.

2.5.3 Other Related Problems

There are other works in rate–distortion theory that relate to this thesis. Here, we briefly summarize a subset of relevant ideas.

Finite blocklength analysis

Recent work to generalize the asymptotic nature of Shannon theory [146] has led to characterization of the rate–distortion function at finite blocklengths [73, 92]. In general, this analysis technique is meaningful for block lengths as low as 100, but is unsuitable for regimes traditionally considered in high-resolution theory. One interesting result is that the rate region of the Slepian–Wolf setup is no longer tight with the outer bound of joint encoding [178].

Remote sources

Most commonly, Shannon-theoretic analyses assume a node makes a perfect measurement of a source and all loss is due to compression. The concept of *remote sources* was introduced to allow for noisy information collection, and provides interesting trade-offs between communication and estimation [11, 193]. Remote sources have been integrated into the multiterminal source coding literature as well [42, 200, 207].

Coding for computing

A central theme of the thesis is caring about computations on source realizations. In Shannon theory, this problem has mostly be considered for discrete-alphabet sources. In the lossless case and a general computation g, the problem has been approached

through classification of functions by [70]. The use of graph entropy [91] can also aid in coding for computing [32, 41, 140]. For the lossy case, computing a function of a source using a helper extends the Wyner–Ziv framework [42, 199]. Here, it is useful to think about the computation as an instance of a remote source.

Interactive source coding

Consider a different network situation in which a pair of nodes communicate to learn each other's data. The communication is done in rounds and the objective is to find a relationship between distortion and the rates of each round. This has been studied in detail both in lossless and lossy situations [85, 138, 139]. Extensions of using interaction for computation were recently proposed in [106, 107]. To the best of our knowledge, high-resolution techniques have never been fruitfully applied to this problem.

Chapter 3

DFSQ Simplified

We have seen in Section 2.3 that distributed functional scalar quantization (DFSQ) theory provides optimality conditions and predicts performance of data acquisition systems in which a computation on acquired data is desired. In this chapter, we address two limitations of previous works: prohibitively expensive decoder design and a restriction to source distributions with bounded support. We show that a much simpler decoder has equivalent asymptotic performance to the conditional expectation estimator studied previously, thus reducing decoder design complexity. The simpler decoder has the feature of decoupled communication and computation blocks. Moreover, we extend the DFSQ framework with the simpler decoder to source distributions with unbounded support.

We begin in Section 3.1 by motivating this chapter. In Sections 3.2 and 3.3, we give distortion analysis and optimal quantizer design results. Finally, we provide examples to demonstrate convergence in Section 3.4 and conclude in Section 3.5.

3.1 Motivation

The central goal of this chapter is to develop a more practical method upon the theoretical foundations of [120]. Specifically, we provide new insight on how a simple decoder can be used in lieu of the optimal one in (2.12). Although the conditional expectations in the joint centroid (JC) conditions are offline computations, they may be

extremely difficult to describe analytically and are computationally infeasible for large N and K. Instead, we construct a decoder that is much more intuitive and practical; it simply applies the desired function g explicitly on the quantized measurements. Remarkably, the same asymptotic performance is obtained with the simpler decoder, so the optimization of quantizer point density is unchanged. To accommodate this change, a different set of conditions is required of g, λ_1^N, and $f_{X_1^N}$.

Additionally, we generalize the theory to infinite-support source variables and vector-valued computations, which expands the space of problems that DFSQ provides guarantees. To allow the generalized class of problems, we derive new conditions on the tail behavior of the source densities and computations that allow the distortion to be stably computed. In Appendix 3.E, we also consider applying the simpler decoder for sources over finite support using relaxed conditions.

Moreover, through simulation results we demonstrate that performance at moderate coding rates is well predicted by the asymptotic analysis. The similarity in distortion behavior of JC reconstruction (2.12) and a much simpler decoder demonstrates the convergence of midpoint and centroid reconstructions in this framework. Later in this chapter, we provide some simulation and discussion on this point.

The analysis presented here uses different assumptions on the source distributions and function than [120]—neither is uniformly more or less restrictive. Unlike in [120], we are able to allow the source variables to have infinite support. In fact, the functional setting allows us to generalize the classes of distributions whose reconstruction performance can be accurately predicted using high-resolution quantization theory. Both analyses contain rather technical conditions, and together they suggest a rather general applicability of DFSQ theory.

3.2 Univariate Functional Quantization

We first discuss the quantization of a scalar random variable X by $Q_{K,\lambda}$ to approximate $g(X)$. As mentioned, the decoder will apply g to the quantizer output $Q_{K,\lambda}(X)$ rather than compute the joint centroid condition like in [120]. We find the dependence

of fMSE on λ and then optimize with respect to λ to minimize fMSE.

Consider the following conditions on the source density f_X, point density λ of a companding quantizer, and computation of interest g:

UF1′. The source pdf f_X is continuous and positive on $\mathbb{R}$.

UF2′. The point density λ is continuous and positive on $\mathbb{R}$.

UF3′. The function g is continuous on $\mathbb{R}$ with everywhere-defined derivatives g' and g''.

UF4′. For $m = 0, 1, 2$,

$$f_X(x)|g''(x)|^m|g'(x)|^{2-m}/\lambda^{2+m}(x)$$

is integrable over $\mathbb{R}$.

UF5′. f_X, g and λ satisfy the tail condition

$$\lim_{y\to\infty} \frac{\int_y^\infty |g(x) - g(y)|^2 f_X(x)\,dx}{\left(\int_y^\infty \lambda(x)\,dx\right)^2} = 0,$$

and the corresponding condition for $y \to -\infty$.

UF6′. Define s as the derivative of the expander function c^{-1}, meaning $s(c(x)) = 1/\lambda(x)$. There exists some $B > 0$ such that $s(c(x))$ is decreasing for $x < -B$, s is increasing for $x > B$, and the tails of s satisfy

$$\int_{-\infty}^{c(-B)} s^{2+m}(c(x)/2)\,|g''(x)|^m\,|g'(x)|^{2-m}\,f_X(x)\,dx < \infty$$
$$\int_{c(B)}^{\infty} s^{2+m}((c(x)+1)/2)\,|g''(x)|^m\,|g'(x)|^{2-m}\,f_X(x)\,dx < \infty$$

for $m = 0, 1, 2$.

The main result of this section is on the fMSE induced by a quantizer $Q_{K,\lambda}$ under these conditions:

Theorem 3.1. *Assume f_X, g, and λ satisfy Conditions UF1′–UF6′. Then the fMSE*

$$D_{\text{fmse}}(K, \lambda) = \mathrm{E}\left[|g(X) - g(Q_{K,\lambda}(X))|^2\right]$$

satisfies the following limit:

$$\lim_{K\to\infty} K^2 D_{\text{fmse}}(K,\lambda) = \frac{1}{12}\,\mathrm{E}\left[\left(\frac{\gamma(X)}{\lambda(X)}\right)^2\right]. \tag{3.1}$$

Proof. See Appendix 3.A. □

3.2.1 Remarks

1. The fMSE in (3.1) is the same as in (2.14). We emphasize that the theorem shows that this fMSE is obtained by simply applying g to the quantized variables rather than using the optimal decoder (2.12). Further analysis on this point is given in Section 3.2.3.

2. One key contribution of this theorem is the additional tail condition for infinite-support source densities, which effectively limits the distortion contribution in the overload region. This generalizes the class of probability densities for which quantization distortion can be analyzed using high-resolution approximations [17, 19, 99].

3. The tail conditions in UF5′ imply the overload contributions to distortion become negligible as K becomes large, which is natural for well-behaved sources, computations and compressor functions. This is used to ensure Taylor's theorem can be successfully applied to bound fMSE. The tail conditions in UF6′ do not have simple interpretations but are necessary to employ the dominated convergence theorems used in the proof of Theorem 3.1 [99]. Both conditions are satisfied in many problems of interest.

4. When the source distribution has finite support, the tail conditions are no longer necessary and UF4′ can be simplified. We provide the proof of this case in Appendix 3.E.

5. When g is monotonic, the performance in (3.1) is as good as quantizing and communicating $g(X)$ [120, Lemma 5]. Otherwise, the use of a regular quantizer results in a distortion penalty, as illustrated in Example 3.1 of Section 3.4.

6. For linear computations, the functional sensitivity profile is flat, meaning the optimal quantizer is the same as in the MSE-optimized case. Hence, functional theory

will lead to new quantizer designs only when the computation is nonlinear.

7. Although we have assumed f_X, g and λ are "nice" in the sense that they are continuous and positive, the proof of Theorem 3.1 could allow f_X to be discontinuous or nondifferentiable at a finite number of points, provided the tail conditions still hold and a minor adjustment is made on how partition boundaries are chosen. Rather than elaborating further, we refer the reader to a similar extension in [120, Section III-F]. A similar argument can also be made for g having a finite number of discontinuities in its first and second derivatives.

8. For the high-resolution assumptions to hold, the point density should be positive where the source distribution is positive. However, a consequence of Theorem 3.1 is that there is no distortion contribution from regions where the functional sensitivity profile is zero, meaning the point density can be zero there. The coding of such "don't-care" intervals must be handled with care, as discussed in [120, Section VII].

3.2.2 Asymptotically Optimal Quantizer Sequences

Since the fMSE of Theorem 3.1 matches (2.14), the optimizing quantizers are the same. Using the recipe of Section 2.2, we can show the optimal point density for fixed-rate quantization is asymptotically

$$\lambda^*_{\text{fmse,fr}}(x) = \frac{\left(\gamma^2(x) f_X(x)\right)^{1/3}}{\int_{-\infty}^{\infty} \left(\gamma^2(t) f_X(t)\right)^{1/3}\, dt} \tag{3.2}$$

over the entire support of X, resulting in distortion

$$D^*_{\text{fmse,fr}}(R) \simeq \frac{1}{12} \left\|\gamma^2 f_X\right\|_{1/3} 2^{-2R}. \tag{3.3}$$

Meanwhile, optimization in the entropy-constrained case yields

$$\lambda^*_{\text{fmse,ec}}(x) = \frac{\gamma(x)}{\int_{-\infty}^{\infty} \gamma(t)\, dt} \tag{3.4}$$

over the entire support of X, resulting in distortion

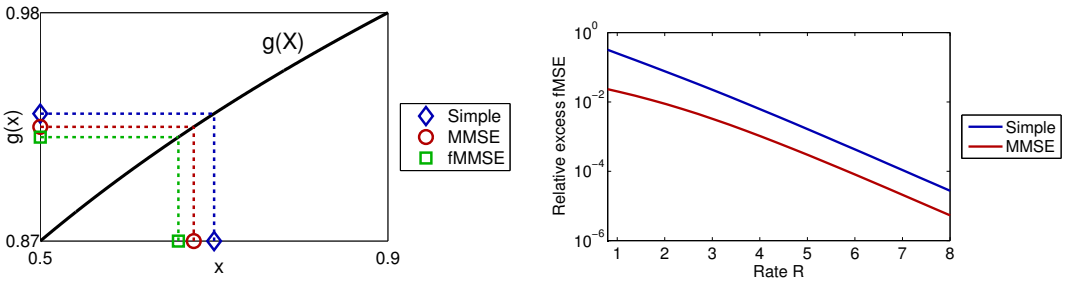
$$D^*_{\text{fmse,ec}}(R) \simeq \frac{1}{12}\, 2^{2h(X)+2\,\mathrm{E}[\log\gamma(X)]}\, 2^{-2R}. \tag{3.5}$$

Observe that while minimization of the distortion-rate expressions provides "optimal" companding quantizers, the distortion-rate expressions themselves are restricted to quantizer point density functions that satisfy UF4′–UF6′. Some of these conditions may be verified quite easily: for instance, UF4′ for $m = 0$ is equivalent to the asymptotic distortion expression being finite. Additionally, if the distribution and functional sensitivities satisfy certain properties—e.g. if the sensitivities possess a positive lower bound over the distribution's support—these conditions may be automatically satisfied. In general, the conditions must be checked on a case-by-case basis for the asymptotic analysis to rigorously hold. As demonstrated in Example 3.5 of Section 3.4, design based on the asymptotic analysis can be sensible even when the technical requirements are not satisfied.

Further care is needed in the entropy-constrained setting. Many computations yield γ that is not integrable over $\mathbb{R}$, making (3.4) invalid; for example, a linear computation leads to constant γ. When the source has finite support, the integral in the denominator of (3.4) can be reduced to one on that finite support, again yielding a valid, optimal normalized point density. Otherwise, one must use an unnormalized point density to represent the asymptotically-optimal companding quantizer sequence. We leave this generalization as future work.

3.2.3 Negligible Suboptimality of Simple Decoder

Recall that the decoder analyzed in this work is the computation g applied to midpoint reconstruction as formulated in (2.3). One may do better by applying g after finding the conditional MMSE estimate of X (using knowledge of the source distribution only) and would do best with the fMMSE estimator (2.12) (incorporating knowledge of the function as well). The codeword placements of the three decoders are visualized through an example in Figure 3-1a. The asymptotic match of the performance of the



(a) (b)

Figure 3-1: **(a)** Codeword placement under simple, MMSE, and fMMSE decoders. The simple decoder performs midpoint reconstruction followed by the application of the computation g. The MMSE decoder applies g to the conditional expectation of X within the cell. Finally, the fMMSE decoder determines (2.12) for the cell. In this example, the source distribution is exponential and the computation is concave. **(b)** Performance loss due to the suboptimal codeword placement with respect to rate. We can see that relative excess fMSE decreases linearly with rate and hence the fMSE of the resulting quantizers are asymptotically equivalent.

simple decoder to the optimal estimator (2.12) is a main contribution of this work.

The simple decoder is suboptimal because it does not consider the source distribution at all, or equivalently assumes the distribution is uniform and the functional sensitivity profile is constant over the cell. High-resolution analysis typically approximates the source distribution as uniform over small cells [66], and the proof of Theorem 3.1 uses the fact that the sensitivity is approximately flat over very small regions as well. Hence, the performance gap between the simple decoder and the fMMSE estimator becomes negligible in the high-resolution regime.

To illuminate the rate of convergence, we study the performance gap as a function of quantization cell width, which is dependent on the communication rate (Figure 3-1b). Through experimental observation, we see the relative excess fMSE (defined as $(D_{\text{dec}} - D_{\text{opt}})/D_{\text{opt}}$) appears exponential in rate, meaning

$$\frac{D_{\text{simple}}}{D_{\text{opt}}} \approx 1 + c_1 e^{-c_2 R}$$

for some constants c_1 and c_2. The speed at which the performance gap shrinks contributes greatly to why the high-resolution theory is successful even at low communication rates.

3.3 Multivariate Functional Quantization

We now describe the main result of the chapter for the scenario shown in Figure 2-4, where N random scalars X_1^N are individually quantized and a scalar computation $g(\widehat{X}_1^N)$ is performed. We will use a codebook size parameter κ and fractional allocations α_1^N such that every $\alpha_n > 0$ and $\sum_n \alpha_n = 1$; the codebook size for quantizer n is then $K_n = \lfloor \alpha_n \kappa \rfloor$. Since we are concerned with an asymptotic result, the use of κ ensures all codebooks grow at the same rate.

Assume the following conditions on the multivariate joint density, computation and quantizers:

MF1′. The joint pdf $f_{X_1^N}$ is continuous and positive on $\mathbb{R}^N$.

MF2′. For every $n \in [1 : N]$, the point density λ_n is continuous and positive on $\mathbb{R}$.

MF3′. The multivariate function g is continuous and twice differentiable in every argument over $\mathbb{R}^N$; that is, the first partial derivative $g_i = \partial g / \partial x_i$ and second partial derivative $g_{i,j} = \partial^2 g / \partial x_j \, \partial x_i$ are well-defined for every $i, j \in [1 : N]$.

MF4′. For any $n \in [1 : N]$,

$$f_{X_n}(x_n) \, |g_n(x_1^N)|^2 / \lambda_n^2(x_n) \tag{3.6}$$

is integrable over $\mathbb{R}$. Moreover, for any $i, j, n \in [1 : N]$,

$$f_{X_1^N}(x_1^N) \frac{|g_n(x_1^N)| \, |g_{i,j}(x_1^N)|}{\lambda_i(x_i)\lambda_j(x_j)\lambda_n(x_n)} \tag{3.7}$$

is integrable over $\mathbb{R}^N$, and, for $i, j, m, n \in [1 : N]$,

$$f_{X_1^N}(x_1^N) \frac{|g_{i,j}(x_1^N)| \, |g_{m,n}(x_1^N)|}{\lambda_i(x_i)\lambda_j(x_j)\lambda_m(x_m)\lambda_n(x_n)} \tag{3.8}$$

is integrable over $\mathbb{R}^N$.

MF5′. For $i, j \in [1 : N]$,

$$\frac{\mathrm{E}[(X_i - Q_{\lambda_i,K_i}(X_i))(X_j - Q_{\lambda_j,K_j}(X_j))]}{\sqrt{D_i\, D_j}} \to 0$$

as $\kappa \to \infty$, where $D_n = \mathrm{E}[|X_n - Q_{\lambda_n,K_n}(X_n)|^2]$.

MF6′. We adopt the notation $x_{\setminus n}$ for x_1^N with the nth element removed; the inverse operator $\tilde{x}(x_n, x_{\setminus n})$ outputs a length-N vector with x_n inserted as the nth element. Then for every index n, the following holds for every $x_{\setminus n}$:

$$\lim_{y\to\infty} \frac{\int_y^\infty |g(\tilde{x}(x, x_{\setminus n})) - g(\tilde{x}(y, x_{\setminus n}))|^2 f_{X_1^N}(\tilde{x}(x, x_{\setminus n}))\, dx}{\left(\int_y^\infty \lambda_n(x)\, dx\right)^2} = 0.$$

An analogous condition holds for the corresponding negative-valued tails.

MF7′. Define s_n as the derivative of the expander function c_n^{-1}, meaning

$$s_n(c_n(x)) = 1/\lambda_n(x).$$

There exists some $B > 0$ such that $s_n(c(x))$ is decreasing for $x < -B$, s_n is increasing for $x > B$, and the tails of s_n satisfy

$$\int_{-\infty}^{c_n(-B)} s_n^2(c_n(x)/2)\, \gamma_n^2(x)\, f_{X_n}(x)\, dx < \infty,$$
$$\int_{c_n(B)}^{\infty} s_n^2((c_n(x)+1)/2)\, \gamma_n^2(x)\, f_{X_n}(x)\, dx < \infty,$$

for all $n \in [1 : N]$. This condition is a generalization of UF6′ for $m = 0$ applied to (3.6). Effectively, it bounds the tail contributions of an integral with the integrand being a modified version of (3.6). We also require similar conditions for (3.7) and (3.8), which are analogous to UF6′ for $m = 1$ and $m = 2$ respectively. We omit the exact form here for clarity of presentation.

Recalling $Q_{K_1^N,\lambda_1^N}$ and λ_1^N represent a set of N quantizers and point densities respectively, we present a theorem similar to Theorem 3.1:

Theorem 3.2. *Assume $f_{X_1^N}$, g, and λ_1^N satisfy conditions MF1′–MF7′. Also assume*

a fractional allocation α_1^N *such that every* $\alpha_n > 0$ *and* $\sum_n \alpha_n = 1$*, meaning a set of quantizers* $Q_{K_1^N,\lambda_1^N}$ *will have* $K_n = \lfloor \alpha_n \kappa \rfloor$ *for some total allocation* κ*. Then the fMSE*

$$D_{\text{fmse}}(K_1^N, \lambda_1^N) = \mathrm{E}\left[|g(X_1^N) - g(Q_{K_1^N,\lambda_1^N}(X_1^N))|^2\right]$$

satisfies the following limit:

$$\lim_{\kappa\to\infty} \kappa^2 D_{\text{fmse}}(K_1^N, \lambda_1^N) = \sum_{n=1}^{N} \frac{1}{12\alpha_n^2} \mathrm{E}\left[\left(\frac{\gamma_n(X_n)}{\lambda_n(X_n)}\right)^2\right]. \tag{3.9}$$

Proof. See Appendix 3.B. □

3.3.1 Remarks

1. Like in the univariate case, the simple decoder has performance that is asymptotically equivalent to the more complicated optimal decoder (2.12).

2. Here, the computation cannot generally be performed before quantization because encoders are distributed. The exception is when the computation is *separable*, meaning it can be decomposed into a linear combination of computations on individual scalars. As a result, for each n the partial derivative of g depends only on X_n and the functional sensitivity profile simplifies to the univariate case, as demonstrated in Example 3.2 of Section 3.4.

3. The strict requirements of MF1′ and MF3′ could potentially be loosened. However, simple modification of individual quantizers like in the univariate case is insufficient since discontinuities may lie on a manifold that is not aligned with the partition boundaries of the Cartesian product of N scalar quantizers. As a result, the error from using a planar approximation through Taylor's theorem may decay at the same rate as in (3.9), which would invalidate Theorem 3.2. However, based on experimental observations, such as in Example 3.5 of Section 3.4, we believe that when these discontinuities exist on a manifold of Jordan measure zero their error may be accounted for. Techniques similar to those in the proofs from [120] could potentially be useful in showing this rigorously.

4. Condition MF5′ is known as the asymptotic whiteness property (AWP). For uniform quantization with midpoint reconstruction and nonuniform quantization with centroid reconstruction, it is shown in [80, 189] that the quantization error for each cell converges to a uniform density sufficiently fast such that the correlation of the quantization error components vanishes faster than the distortion under mild regularity conditions. We leave the AWP as a condition, but mention that establishing it under general conditions for companding quantizers with midpoint reconstruction is an interesting open problem. The solution may rely on extending Theorem 1 of [189] to hold after the expansion step of the compander. To prove the convergence of the quantization error correlation to zero, it may be necessary to consider midpoint reconstruction both before and after expansion using techniques developed in [113].

5. When the joint source density has finite support, the tail conditions are no longer necessary. The exact proof parallels the discussion in Appendix 3.E.

3.3.2 Asymptotically Optimal Quantizer Sequences

As in the univariate case, the optimal quantizers match those in previous DFSQ work since the distortion equations are the same. Using Hölder's inequality, the optimal point density for fixed-rate quantization for each source n (communicated with rate R_n) is asymptotically

$$\lambda^*_{n,\text{fmse,fr}}(x) = \frac{\left(\gamma_n^2(x) f_{X_n}(x)\right)^{1/3}}{\int_{-\infty}^{\infty} \left(\gamma_n^2(t) f_{X_n}(t)\right)^{1/3}\, dt} \tag{3.10}$$

over the support of X_n, with fMSE

$$D^*_{\text{fmse,fr}}(R_1^N) \simeq \frac{1}{12} \sum_{n=1}^{N} \|\gamma_n^2 f_{X_n}\|_{1/3}\, 2^{-2R_n}. \tag{3.11}$$

Similarly, the best point density for the entropy-constrained case is asymptotically

$$\lambda^*_{n,\text{fmse,ec}}(x) = \frac{\gamma_n(x)}{\int_{-\infty}^{\infty} \gamma_n(t)\, dt} \tag{3.12}$$

over the support of X_n, leading to a fMSE of

$$D^*_{\text{fmse,ec}}(R_1^N) \simeq \frac{1}{12}\sum_{n=1}^{N} 2^{2h(X_n)+2\,\mathrm{E}[\log\gamma(X_n)]}\, 2^{-2R_n}. \tag{3.13}$$

We present performance while leaving the fractional allocation α_1^N as a parameter. Given a total communication rate constraint R, we can also optimize α_1^N. Rather than repeat the results here, we point to similar work in [120, Lemma 4].

As in the univariate case, this optimization arrives with the caveat that conditions MF4′–MF7′ must be satisfied by the resulting point density functions. In general this must be verified in a case-by-case basis, but as noted in Section 3.2.2, in practice the analysis can be useful even when the requirements are not satisfied.

3.3.3 Vector-Valued Functions

In Theorem 3.2, we assumed the computation g is scalar-valued. For completeness, we now consider vector-valued functions, where the output of g is a vector in $\mathbb{R}^M$. Here, the distortion measure is a weighted fMSE:

$$D_{\text{fmse}}(K_1^N, \lambda_1^N, \beta_1^M) = \sum_{m=1}^{M} \beta_m\, \mathrm{E}\left[\left|g^{(m)}(X_1^N) - g^{(m)}(Q_{K_1^N,\lambda_1^N}(X_1^N))\right|^2\right],$$

where β_1^M is a set of scalar weights and $g^{(m)}$ is the mth entry of the output of g. Through a natural extension of the proof of Theorem 3.2, we can find the limit of the weighted fMSE assuming each entry of the vector-valued function satisfies MF1′–MF7′.

Corollary 3.1. *The weighted fMSE of a source $f_{X_1^N}$, computation g, set of scalar quantizers $Q_{K_1^N,\lambda_1^N}$, and fractional allocation α_1^N satisfies the following limit:*

$$\lim_{\kappa\to\infty} \kappa^2 D_{\text{fmse}}(K_1^N, \lambda_1^N, \beta_1^M) = \sum_{n=1}^{N} \frac{1}{12\alpha_n^2}\, \mathrm{E}\left[\left(\frac{\gamma_n(X_n, \beta_1^M)}{\lambda_n(X_n)}\right)^2\right], \tag{3.14}$$

where the combined functional sensitivity profile *is*

$$\gamma_n(x, \beta_1^M) = \left(\sum_{m=1}^{M} \beta_m \, \mathrm{E}\left[|g_n^{(m)}(X_1^N)|^2 \,\middle|\, X_n = x \right] \right)^{1/2}.$$

The point densities given in (3.10) and (3.12) are again optimal under this new definition of γ_n.

3.4 Examples

In this section, we present examples for both univariate and multivariate functional quantization using asymptotic expressions and empirical results from sequences of real quantizers. The empirical results are encouraging since the convergence to asymptotic limits is fast, usually when the quantizer rate is about 4 bits per source variable. This is because the Taylor remainder term in the distortion calculation decays with an extra κ factor, which is exponential in the rate.

3.4.1 Examples for Univariate Functional Quantization

Below we present an example of functional quantization in the univariate case. The theoretical results follow directly from Section 3.2.

Example 3.1. Assume $X \sim \mathcal{N}(0, 1)$ and $g(x) = x^2$, yielding a functional sensitivity profile $\gamma(x) = 2|x|$. We consider uniform quantizers, optimal "ordinary" quantizers (quantizers optimized for distortion of the source variable rather than the computation) given in Section 2.2, and optimal functional quantizers given in Section 3.2.2, for a range of rates. The point densities of these quantizers, the source density f_X, and computation g satisfy UF1′–UF6′ and hence we use Theorem 3.1 to find asymptotic distortion performance. We also design practical quantizers for a range of R and find the empirical fMSE through Monte Carlo simulations. In the fixed-rate case, theoretical and empirical performance are shown (Figure 3-2a).

The distortion-minimizing uniform quantizer has a granular region that depends

on R, which was explored in [72]. Here, we simply perform a brute-force search to find the best granular region and the corresponding distortion. Surprisingly, this choice of the uniform quantizer performs better over moderate rate regions than the MSE-optimized quantizer. This is because the computation is less meaningful where the source density is most likely and the MSE-optimized quantizer places most of its codewords. Hence, one lesson from DFSQ is that using standard high-resolution theory may yield *worse* performance than a naive approach for some computations. Meanwhile, the functional quantizer optimizes for the computation and gives an additional 3 dB gain over the optimal ordinary quantizer. There is still a loss in using regular quantizers due to the computation being non-monotonic. In fact, if the computation can be performed prior to quantization, we gain an extra bit for encoding the magnitude and thus 6 dB of performance. This illustrates Remark 2 of Section 3.2.1.

In the fixed-rate case, the empirical performance approaches the distortion limit described by Theorem 3.1. The convergence is fast and the asymptotic results predict practical quantizer performance at rates as low as 4 bits/sample.

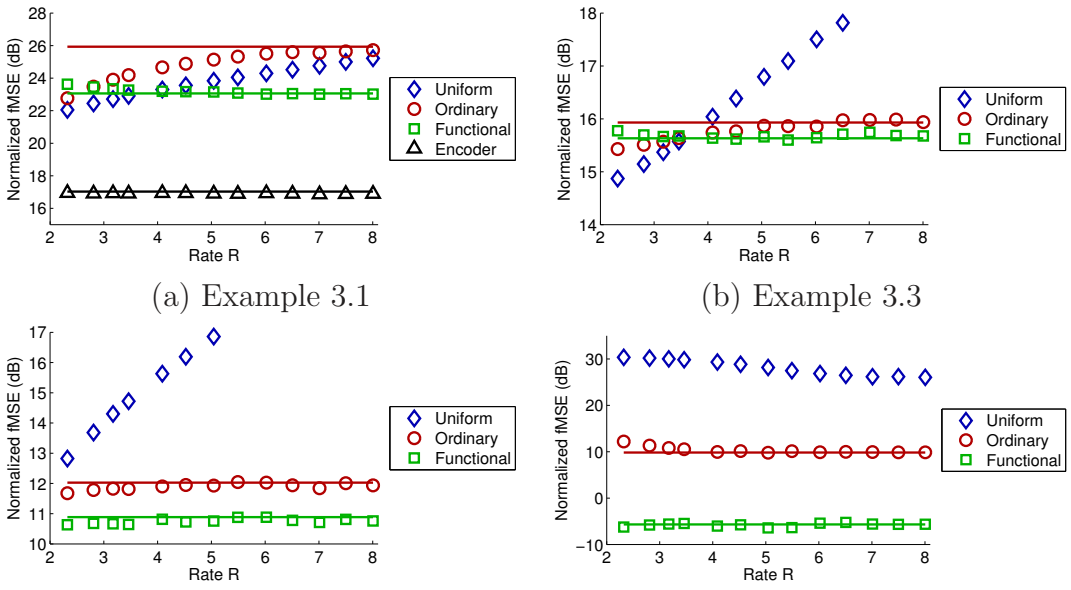
3.4.2 Examples for Multivariate Functional Quantization

We next provide four examples that follow from the theory of Section 3.3.

Example 3.2. Let N sources be iid standard normal random variables and the computation be $g(x_1^N) = \|x_1^N\|_2^2$. Since the computation is separable, the functional sensitivity profile of each source is $\gamma_n(x) = 2|x|$, and the quantizers are the same as in Example 3.1. The distortion is also the same, except now scaled by N.

Example 3.3. We now consider a more interesting extension of Example 3.2 where the sources are correlated and the computation is $g(x_1^N) = \|x_1^N\|_2$. Because the norm is not squared, the computation is no longer separable. For two jointly Gaussian random variables distributed $\mathcal{N}(0,1)$, a correlation coefficient of ρ implies that

$$X_2 = \rho X_1 + \sqrt{1-\rho^2}N,$$



(a) Example 3.1 (b) Example 3.3

(c) Example 3.4 (d) Example 3.5

Figure 3-2: Empirical and theoretical performance for the ordinary and functional quantizers for: **(a)** a scalar Gaussian source and $g(x) = x^2$; **(b)** jointly Gaussian sources with correlation coefficient 0.5 and $g(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$; **(c)** exponential sources with parameter $\lambda = 1$ and $g(x_1, x_2) = x_1/(1 + x_2)$; and **(d)** $N = 10$ exponential sources and $g(x_1^N) = \min(x_1^N)$. Note that we also include empirical results for uniform quantizers that have different granular regions depending on the quantization rate and the case when the computation is performed before quantization in (a), labeled "Encoder." Theoretical performance is determined using Theorem 3.2 and are represented by solid lines. Experimental validation is determined by designing real quantizers using the compander model and running Monte Carlo simulations; the resulting fMSE is represented by markers. To emphasize the gap between the results and to illustrate convergence to the high-resolution approximation, we normalize the plots by multiplying fMSE by 2^{2R}.

where N is standard normal and independent of X_1. The functional sensitivity profile then becomes

$$\begin{aligned}\gamma_1^2(x) &= \left(\mathrm{E}[|g_1(X_1, X_2)|^2 \,|\, X_1 = x]\right)^{1/2} \\ &= \mathrm{E}\left[\frac{X_1^2}{X_1^2 + X_2^2} \,\middle|\, X_1 = x\right] \\ &= \mathrm{E}_N\left[\frac{x^2}{x^2 + (\rho x + \sqrt{1-\rho^2}N)^2}\right]\end{aligned}$$

In Figure 3-2b, we demonstrate the convergence of the distortion from sequences of companding quantizers to the asymptotic behavior for $\rho = 0.5$. Similar results can be obtained for other choices of ρ.

Example 3.4. Consider two iid exponential sources X_1 and X_2 with parameter $\lambda = 1$; we wish to compute $g(x_1, x_2) = x_1/(a + x_2)$, where we let $a = 1$. Using (2.13), the functional sensitivity profiles are

$$\gamma_1(x) = \int_0^\infty e^{-x_2} \cdot (1 + x_2)^{-2}\, dx_2 \approx 0.635$$

and $\gamma_2(x) = (1 + x)^{-2}/\sqrt{3}$. In Figure 3-2c, we experimentally verify that sequences of real quantizers approach the predicted distortion–rate trade-off.

Example 3.5. Let N sources be iid exponential with parameter $\lambda = 1$ and the computation be $g(x_1^N) = \min(x_1^N)$. In this case, Condition MF3′ is not satisfied since there exists $N(N-1)/2$ two-dimensional planes where the derivative is not defined. However, as discussed in the remarks on Theorem 3.2, we strongly suspect we can disregard the distortion contributions from these surfaces. The overall performance, ignoring the violation of condition MF3′, may be analyzed using the functional sensitivity profile:

$$\begin{aligned}\gamma_n(x) &= \left(\mathrm{E}[|g_n(X_1^N)|^2 \,|\, X_n = x]\right)^{1/2} \\ &= \left(\Pr\{\min(X_1^N) = X_n \,|\, X_n = x\}\right)^{1/2} \\ &= \left(e^{-\lambda x}\right)^{(N-1)/2},\end{aligned}$$

where the third line follows from the cdf of exponential random variables.

In Figure 3-2d, we experimentally verify that the asymptotic predictions are precise. This serves as evidence that MF3′ may be loosened.

3.5 Conclusions

In this work, we have extended distributed functional scalar quantization to a general class of finite- and infinite-support distributions, and demonstrated that a simple decoder, performing the computation directly on the quantized measurements, achieves asymptotically equivalent performance to the fMMSE decoder. Although there are some technical restrictions on the source distributions and computations to ensure the high-resolution approximations are legitimate, the main goal of the chapter is to show that DFSQ theory is widely applicable to distributed acquisition systems without requiring a complicated decoder. Furthermore, the asymptotic results give good approximations for the performance at moderate quantization rates.

DFSQ has immediate implications in how sensors in acquisition networks collect and compress data when the designer knows the computation to follow. Using both theory and examples, we demonstrate that knowledge of the computation may change the quantization mapping and improve fMSE. Because the setup is very general, there is potential for impact in areas of signal acquisition where quantization is traditionally considered as a black box. Examples include multi-modal imaging technologies such as 3D imaging and parallel MRI. This theory can also be useful in collecting information for applications in machine learning and data mining. In these fields, large amounts of data are collected but the measure of interest is usually some nonlinear, low-dimensional quantity. DFSQ provides insight on how data should be collected to provide more accurate results when the resources for acquiring and storing information are limited.

3.A Proof of Theorem 3.1

Taylor's theorem states that a function g that is $n+1$ times continuously differentiable on a closed interval $[a, x]$ takes the form

$$g(x) = g(a) + \left(\sum_{i=1}^{n} \frac{g^{(i)}(a)}{i!} (x-a)^i \right) + R_n(x, a),$$

with a Taylor remainder term

$$R_n(x, a) = \frac{g^{(n+1)}(\xi)}{(n+1)!} (x-a)^{n+1}$$

for some $\xi \in [a, x]$. More specific to our framework, for any $x \in [c_k, p_k)$, the first-order remainder is bounded as

$$|R_1(x, c_k)| \leq \frac{1}{2} \max_{\xi \in [c_k, p_k]} |g''(\xi)| \, (p_k - c_k)^2. \tag{3.15}$$

We will denote the length of the partition corresponding to the kth codeword as $I_k = p_k - p_{k-1}$ and let $I(x) = I_k$ if $x \in P_k$. Moreover, we define $\tilde{g}$ as a piecewise-constant upper bound to the second derivative of g over the partition of $Q_{K,\lambda}$:

$$\tilde{g}(x) = \sup_{t \in P_k} |g''(t)| \quad \text{if } x \in P_k,\ k \in [1:K] \tag{3.16}$$

Since c_k is at the midpoint between p_k and p_{k-1}, we can rewrite the Taylor remainder term as

$$|R_1(x, c_k)| \leq \frac{1}{8} \tilde{g}(x) I^2(x). \tag{3.17}$$

Consider expansion of $D_{\text{fmse}}(K, \lambda)$ by total expectation:

$$D_{\text{fmse}}(K, \lambda) = \sum_{k=0}^{K-1} \int_{p_k}^{p_{k+1}} |g(x) - g(c_k)|^2 f_X(x) \, dx.$$

We would like to eliminate the first and last components of the sum because the unbounded interval of integration would cause problems with the Taylor expansion

employed later. The last component is

$$\int_{p_{K-1}}^{\infty} |g(x) - g(p_{K-1})|^2 f_X(x)\, dx, \tag{3.18}$$

where we have used $c_K = p_{K-1}$. By Condition UF5′, this is asymptotically negligible in comparison to

$$\left(\int_{p_{K-1}}^{\infty} \lambda(x)\, dx \right)^2 = \frac{1}{K^2}.$$

Thus (3.18) does not contribute to $\lim_{K\to\infty} K^2 D_{\text{fmse}}(K, \lambda)$. We can similarly eliminate the first term, yielding

$$D_{\text{fmse}}(K, \lambda) \simeq \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} |g(x) - g(c_k)|^2 f_X(x)\, dx, \tag{3.19}$$

where we recall $\simeq$ indicates that the ratio of the two expressions approaches one as K increases. Effectively, UF5′ promises that the tail of the source distribution is decaying fast enough that we can ignore the distortion contributions outside the extremal codewords.

Assuming UF3′, further expansion of (3.19) using Taylor's theorem yields:

$$\begin{aligned}
K^2 D_{\text{fmse}}(K, \lambda) &\simeq K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} |g'(c_k)(x - c_k) + R_1(x, c_k)|^2 f_X(x)\, dx \\
&\leq \underbrace{K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} |g'(c_k)|^2 \, |x - c_k|^2 f_X(x)\, dx}_{A} && (3.20) \\
&\quad + \underbrace{K^2 \sum_{k=1}^{K-2} 2 \int_{p_k}^{p_{k+1}} |R_1(x, c_k)| \, |g'(c_k)| \, |x - c_K| f_X(x)\, dx}_{B} && (3.21) \\
&\quad + \underbrace{K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} R_1(x, c_k)^2 f_X(x)\, dx}_{C}. && (3.22)
\end{aligned}$$

Of the three terms, only term A has a meaningful contribution, which has the following

asymptotic form:

$$\begin{aligned}
\lim_{K\to\infty} K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} & |g'(c_k)|^2 \, |x - c_k|^2 \, f_X(x)\, dx \\
&\overset{(a)}{=} \lim_{K\to\infty} K^2 \int_{p_1}^{p_{K-1}} |g'(x)|^2 \, |x - Q_{K,\lambda}(x)|^2 \, f_X(x)\, dx \\
&\overset{(b)}{=} \lim_{K\to\infty} K^2 \int_{\mathbb{R}} |g'(x)|^2 \, |x - Q_{K,\lambda}(x)|^2 \, f_X(x)\, dx \\
&\overset{(c)}{=} \frac{1}{12} \int_{\mathbb{R}} \left(\frac{g'(x)}{\lambda(x)} \right)^2 f_X(x)\, dx, \qquad (3.23)
\end{aligned}$$

where (a) follows from the definition of $Q_{K,\lambda}$; (b) from $p_1 \to -\infty$ and $p_{K-1} \to \infty$; and (c) from an extension of the proof by Linder [99], which is given in Theorem 3.3 in Appendix 3.C. Conditions UF2′, UF4′ and UF6′ for $m = 0$ are used here. Noting that $\gamma(x) = |g'(x)|$ gives (3.1).

The higher-order error terms become negligible with increasing K using the bound reviewed in (3.15):

$$\begin{aligned}
\lim_{K\to\infty} K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} & |R_1(x, c_k)| \, |g'(c_k)| \, |x - c_K| \, f_X(x)\, dx \\
&\overset{(a)}{=} \lim_{K\to\infty} \frac{K^2}{4} \int_{p_1}^{p_{K-1}} |\tilde{g}(x)| \, I^2(x) \, |g'(x)| \, |x - Q_{K,\lambda}(x)| \, f_X(x)\, dx \\
&\overset{(b)}{=} \lim_{K\to\infty} \frac{K^2}{4} \int_{\mathbb{R}} |\tilde{g}(x)|^2 \, I^2(x) \, |g'(x)| \, |x - Q_{K,\lambda}(x)| \, f_X(x)\, dx \\
&\overset{(c)}{=} \lim_{K\to\infty} \frac{1}{4K} \int_{\mathbb{R}} \frac{|\tilde{g}(x)|^2 \, |g'(x)|}{\lambda^3(x)} f_X(x)\, dx \\
&\overset{(d)}{=} 0, \qquad (3.24)
\end{aligned}$$

where (a) follows from bounding $R_1(x, c_k)$ using (3.17); (b) from $p_1 \to -\infty$ and $p_{K-1} \to \infty$; (c) from a similar extension of Theorem 3.3 (see Appendix 3.D), using UF2′ and UF6′ for $m = 1$; and (d) from UF4′ for $m = 1$. Compared to (3.23), there is an extra $1/K$ factor arising from the second-order Taylor error, which drives term B to zero. A similar analysis can be used to show that expansion term C scales as $1/K^2$ with growing codebook size and is therefore also negligible. Here, conditions

UF4′ and UF6′ for $m = 2$ are needed.

3.B Proof of Theorem 3.2

We parallel the proof of Theorem 3.1 using Taylor expansion and bounding the distortion contributions of each granular cell. By the first-order version of the multivariate Taylor's theorem, a function that is twice continuously differentiable on a closed ball containing a_1^N takes the form

$$g(x_1^N) = g(a_1^N) + \sum_{n=1}^{N} \left[g_n(a_1^N)(x_n - a_n) \right] + R_1(x_1^N, a_1^N),$$

where we recall that $g_n(x_1^N)$ is the partial derivative of g with respect to the nth argument evaluated at the point x_1^N. The remainder term is bounded by

$$\left| R_1(x_1^N, a_1^N) \right| \leq \sum_{i=1}^{N} \sum_{j=1}^{N} |x_i - a_i| \, |x_j - a_j| \, |g_{i,j}(x_1^N)|, \tag{3.25}$$

where $g_{i,j}$ is the second-order partial derivation with respect to x_i first and then x_j evaluated at x_1^N.

Let $\mathcal{T}_N$ be an indexing of the cells in the Cartesian product of N scalar quantizers, excluding the overload regions. By total expectation, we find the distortion of each partition cell and sum their contributions. By Condition MF5′, the distortion from overload cells become negligible with increasing κ and can be ignored. Using Taylor's theorem and MF4′, the scaled total distortion becomes

$$\kappa^2 D_{\text{fmse}}(K_1^N, \lambda_1^N) \leq A + 2B + C, \tag{3.26}$$

where

$$A = \kappa^2 \sum_{t \in \mathcal{T}_N} \int_{x_1^N \in t} \sum_{i=1}^{N} \sum_{j=1}^{N} |g_i\left((c_t)_1^N\right)| \, |g_j\left((c_t)_1^N\right)| \, |x_i - c_{t,i}| \, |x_j - c_{t,j}| \, f_{X_1^N}(x_1^N) \, dx_1^N,$$

$$B = \kappa^2 \sum_{t \in \mathcal{T}_N} \int_{x_1^N \in t} \sum_{n=1}^{N} |g_n\left((c_t)_1^N\right)| \, |x_n - c_{t,n}| \, |R_1\left(x_1^N, (c_t)_1^N\right)| \, f_{X_1^N}(x_1^N) \, dx_1^N,$$

$$C = \kappa^2 \sum_{t \in \mathcal{T}_N} \int_{x_1^N \in t} R_1^2\left(x_1^N, (c_t)_1^N\right) f_{X_1^N}(x_1^N) \, dx_1^N.$$

Let us consider the summands of A where $i = j$:

$$\sum_{t \in \mathcal{T}_N} \sum_{n=1}^{N} \int_{x_1^N \in t} |g_n\left((c_s)_1^N\right)|^2 \, |x_n - c_{t,n}|^2 f_{X_1^N}(x_1^N) \, dx_1^N. \tag{3.27}$$

We note that these distortion contributions are equivalent to those in the univariate case and can apply the derivations in Theorem 3.1. Using Conditions MF2′, MF3′ and MF7′, (3.27) approaches the integral expression

$$\sum_{n=1}^{N} \frac{1}{12\alpha_n^2} \, \mathrm{E}\left[\left(\frac{g_n\left(X_n\right)}{\lambda_n(X_n)}\right)^2\right] = \sum_{n=1}^{N} \frac{1}{12\alpha_n^2} \, \mathrm{E}\left[\left(\frac{\gamma_n(X_n)}{\lambda_n(X_n)}\right)^2\right],$$

where the expectation on the left-hand side is with respect to the joint density $f_{X_1^N}$. Using the definition of functional sensitivity profile in (2.13), we get the right-hand side, where the expectation is only with respect to X_n.

We now consider the remaining summands of A where $i \neq j$, corresponding to the correlation between quantization errors in the granular region. Under the asymptotic whiteness property MF5′, the distortion contributions from these terms decay *faster* than in the terms in (3.27) in the granular region; therefore, they do not contribute to the asymptotic distortion. In Remark 3 of Section 3.3.1, we discuss generalizing to discontinuous densities and computations. Some care is needed so that this does not violate the validity of the asymptotic whiteness property.

We will now parallel the results of Appendix 3.A to show the higher-order error terms B and C are negligible with large κ. We denote the length of the partition

corresponding to the kth codeword of the nth quantizer as $I_{n,k}$ and let $I_n(x) = I_{n,k}$ if $x \in P_{n,k}$. Moreover, we define $\tilde{g}_{i,j}$ as a piecewise-constant upper bound to the second-order partial derivative of g over the partition of $Q_{K,\lambda}$:

$$\tilde{g}_{i,j}(x_1^N) = \sup_{\hat{x}_1^N \in t} |g_{i,j}(\hat{x}_1^N)| \quad \text{if } x_1^N \in t,$$

where t is an N-dimensional cell in $\mathcal{T}_N$. We can then bound (3.25):

$$\left|R_1(x_1^N, a_1^N)\right| \leq \frac{1}{4} \sum_{i=1}^{N} \sum_{j=1}^{N} I_j(x_i) I_i(x_j) \tilde{g}_{i,j}(x_1^N). \tag{3.28}$$

We now consider B:

$$\begin{aligned}
&\lim_{\kappa\to\infty} \kappa^2 \sum_{t\in\mathcal{T}_N} \int_{x_1^N \in t} \sum_{n=1}^{N} \left|g_n\left((c_t)_1^N\right)\right| |x_n - c_{t,n}| \left|R_1\left(x_1^N, (c_t)_1^N\right)\right| f_{X_1^N}(x_1^N)\, dx_1^N, \\
&\qquad \overset{(a)}{\leq} \lim_{\kappa\to\infty} \frac{\kappa^2}{4} \int_{\mathbb{R}^N} \left(\sum_{n=1}^{N} |g_n(x_1^N)| \, |x_n - Q_{K_n,\lambda_n}(x_n)| \right) \\
&\qquad\qquad \cdot \left(\sum_{i=1}^{N} \sum_{j=1}^{N} I_i(x_i)\, I_j(x_j)\, \tilde{g}_{i,j}(x_1^N) \right) f_{X_1^N}(x_1^N)\, dx_1^N, \\
&\qquad = \lim_{\kappa\to\infty} \frac{\kappa^2}{4} \int_{\mathbb{R}^N} \left(\sum_{n=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |g_n(x_1^N)| \, |x_n - Q_{K_n,\lambda_n}(x_n)| \right. \\
&\qquad\qquad \left. \cdot\, I_i(x_i)\, I_j(x_j)\, \tilde{g}_{i,j}(x_1^N) \right) f_{X_1^N}(x_1^N)\, dx_1^N, \\
&\qquad \overset{(b)}{=} \lim_{\kappa\to\infty} \frac{1}{4\kappa\alpha_i\alpha_j\alpha_n} \sum_{n=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{\mathbb{R}^N} \frac{|g_n(x_1^N)| \, |g_{i,j}(x_1^N)|}{\lambda_i(x_i)\lambda_j(x_j)\lambda_n(x_n)} f_{X_1^N}(x_1^N)\, dx_1^N, \\
&\qquad = 0,
\end{aligned}$$

where (a) follows from bounding $R_1(x_1^N, c_t)$ using (3.28) and the fact that the limits of integration converge to $\mathbb{R}^N$; and (b) from a generalization of the proof by Linder [99], which relies on the dominated convergence theorem to show how interval lengths can converge to the reciprocal of the point density. For this case, there is an extra $1/\kappa$ factor which drives B to zero, using conditions MF4′ and MF7′. Note that for general vector quantizers, a companding function may not exist. However, the simple

structure arising from a Cartesian product of N scalar quantizers is nicely represented, which allows Linder's method to be adequate.

Remainder term C is negligible in a similar manner (vanishing with $1/\kappa^2$), which proves the theorem.

3.C Quantization for Weighted Distortion

In this section, we prove a modest extension to Linder's rigorous results [99] on the distortion of companding quantizers on sources with infinite support. The addition here is a weighting function w inside the integral of the MSE distortion:

$$\int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2 \, w(x) \, f_X(x) \, dx.$$

Linder's result for MSE depends heavily on the dominated convergence theorem and its generalization. We will follow a similar strategy, except on a "weighted" probability density that is not required to integrate to one.

Recall that a scalar companding quantizer $Q_{K,\lambda}$ is specified by the codebook size K and point density λ, where λ is the derivative of the compressor function c. In this section, we will be explicit that we are considering a sequence of quantizers indexed by K that are constructed using the companding model. The partition points of $Q_{K,\lambda}$ are defined as $p_{k,K} = c^{-1}(k/K)$ and the codewords are determined using midpoint reconstruction, $c_{k,K} = (p_{k-1,K} + p_{k,K})/2$, except for the extremal codewords. We additionally define the derivative of the expander function c^{-1} as s, where $s(c(x)) = 1/\lambda(x)$, and the interval that is mapped to codeword $c_{k,K}$ as $I_{k,K} = [p_{k-1,K}, p_{k,K})$. We let μ denote the Lebesgue measure.

We impose the following conditions on f_X, w, and λ:

LC1. The point density λ is continuous and positive on $\mathbb{R}$.

LC2. $f_X(x)w(x)/\lambda^2(x)$ is Lebesgue integrable over $\mathbb{R}$.

LC3. There exists some $B > 0$ such that $\lambda(x)$ is increasing for $x < -B$ and is decreasing for $x > B$.

LC4. The inverse of λ, s, satisfies

$$\int_{-\infty}^{B} s^2(c(x)/2)\, w(x)\, f_X(x)\, dx < \infty,$$
$$\int_{-B}^{\infty} s^2((c(x)+1)/2)\, w(x)\, f_X(x)\, dx < \infty.$$

Before stating the main result, we define several sequences of functions that will be needed in the proof.

Definition 3.1. Consider a function h that is continuous, positive and integrable. The piecewise constant and truncated approximation to h over the partition induced by quantizer $Q_{K,\lambda}$ is defined as

$$\tilde{h}_K(x) = \begin{cases} \frac{1}{A_K(h)\mu(I_{k,K})} \int_{I_{k,K}} h(t)\, dt, & \text{for } x \in I_{k,K},\ k \in [2 : (K-1)]; \\ 0, & \text{otherwise,} \end{cases}$$

where

$$A_K(h) = \frac{\int_{p_{1,K}}^{p_{K-1,K}} h(x)\, dx}{\int_{-\infty}^{\infty} h(x)\, dx}.$$

Using the Lebesgue differentiation theorem, $\tilde{h}_K \to h$ as $K \to \infty$ a.e. with respect to μ.

Definition 3.2. We define as an approximation to s a function $s_K : (0,1) \to (0,\infty)$:

$$s_K(y) = \begin{cases} \sup_{t \in [\frac{k-1}{K}, \frac{k}{K})} s(t), & \text{for } y \in I_{k,K},\ k \in [2 : (K-1)]; \\ s(y), & \text{otherwise.} \end{cases}$$

The approximation s_K is the piecewise-constant function that most tightly upper bounds s on the granular region. We note that $s_K \to s$ as $K \to \infty$ by the continuity of s, which follows from LC1. Notice a slight modification in the definition of s_K from that in [99], due to the different placement of codewords in the extremal quantization cells.

Definition 3.3. We define as an approximation to s a function $\hat{s}_K : (0,1) \to (0,\infty)$:

$$\hat{s}_K(y) = \begin{cases} K\mu(I_{k,K}), & \text{for } y \in I_{k,K},\ k = [2 : (K-1)]; \\ 0 & \text{otherwise.} \end{cases}$$

Intuitively, $\hat{s}_K$ is a piecewise-constant approximation of s with points of discontinuity determined by the partition P_K.

We now introduce some lemmas that we will combine to prove the theorem. First, we relate the distortion integrals with respect to s and s_K:

Lemma 3.1. *The integral with respect to s_K converges to the integral with respect to s in the following manner:*

$$\lim_{K\to\infty} \int_{\mathbb{R}} s_K^2(c(x))\, w(x)\, f_X(x)\, dx = \int_{\mathbb{R}} s^2(c(x))\, w(x)\, f_X(x)\, dx.$$

Proof. The change of variables $y = c(x)$ yields an alternative form for the LHS integral:

$$\int_{\mathbb{R}} s_K^2(c(x))\, w(x)\, f_X(x)\, dx = \int_0^1 s_K^2(y)\, w(c^{-1}(y))\, p(y)\, dy,$$

where $p(y) = f_X(c^{-1}(y)) \,/\, \lambda(c^{-1}(y))$.

Note that LC3 implies that there exists some $\varepsilon = c(B)$ such that $s(y)$ is decreasing on $y \in (0,\varepsilon)$. Using the inequality $(k+1)/(2K) < k/K$ and the definition of s_K, we can see $s(y/2) \geq s_K(y)$ for all $y \in (0,\varepsilon)$. Using the continuity of s and LC4, we can use the Lebesgue Dominated Convergence Theorem [159, Section 4.4] and $s_K \to s$ as $K \to \infty$ to show

$$\lim_{K\to\infty} \int_0^{\varepsilon} s_K^2(y)\, w(c^{-1}(y))\, p(y)\, dy = \int_0^{\varepsilon} s^2(y)\, w(c^{-1}(y))\, p(y)\, dy. \tag{3.29}$$

Similarly, we can parallel the above proof for $y \in (1-\varepsilon, 1)$ to show

$$\lim_{K\to\infty} \int_{1-\varepsilon}^1 s_K^2(y)\, w(c^{-1}(y))\, p(y)\, dy = \int_{1-\varepsilon}^1 s^2(y)\, w(c^{-1}(y))\, p(y)\, dy. \tag{3.30}$$

Because s is bounded on $[\varepsilon, 1-\varepsilon]$ by LC1,

$$\lim_{K\to\infty} \int_{\varepsilon}^{1-\varepsilon} s_K^2(y)\, \gamma^2(c^{-1}(y))\, p(y)\, dy = \int_{\varepsilon}^{1-\varepsilon} s^2(y)\, \gamma^2(c^{-1}(y))\, p(y)\, dy. \tag{3.31}$$

Combining (3.29)–(3.31) proves the lemma. □

Next we relate quantization error and s_K:

Lemma 3.2. *For large K and $x \in I_{k,K}$, $k \in [1:K]$,*

$$K^2|x - Q_{K,\lambda}(x)| \;\le\; K^2\mu^2(I_{k,K}) \;\le\; s_K^2(c(x)).$$

Proof. The left inequality is trivial. By the mean-value theorem of differentiation, there exists some $v_k \in I_{k,K}$ such that

$$s(c(v_k)) = \frac{c^{-1}(k/K) - c^{-1}((k-1)/K)}{k/K - (k-1)/K} = K\mu(I_{k,K}).$$

Using the definition of s_K yields the right inequality for K large enough such that Condition LC3 ensures s is monotonic in the extremal partitions. □

Finally, we introduce a lemma that relates the truncated source to the integrable form of the distortion:

Lemma 3.3. *The following limit holds:*

$$\lim_{K\to\infty} K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2\, \tilde{h}_K(x)\, dx = \frac{1}{12}\int_{\mathbb{R}} \frac{h(x)}{\lambda^2(x)}\, dx.$$

Proof. We can show that

$$\begin{aligned} K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2\, \tilde{h}_K(x)\, dx &= K^2 \sum_{k=2}^{K-1} \frac{1}{12}\mu(I_{k,K})^3 \frac{1}{A_K(h)\,\mu(I_{k,K})} \int_{I_{k,K}} h(x)\, dx \\ &= \frac{1}{12A_K(h)} \int_{\mathbb{R}} \hat{s}_K^2(c(x))\, h(x)\, dx, \end{aligned}$$

where the first line comes from variance of uniform noise on an interval and the definition of $\tilde{h}_K$, and the second line comes from the definition of $\hat{s}_K$.

From Lemma 3.2, we find $s_K(y)$ dominates $\hat{s}_K(y)$, i.e. $\hat{s}_K(y) \leq s_K(y)$ for $y \in (0,1)$. Using Lemma 3.1, we see $s_K^2(c(x))h(x)$ is Lebesgue integrable. Combining the General Dominated Convergence Theorem [159, Section 4.4] and the fact that $\hat{s}_K \to s$ as $K \to \infty$ for all $y \in (0,1)$,

$$\begin{aligned}\lim_{K\to\infty} \frac{1}{12A_K(h)} \int_{\mathbb{R}} \hat{s}_K^2(c(x))\, h(x)\, dx &= \frac{1}{12} \int_{\mathbb{R}} s^2(c(x))\, h(x)\, dx \\ &= \frac{1}{12} \int_{\mathbb{R}} \frac{h(x)}{\lambda^2(x)}\, dx,\end{aligned}$$

where we use LC2 to ensure the existence of the right-hand side. □

We now prove the main theorem:

Theorem 3.3. *Suppose the source density f_X, weighting function w, and point density λ satisfy Conditions LC1–4. Then*

$$\lim_{K\to\infty} K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2\, w(x)\, f_X(x)\, dx = \frac{1}{12} \int_{\mathbb{R}} \frac{w(x)}{\lambda^2(x)}\, f_X(x)\, dx.$$

Proof. Let $h(x) = w(x)f_X(x)$. We want to show that

$$\lim_{K\to\infty} K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2\, h(x)\, dx = \lim_{K\to\infty} K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2\, \tilde{h}_K(x)\, dx. \tag{3.32}$$

To prove (3.32), we note

$$\begin{aligned}K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2\, |h(x) - \tilde{h}_K(x)|\, dx &\leq \int_{\mathbb{R}} s_K^2(c(x))\, |h(x) - \tilde{h}_K(x)|\, dx \\ &\leq \left(1 + \frac{1}{A_K(h)}\right) \int_{\mathbb{R}} s_K(c(x))^2\, h(x)\, dx \\ &\leq 3 \int_{\mathbb{R}} s_K(c(x))^2\, h(x)\, dx,\end{aligned}$$

where the last inequality holds only for large K since $A_K(h)$ approaches one from above. We also recall $\tilde{h}_K \to h$ as $K \to \infty$ a.e. with respect to μ. Hence, we can again employ the General Lebesgue Dominated Convergence Theorem, this time using the fact $|h(x) - \tilde{h}_K(x)| \leq f(x)$, along with Lemma 3.1 to show (3.32).

To complete the proof of the theorem, we combine Lemma 3.3 and (3.32). □

3.D Generalizing Theorem 3.3

We also need a Linder-style proof to bound the higher-order distortion terms (3.21) and (3.22). Here, we provide only a brief sketch on how to extend Theorem 3.3. Consider the integral

$$K^2 \int_{\mathbb{R}} I^2(x)\, w(x)\, |x - Q_{K,\lambda}(x)|\, f_X(x)\, dx, \tag{3.33}$$

where $I(x) = \mu(I_{k,K})$ if $x \in I_{k,K}$. We can rewrite (3.33) as

$$\int_{\mathbb{R}} \hat{s}_K^2(c(x))\, w(x)\, |x - Q_{K,\lambda}(x)|\, f_X(x)\, dx \leq \frac{1}{K} \int_{\mathbb{R}} s_K^3(x)\, w(x)\, f_X(x)\, dx,$$

where the first line uses the definition of $\hat{s}_K$ and the second uses Lemma 3.2. Ensuring that the right-hand side is integrable is sufficient to show that (3.33) becomes negligible as K becomes large. The success of convergence with K depends on a condition analogous to LC4.

3.E Sources over Finite Support

In [120], DFSQ results are given for quantization of sources over finite support using joint centroid reconstruction. In the above analysis, we rigorously demonstrated how similar performance is achieved by a simpler decoder over sources with infinite support. In this appendix, we briefly argue that a similar technique can be used to bound the distortion of sources over finite support using the simpler decoder. We still use Taylor's theorem, but no longer require tail conditions. As a result, we need fewer conditions and can derive a proof that does not rely on Lebesgue integration. We will only show the univariate case here, but the multivariate case follows easily. The proof is specific to sources that lie on the unit cube $[0, 1]^N$, but we note that sources on more general spaces can be generalized easily.

Assume a companding quantizer with point density λ defined on $[0,1]$. Consider the following conditions on the computation g and the density f_X of the source:

UF1″. The source pdf f_X is continuous and positive on $[0,1]$.

UF2″. The point density λ is continuous and positive on $[0,1]$.

UF3″. The function g is continuous on $[0,1]$ with both $|g'|$ and $|g''|$ defined and bounded by a finite constant $C_{u,K}$.

UF4″. $f_X(x)|g'(x)|^m/\lambda^2(x)$ is Riemann integrable over $[0,1]$ for $m = 0, 1, 2$.

The fMSE induced by a quantizer $Q_{K,\lambda}$ under these conditions is bounded as follows:

Theorem 3.4. *Assume f_X, g, and λ satisfy conditions UF1″–UF4″. Then the fMSE*

$$D_{\text{fmse}}(K,\lambda) = \mathrm{E}\left[|g(X) - g(Q_{K,\lambda}(X))|^2\right]$$

satisfies the following limit:

$$\lim_{K\to\infty} K^2 D_{\text{fmse}}(K,\lambda) = \frac{1}{12}\,\mathrm{E}\left[\left(\frac{\gamma(X)}{\lambda(X)}\right)^2\right].$$

Proof. Here, we require the first mean-value theorem for integrals, which states that for a continuous function $r : [a,b] \to \mathbb{R}$ and integrable function $s : [a,b] \to [0,\infty)$ that does not change sign, there exists a value $x \in [a,b]$ such that

$$\int_a^b r(t)s(t)\,dt = r(x)\int_a^b s(t)\,dt. \tag{3.34}$$

For the case of the companding quantizers, combining this with (2.1) means

$$\frac{1}{K} = \int_{p_k}^{p_{k+1}} \lambda(x)\,dx = \lambda(y_k)(p_{k+1} - p_k) = \lambda(y_k)\Delta_k \tag{3.35}$$

for some $y_k \in (p_k, p_{k+1}]$, where we have defined the kth quantizer cell length $\Delta_k = p_{k+1} - p_k$. The relationship between K, λ, and Δ_k is central to this proof.

Like the proof of Theorem 3.1, we consider the expansion of $D_{\text{fmse}}(K,\lambda)$ by total

expectation:

$$D_{\text{fmse}}(K,\lambda) = \sum_{k=0}^{K-1} \int_{p_k}^{p_{k+1}} |g(x) - g(c_k)|^2 f_X(x)\, dx, \tag{3.36}$$

where we let $p_0 = 0$ and $p_K = 1$ since the support of the source is $[0,1]$. Further expansion of (3.36) using Taylor's theorem yields:

$$\begin{aligned} K^2 D_{\text{fmse}}(K,\lambda) &= K^2 \sum_{k=0}^{K-1} \int_{p_k}^{p_{k+1}} |g'(c_k)(x-c_k) + R_1(x,c_k)|^2 f_X(x)\, dx \\ &\leq \underbrace{K^2 \sum_{k=0}^{K-1} \int_{p_k}^{p_{k+1}} |g'(c_k)|^2 \, |x-c_k|^2 f_X(x)\, dx}_{A} \\ &\quad + \underbrace{K^2 \sum_{k=0}^{K-1} 2 \int_{p_k}^{p_{k+1}} |R_1(x,c_k)| \, |g'(c_k)| \, |x-c_K| f_X(x)\, dx}_{B} \\ &\quad + \underbrace{K^2 \sum_{k=0}^{K-1} \int_{p_k}^{p_{k+1}} R_1(x,c_k)^2 f_X(x)\, dx}_{C}. \end{aligned}$$

Of the three expressions, only term A is meaningful to overall distortion for large K and has the following asymptotic form:

$$\begin{aligned} K^2 \sum_{k=0}^{K-1} \int_{p_k}^{p_{k+1}} |g'(c_k)|^2 \, |x-c_k|^2 f_X(x)\, dx &\overset{(a)}{=} K^2 \sum_{k=0}^{K-1} |g'(c_k)|^2 f_X(v_k) \int_{p_k}^{p_{k+1}} |x-c_k|^2\, dx \\ &\overset{(b)}{=} K^2 \sum_{k=0}^{K-1} |g'(c_k)|^2 f_X(v_k) \frac{\Delta_k^3}{12} \\ &\overset{(c)}{=} \frac{1}{12} \sum_{k=0}^{K-1} f_X(v_k) \left(\frac{|g'(c_k)|^2}{\lambda^2(y_k)} \right) \Delta_k \\ &\xrightarrow{(d)} \frac{1}{12} \int_0^1 \left(\frac{g'(x)}{\lambda(x)} \right)^2 f_X(x)\, dx, \end{aligned}$$

where (a) arises from using (3.34), where v_k is some point in the kth quantizer cell; (b) is evaluation of the integral, recalling (2.3); (c) follows from (3.35); and (d) holds as $K \to \infty$ by the convergence of Riemann rectangles to the integral (assumption UF4″)

The higher-order error terms are negligible using the bound reviewed in (3.15). We show that term B goes to zero and just mention that a similar analysis will bound term C:

$$
\begin{aligned}
K^2 \sum_{k=0}^{K-1} 2 \int_{p_k}^{p_{k+1}} & |R_1(x, c_k)| \, |g'(c_k)| \, |x - c_K| f_X(x) \, dx \\
& \overset{(a)}{\leq} K^2 \overline{\Delta} \sum_{k=0}^{K-1} C_{u,K} \Delta_k |g'(c_k)| \int_{p_k}^{p_{k+1}} |x - c_k| f_X(x) \, dx \\
& \overset{(b)}{=} K^2 \overline{\Delta} \sum_{k=0}^{K-1} C_{u,K} \Delta_k |g'(c_k)| f_X(v_k) \frac{\Delta_k^2}{4} \\
& \overset{(c)}{=} C_u \overline{\Delta} \sum_{k=1}^{K-2} f_X(v_k) \frac{|g'(c_k)|}{\lambda^2(y_k)} \Delta_k \\
& \overset{(d)}{\longrightarrow} 0,
\end{aligned}
$$

where (a) follows from bounding $R_1(x, c_k)$ using (3.15) and by defining $\overline{\Delta}$ as the longest quantizer interval length; (b) arises from using (3.34); (c) follows from (3.35); and (d) holds as $K \to \infty$ by the convergence of Riemann rectangles to the integral (assumption UF3′) and the vanishing of $\overline{\Delta}$. Hence, the distortion contribution becomes negligible as K increases. Note that expansion term C will scale with $\overline{\Delta}^2$ with growing codebook size and becomes negligible faster.

□

Chapter 4

Performance Loss of DFSQ

Distributed functional scalar quantization (DFSQ) was proposed as a framework to characterize the performance of lossy compression in a distributed setting for when a computation is desired at the fusion center, addressing practical concerns such as code block length, buffer size and decoder complexity. This is useful for goals such as parameter estimation, coding for perception, and information extraction, which are common applications for distributed systems. In general, high-resolution quantization theory is useful in determining closed-form performance expressions for a wide range of source models and fidelity criteria in situations where the Shannon rate–distortion function (2.22) is difficult to compute.

Having formed a rigorous foundation for DFSQ theory from Chapters 2 and 3, we now try to understand the performance gap between this achievable scheme and the fundamental limits of compression using large-blocklength analysis. Our hope is to demonstrate that this gap is small, which solidifies the usefulness of DFSQ. The fast convergence of the performance of scalar quantizers at finite rates to the low-distortion asymptotic underlies the importance of this theory to real-world applications, a point highlighted by Berger and Gibson in their survey of lossy source coding [12]. In a network scenario, the importance of constraints to block length and coding complexity is magnified due to resource limitations in practical implementations.

We begin with some motivation why the performance gap between DFSQ and the rate–distortion outer bound is expected to be small in Section 4.1. We then provide

the main results in Section 4.2 and 4.3. Finally, we conclude the chapter in Section 4.4 with some suggestions for future work.

4.1 Motivation

We begin by reviewing a well-known result from lossy source coding of memoryless Gaussian sources. Consider a discrete-time continuous-valued source $X \sim \mathcal{N}(0,1)$. Using the high-resolution analysis, the distortion–rate trade-off from applying an optimal nonuniform fixed-rate scalar quantizer (FRSQ) on this source is asymptotically

$$D^*_{\mathrm{mse,fr}}(R) \simeq \frac{\pi\sqrt{3}}{2} 2^{-2R}$$

using (2.8). When entropy-constrained scalar quantization (ECSQ) is used on the same source, the distortion is asymptotically

$$D^*_{\mathrm{mse,ec}}(R) \simeq \frac{\pi e}{6} 2^{-2R}$$

using the analogue of (2.10) with unnormalized point densities [56, 57]. Finally, the fundamental limit for a Gaussian source is

$$\delta(R) = 2^{-2R}$$

for any choice of R (note this is asymptotic in the block length L, not rate R).[1] Hence, the performance of ECSQ is simply 1.53 dB more or 0.255 bits/sample less than the optimal coding strategy for high rate, or large R. The loss from FRSQ is more, and the performance gap for both settings is well-approximated by a constant as R increases.

It turns out that the 0.255 bits/sample rate gap is true for most sources when the rate is large due to the tightness of the Shannon lower bound in the low-distortion limit

[1] A Gaussian source is a rare instance where the Shannon distortion–rate function has a clean form.

(see Section 2.5.2). This specifies the gap between scalar quantization with lossless entropy coding, which is easy to implement in practice, and vector quantization with lossy entropy coding, which is the fundamental limit. In this chapter, we establish similar Shannon-theoretic rate–distortion results for fMSE and determine the rate loss that would be incurred by using the simple implementation of DFSQ for small distortion. We find the gap is small under many scenarios.

4.2 Rate–Distortion Behavior of Distributed Systems for W-MSE

We first analyze the rate–distortion behavior under the W-MSE measure (2.25):

$$d(x_1^N, \hat{x}_1^N) = \left\| W(x_1^N)(x_1^N - \hat{x}_1^N) \right\|^2.$$

We consider the setting where N scalar sources are separately encoded in an architecture similar to Figure 2-5. Note that this is different from the the setup of [101], which used joint coding of an N-dimensional source. In this new setting, we restrict attention to a subclass of the distortion measure where the input-dependent weighting matrix W is nonnegative and diagonal, such that the nth diagonal entry of W depends only on X_n. The eventual result of this section, Theorem 4.1, will then provide the bridge between prior work on nondifference distortion measures and fMSE.

We impose the following regularity conditions on the joint density of X_1^N and the input-weighted matrix W:

WC1. $W(x_1^L)$ is strictly nonnegative for all inputs x_1^L and equal to zero at most on a set of Jordan measure zero.

WC2. $W(x_1^N)$ is diagonal and the nth diagonal entry, denoted $w_n(x_1^N)$, depends only on x_n.

WC3. $W(x_1^N)$ is piecewise continuously differentiable with respect to x_1^N.

WC4. X_1^N have finite differential entropy and second moment.

WC5. For all $n \in [1:N]$,

$$\mathrm{E}\left[\log w_n(X_1^N)\right] < \infty.$$

WC6.

$$\mathrm{E}\left[\left\{\mathrm{tr}(W^{-1}(X_1^N))\right\}^3\right] < \infty.$$

Then the minimum sum rate of the achievable rate region is characterized asymptotically by the following theorem:

Theorem 4.1. *Consider N correlated sources that are encoded separately and an allowed W-MSE distortion D. If the joint source density $f_{X_1^N}$ and input-dependent weighting matrix W satisfy Conditions WC1–WC6, then the minimum sum rate r_{sum} of the achievable rate region under the W-MSE distortion satisfies:*

$$\lim_{D\to 0}\left(r_{\mathrm{sum}}(D) + \frac{N}{2}\log(2\pi e D/N)\right) = h(X_1^N) + \sum_{n=1}^{N}\mathrm{E}[\log w_n(X_n)], \tag{4.1}$$

Proof. The outer bound follows directly from (2.26), which requires Conditions WC3–WC6 hold [101].[2] We now devise an achievability scheme and sketch out a proof to confirm that the outer bound becomes tight in the low-distortion limit. We emphasize that the encoders of the N sources are separate and can use long blocks in this setup.

Let us define monotonic scalar functions $\widetilde{g}_n$ satisfying $\widetilde{g}_n'(x) = w_n(x)$. Since $w_n(x)$ is nonnegative, $\widetilde{g}_n$ always exists and is unique up to a constant offset. Moreover, since $w_n(x)$ is positive except on a set of Jordan measure zero (Condition WC1), $\widetilde{g}_n$ is strictly increasing.

Consider a set of transformed sources $\{Y_n\}_{n=1}^N$, defined as $Y_n = \widetilde{g}_n(X_n)$. In general, the set $\{Y_n\}_{n=1}^N$ is correlated if $\{X_n\}_{n=1}^N$ is correlated. We now solve a modified problem, which is to determine an achievable rate region for distributed compression of Y_1^N under the MSE fidelity criterion. Using the direct-coding proof of Zamir and

[2]Conditions WC5 and WC6 are hard to intuitively explain but are necessary for the proofs in [101].

Berger [207, Theorem 2], we see that the sum rate can asymptotically achieve

$$R(D_1^N) \lesssim h(Y_1^N) - \frac{1}{2}\sum_{n=1}^{N} \log(2\pi e D_n) + \varepsilon(D_1^N),$$

where D_n is the allowable distortion between Y_n and $\widehat{Y}_n$, and $\varepsilon(D_1^N) \to 0$ as $D_1^N \to 0$. The notation $\lesssim$ indicates the ratio of the left and right expressions is asymptotically less than one as D decreases.

Using Taylor's theorem to expand $\widetilde{g}_i$ around X_i, we can approximate the MSE of compressing Y_n as:

$$\begin{aligned} D_n(Y_n, \widehat{Y}_n) &= \mathrm{E}[|Y_n - \widehat{Y}_n|^2] \\ &= \mathrm{E}[|\widetilde{g}_n(X_n) - \widetilde{g}_n(\widehat{X}_n)|^2] \\ &\approx \mathrm{E}[|\widetilde{g}_n'(X_n)|^2 (X_n - \widehat{X}_n)^2]. \end{aligned}$$

We see that this becomes equivalent to the fidelity requirements of the W-MSE distortion as the distortion decreases. Setting $D_n = D/N$ for every n yields the achievable sum rate

$$R_{\mathrm{sum}}(D) \lesssim h(Y_1^N) - \frac{N}{2}\log(2\pi e D/N) + \varepsilon(D_1^N).$$

In general, if $Y_1^N = u(X_1^N)$, then

$$h(Y_1^N) \leq h(X_1^N) + \mathrm{E}[\log |\partial u / \partial X_1^N|]$$

with equality if and only if u is bijective [11]. Since each $\widetilde{g}_i$ is strictly increasing and hence invertible, the equality holds for this choice of u and we can rewrite the asymptotic achievable rate as

$$R_{\mathrm{sum}}(D) \lesssim h(X_1^N) - \frac{L}{2}\log(2\pi e D/N) + \sum_{n=1}^{N} \mathrm{E}[\log w_n(X_n)] + \varepsilon(D_1^N).$$

This matches the outer bound as D becomes small, providing a tight characterization of the sum rate $\square$

Presented here is only the sum-rate characterization of the rate–distortion region. As in all multiterminal source coding problems, we must also bound all individual rates as well as the sums of all subsets of rates. We do not present these calculations here because they do not contribute understanding to sum rate, but they are easily found following the same analysis.

The achievability scheme of transforming the source before applying Slepian–Wolf coding is highly reminiscent of companding quantizers. Hence, one nice interpretation is that fine nonuniform quantization followed by block coding can achieve the optimal sum rate in a distributed setting for the W-MSE distortion with diagonal weighting matrices, provided that the target distortions for all sources are small.

4.3 Rate Loss of DFSQ

We recall the fidelity criterion of interest is fMSE as defined in (2.11) and the performance of nonuniform scalar quantizers is given in (2.17), (2.20) and (2.21). In this section, we will understand the achievable rate region with respect to fMSE using Theorem 4.1 and compare the sum rate to the DFSQ results; this determines the rate loss from using scalar quantizers.

We require the following conditions on the source density $f_{X_1^N}$ and computation g, recalling $g_n(x)$ denotes the partial derivative of g with respect to its nth argument evaluated at the point x.

FC1. For all $n \in [1 : N]$, $g_n(x_1^N)$ is nonzero except on a set of Jordan measure zero.

FC2. $g(x_1^N)$ is twice differentiable with respect to x_1^N.

FC3. X_1^N have finite differential entropy and second moment.

FC4. For all $n \in [1 : N]$,

$$\mathrm{E}[\log(|g_n(X_1^N)|^2)] < \infty.$$

FC5.

$$\mathrm{E}\left[\left\{\sum_{n=1}^{N}\left|g_n^{-2}(X_1^N)\right|\right\}^3\right] < \infty.$$

Then the fMSE rate–distortion function is characterized by the following theorem

Theorem 4.2. *Consider N correlated sources that are encoded separately and an allowed fMSE distortion D. If the joint source density $f_{X_1^N}$ and computation g satisfy Conditions FC1–FC5, then the minimum sum rate r_{sum} of the rate–distortion region under the fMSE satisfies:*

$$\lim_{D\to 0}\left(r_{\mathrm{sum}}(D)+\frac{N}{2}\log(2\pi eD/N)\right) = h(X_1^N)+\sum_{n=1}^{N}\mathrm{E}[\log\gamma_n(X_n)],$$

where γ_n is defined in (2.13)*.*

Proof. In the first step of the proof sketch, we demonstrate that the fMSE can be approximated with a W-MSE distortion in the small-distortion regime. Using Taylor's theorem, we rewrite fMSE as:

$$\mathrm{E}[|g(X_1^N)-g(\widehat{X}_1^N)|^2] = \mathrm{E}\left[\left|\sum_{n=1}^{N} g_n(X_1^N)(X_n-\widehat{X}_n)+R_1(X_n-\widehat{X}_n)\right|^2\right],$$

where $R_1(x)$ is a remainder term of order $\mathcal{O}(|x-\hat{x}|^2)$. As the distortion decreases, the terms in the product that contain R_1 decay quickly and become negligible in the small-distortion limit (see [176]). Moreover, all terms containing $(X_i-\widehat{X}_i)(X_j-\widehat{X}_j)$ for $i\neq j$ also become negligible as quantization errors become white in the high-resolution limit [189]. Hence, the resulting fMSE distortion can be simplified to

$$\begin{aligned}\mathrm{E}[|g(X_1^N)-g(\widehat{X}_1^N)|^2] &\approx \sum_{n=1}^{N}\mathrm{E}\left[|g_n(X_1^N)|^2(X_n-\widehat{X}_n)^2\right]\\ &= \sum_{n=1}^{N}\mathrm{E}\left[\gamma_n^2(X_n)(X_n-\widehat{X}_n)^2\right],\end{aligned}$$

where the relative error in the approximation can be made arbitrarily small with increasing codebook size. We see that this distortion corresponds to the diagonal

W-MSE distortion studied in Section 4.2. In particular, the nth diagonal entry of the weighting matrix W is $\gamma_n^2(X_n)$, ensuring it is nonnegative. Conditions FC1–FC4 then satisfy the sufficient conditions for Theorem 4.1, thereby giving the minimum sum rate of the achievable rate region. □

We can now determine the gap between the minimum sum rate using nonuniform scalar quantization followed by Slepian–Wolf coding (SWSQ), as described in (2.21), and the optimum solution shown in Theorem 4.2:

Corollary 4.1. *The sum rate from using* optimal *nonuniform scalar quantization followed by Slepian–Wolf coding becomes exactly* $(N/2)\log(\pi e/6)$ *bits worse than the sum rate* r_{sum} *of the fMSE rate–distortion function in the limit of small distortion. This corresponds to a rate loss of 0.255 bits/sample, which matches the well-known result in the point-to-point setting.*

We can also bound the rate loss of ECSQ:

Corollary 4.2. *The sum rate from using* optimal *nonuniform scalar quantization followed by entropy coding becomes exactly*

$$(N/2)\log(\pi e/6) + I(X_1;\ldots;X_N) \text{ bits},$$

worse than the sum rate r_{sum} *of the fMSE rate–distortion function in the limit of small distortion, where* $I(X_1;\ldots;X_N)$ *denotes the multivariate mutual information, or the multiinformation [171].*

We can similarly derive a rate gap for FRSQ, but the closed-form expression is less intuitive because of dependencies on the source and computation.

4.3.1 Remarks

1. Conditions FC1–FC4 and MF1–MF3 are not the most general that can be derived. In particular, it is possible to generalize the results to infinite-support densities provided the differential entropy is finite. The precise conditions will parallel the analysis in Chapter 3.

2. The results in this section match existing results when the problem is constrained in certain ways. For example, when sources are independent, the problem reduces to the point-to-point SLB with a scalar W-MSE distortion that is a function of the functional sensitivity profile. When the computation is linear, i.e. the functional sensitivity profile is flat, the model reduces to the case studied by Zamir and Berger [207]. When $L = 1$, the result is the scalar W-MSE result [101].

3. Although the gap between DFSQ and the rate–distortion function can be small when the correlation between the quantized observations is coded properly, the multiinformation gap can be quite large. Proper classification of this gap between ECSQ and SWSQ helps motivate engineering choices when designing sensor networks.

4. In (2.17), (2.20) and (2.21), the optimal sum rate is achieved using bit allocation. For a given distortion, there usually exists a unique point in the achievable region where the sum rate is minimized, meaning this is unique set of rates that will achieve the rate gap. This is in comparison to the Slepian–Wolf result, where the minimum sum rate can be achieved by many sets of rates.

4.4 Future Work

In this chapter we have characterized the sum-rate gap between coded scalar quantization and the achievable rate region when the fidelity criterion is fMSE. This result is especially encouraging in the distributed setting because it demonstrates that simple compression structures are efficient, even with strict latency constraints and power considerations. This bridge between Shannon source coding theory and high-resolution quantization theory can impact compression design for general network architectures and fidelity criteria. In particular, it motivates the need to design efficient entropy codes and Slepian–Wolf codes on top of scalar-quantizer outputs.

It may also be useful to compare DFSQ to other known achievable results that use coding over long blocks for specific models where the analysis is tractable, e.g. quadratic Gaussian [201]. Additionally, it is interesting to consider extensions to remote sources [42].

In Chapter 5, we demonstrate how intersensor interaction can change the design principles in distributed networks operating in the small blocklength regime [174, 175]. One exciting line of research is to determine the rate-loss bounds for this generalization, building off rate–distortion results in [86]. Such "chatting" networks greatly expand the possibilities of network topologies, especially when the cost of communicating a bit may vary in the network.

Chapter 5

Chatting Networks

One interesting message in Misra *et al.* [120] is that a small amount of interaction between two quantizers in the DFSQ framework can yield substantial gains. This small technical result has influenced the work in this chapter, which explores how "chatting" can affect the compression of a large network. The results here have implications in the design of sensors in networks when there are opportunities to leverage different communication protocols that may have varying costs of communication and bandwidth constraints. It also combats the optimistic intuition of the Slepian–Wolf theorem in practical settings where the result does not hold. We summarize the key motivations of this work in Section 5.1 and introduce some preliminaries in Section 5.2. In Sections 5.3 and 5.4, we describe the problem model and provide the main performance results respectively. We then expand on joint optimization of heterogeneous systems in Section 5.5 and provide a detailed example to provide intuition of our contributions in Section 5.6.

Parts of this work were discussed in [174] and [175].

5.1 Motivation

A longstanding consideration in distributed compression systems is whether sensors wishing to convey information to a fusion center should communicate with each other to improve efficiency. Architectures that only allow communication between indi-

vidual sensors and the fusion center simplify the network's communication protocol and decrease sensor responsibilities. Moreover, information theoretic results such as the Slepian–Wolf theorem (2.23) demonstrate distributed compression can perform as well as joint compression for lossless communication of correlated information sources [168]. Although this surprising and beautiful result does not extend fully, comparable results for lossy coding show that the rate loss from separate encoding can be small using Berger–Tung coding (see, e.g., [206]), again suggesting communication between sensors has little or no utility.

Although it is tempting to use results from information theory to justify simple communication topologies, it is important to note the Slepian–Wolf result relies on coding over long blocks; in the finite-blocklength regime, the optimality of distributed encoding does not hold [178]. Moreover, investigations on lossy extensions of the Slepian–Wolf theorem usually focus on compression fidelity of the source amplitudes rather than more general computations that are of interest in practical sensor networks, where separation may be suboptimal.

This chapter examines the use of communication among sensors when the block length is one, a regime where intersensor collaboration, or *chatting*, can greatly decrease the aggregate communication from sensors to the fusion center to meet a distortion criterion as compared to a strictly distributed network. This is especially true when the fusion center's objective is to compute a function of the sources with high fidelity rather than determine the sources themselves. We analyze these networks using the distributed functional scalar quantization (DFSQ) framework, as introduced in Section 2.3. In our problem model (Figure 1-1), N correlated but memoryless continuous-valued, discrete-time stochastic processes produce scalar realizations $X_1^N(t) = (X_1(t), \ldots, X_N(t))$ for $t \in \mathbb{Z}$. For each t, realizations of these sources are scalar quantized by sensors and transmitted to a fusion center at rates R_1^N. To aid this communication, sensors can collaborate with each other via a side channel that is unobservable to the fusion center.

The side channel facilitating intersensor communication has practical implications. In typical communication systems, the transmission power needed for reliable com-

munication increases superlinearly with distance and bandwidth [181]. Hence, it is much cheaper to design short and low-rate links between sensors than reliable and high-rate links to a fusion center. Moreover, milder transmission requirements provide more flexibility in determining the transmission media or communication modalities employed, which can allow intersensor communication to be orthogonal to the main network. One such example is cognitive radio, a paradigm where the wireless spectrum can have secondary users that communicate only when the primary users are silent [203]. This means secondary users have less priority and hence lower reliability and rate, which is adequate for intersensor communication.

The main contributions of this work are to precisely characterize the distortion performance of a distributed network when chatting is allowed and to identify the optimal quantizer design for each sensor. We show that collaboration can have significant impact on performance; in some cases, it can dramatically reduce distortion even when the chatting has extremely low rate. We also give necessary conditions on the chatting topology and protocol for successful decodability in the DFSQ framework, thus providing insight into the architecture design for chatting networks. Finally, we recognize that intersensor communication can occur on low-cost channels and solve the rate allocation problem in networks with heterogeneous links and different costs of transmission.

5.2 Preliminaries

5.2.1 Prior Work

The notion of collaborating sensors or agents has been explored in the information theory, distributed estimation and social learning communities. We will provide a cursory summary of some of these works that have inspired the chapter.

Beyond the celebrated Slepian–Wolf result, there are several interesting formulations of chatting. Kaspi and Berger provided inner bounds for the rate region of a two-encoder problem where one encoder can send information for the other us-

ing compress-and-forward techniques [86]. Recently, this bound has been generalized in [160], but the exact rate region is still unknown except in special cases. Chatting is related to source coding problems such as interaction [106,107], omniscience [130] and data exchange [24]. However, these settings are more naturally suited for discrete-alphabet sources and existing results rely on large-blocklength analysis. There is also an interesting relationship between chatting and specific subclasses of multidimensional companding quantizers [102] where the vector quantizers are Cartesian products of scalar partitions.

The structure of chatting also arises in the context of distributed detection and estimation, especially for consensus and gossip problems. Here, multiple rounds of interactions have been studied in the presence of quantization and noise [5, 38, 83, 84, 126, 202]. More recently, new models incorporate quantization or noise in estimation parameters for group or sequential decision-making [155, 156, 184].

5.2.2 Chatting in DFSQ

In [120, Section VIII], chatting is introduced in the setting where one sensor sends exactly one bit to another sensor (see Figure 5-2a). Under fixed-rate quantization, this collaboration can at most decrease the distortion by a factor of four using a property of $\mathcal{L}_{1/3}$ quasi-norms.[1] Because using that bit to send additional information to the fusion center would decrease distortion by exactly a factor of four, this is considered a negative result. Here, there is an implicit assumption that links have equal cost per bit and the network wishes to optimize a total cost budget. In the entropy-constrained setting, chatting may be useful even when links have equal costs. One example was given to demonstrate a single bit of chatting can decrease the distortion arbitrarily close to zero by changing model parameters; more generally, the benefit of chatting varies depending on the source joint distribution and decoder computation.

In previous work, there is no systematic theory on performance and quantizer design of chatting. Moreover, collaboration in larger networks was still an open

[1] $\mathcal{L}_p$ as defined in Chapter 2 does not satisfy the triangle inequality for $0 < p < 1$. Hence it is a quasi-norm rather than a norm.

problem. We extend previous results and provide a more complete discussion on how a chatting channel affects a distributed quantization network. A sample result is that chatting can be beneficial in the fixed-rate setting if the cost of communicating a bit to another sensor is lower than the cost of communicating a bit to the fusion center.

5.2.3 Don't-Care Intervals in DFSQ

When the computation induces the functional sensitivity profile to be zero on some subintervals of the support, the high-resolution assumptions are violated and the asymptotic distortion performance may not be described by (2.14). This issue is addressed by carefully coding when the source is in such a "don't-care" interval [120, Section VII] and then applying high-resolution theory to the remaining support. This consideration is particularly relevant because chatting among sensors can often induce the functional sensitivity profile to be zero, and proper coding can lead to greatly improved performance.

Consider Λ_n don't-care intervals in γ_n and let A_n be the event that the source realization x_n is *not* in the union of them. In the fixed-rate setting, one codeword is allocated to each don't-care interval, and the remaining $K_n - \Lambda_n$ codewords are used to form reconstruction points in the nonzero intervals of the scalar quantizer. There is a small degradation in performance from the loss corresponding to Λ_n, but this quickly becomes negligible as K_n increases. In the entropy-constrained case, the additional flexibility in coding allows for the encoder to split its message and reduce cost. The first part is an indicator variable revealing whether the source is in a don't-care interval and can be coded at rate $I_A \triangleq H_B(\Pr(A_n))$, where H_B is the binary entropy function. The actual reconstruction message is only sent if event A_n occurs, and its rate is amplified to $(R_n - I_A)/\Pr(A_n)$ to meet the average rate constraint. The multiplicative factor $1/\Pr(A_n)$ is called the *rate amplification*.

5.3 Problem Model

We begin by defining the capabilities and restrictions of the sensors and fusion center in Figure 1-1, recalling that the N sensors observe realizations from correlated and memoryless sources. The quantizer at each sensor is scalar, meaning its output depends on the current observation. In this model, we assume that the quantizer's mapping can also be affected by information it receives from other sensors via the chatting channel, but that information is limited to their current observations as well. Because there is no intertemporal communication, we remove the time index and model the sources as being drawn from a joint distribution $f_{X_1^N}$ at each t. We first describe the notation used to model the chatting channel, then summarize what each sensor is allowed to communicate and finally conclude the section with a simple example.

We model the chatting channel in Figure 1-1 as a directed graph $\mathcal{G}^c = (\mathcal{V}, \mathcal{E})$, where the set of nodes $\mathcal{V}$ is the set of all sensors and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of noiseless, directed chatting links (Figure 5-1). If $(i, n) \in \mathcal{E}$, then for each source realization, Sensor i sends to Sensor n a chatting message $M_{i\to n}$ with codebook size $K_{i\to n}$. The parent and children sets of a sensor $n \in \mathcal{V}$ are denoted $\mathcal{N}_p(n)$ and $\mathcal{N}_c(n)$ respectively; when $(i, n) \in \mathcal{E}$, i is a parent of n and n is a child of i. The set of all chatting messages is $M^c = \{M_{i\to n}\}_{(i,n)\in\mathcal{E}}$ and the set of corresponding codebook sizes is $K^c = \{K_{i\to n}\}_{(i,n)\in\mathcal{E}}$. Modeling the chatting channel as a graph becomes useful later when we analyze the topology of intersensor collaboration for successful communication with the fusion center.

The chatting messages are communicated in sequential order according to a schedule that the sensors and the fusion center know in advance; the set of chatting messages M^c can therefore also be thought of as a sequence. This communication occurs quickly in that all chatting is completed before the next discrete time instant, at which point new realizations of X_1^N are measured. We assume that each chatting link can support one message per source realization and an outgoing chatting message from Sensor n can only depend on X_n and the chatting messages received from the sensor's

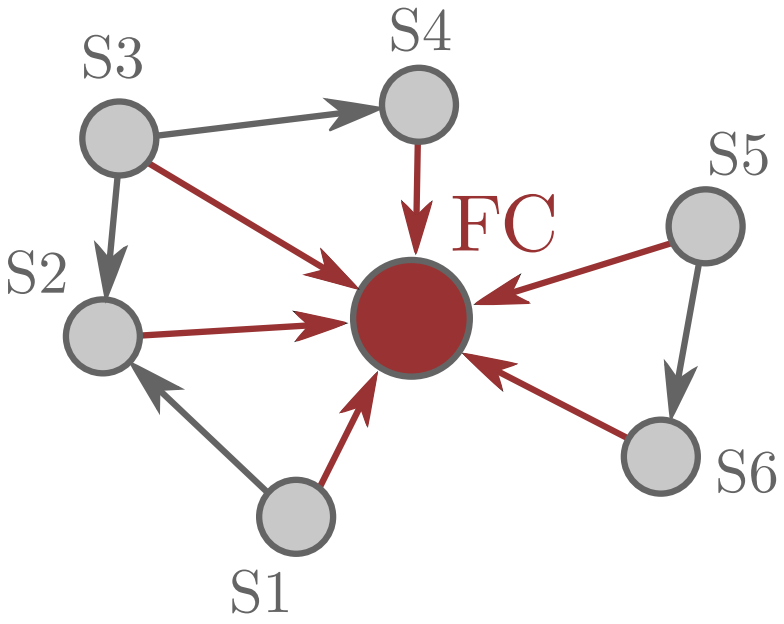
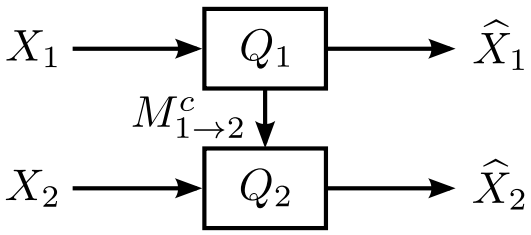


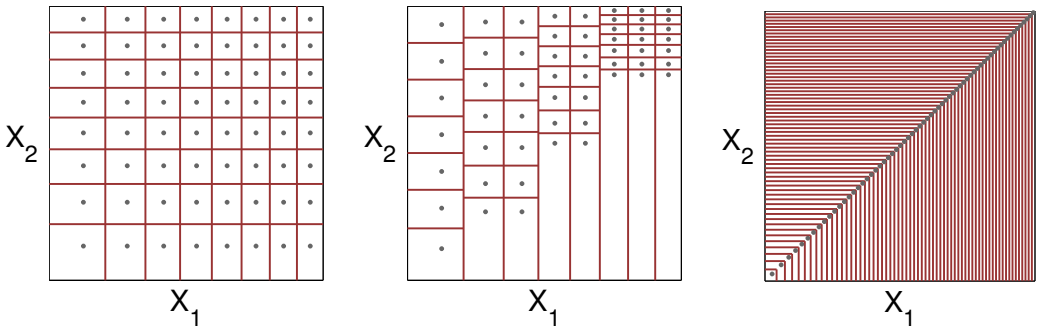
Figure 5-1: A stylized example of the communication and chatting links between sensors (S) and the fusion center (FC). Only the gray lines, corresponding to chatting links, are included in the edge set of $\mathcal{G}^c$.

parent set $\mathcal{N}_p(n)$. After chatting is complete, Sensor n compresses its observation X_n into a message M_n using a codebook dependent on the information gathered from chatting messages, which is then noiselessly communicated to the fusion center. In the most general setting, both the codebook mapping and size, i.e. communication rate, may depend on incoming chatting messages. The fusion center then estimates the computation $g(X_1^N)$ using M_1^N but not M^c. While the above treatment of the chatting channel is very general, the final assumption that the fusion center cannot observe chatting messages directly restricts the type of communication can occur. In the next section, we will discuss in greater detail what type of communication schedules are allowed to optimize compression performance.

We now present a simple example based on a two-sensor network shown in Figure 5-2a. Here, the computation of interest is $Y = \max(X_1, X_2)$ and two sensors compress iid uniform random variables X_1 and X_2 using fixed-rate scalar quantizers. Naïvely, one may predict that the best scalar quantizer should be uniform using (2.5). However, larger amplitudes are more likely to be meaningful for the max computation and the best fixed-rate scalar quantizer is found using (2.15); the quantizer mapping



(a)



(b) Distributed (c) Chatting (d) Joint encoding

Figure 5-2: A two-sensor network demonstrating chatting. In this example, Sensor 1 communicates some information about its realization to Sensor 2 with some chatting rate R_c, as illustrated in **(a)**. The quantizer mappings for the **(b)** distributed, **(c)** chatting, and **(d)** joint-encoding cases are shown.

is shown in Figure 5-2b. If the chatting channel has very high rate, then Sensor 2 can effectively compute and compress Y and achieve joint-encoding performance, as shown in Figure 5-2d. The chatting described in this chapter falls between the distributed and joint-encoding scenarios; the main regime of interest is when the chatting channel has low rate, e.g. 2 bits/sample in Figure 5-2c. We will revisit this example in a more general setting later in the chapter and show how rate allocation, and chatting codebook optimization affects compression performance. Note that the example uses independent sources for simplicity, but the same analysis can account for correlated sources. The correlation will influence the functional sensitivity profile, which alters the shape of the companding quantizers.

5.4 Performance and Design of Chatting Networks

With the chatting model explicit, we present performance of $Q_{K_1^N,\lambda_1^N}$ in the fixed-rate and entropy-constrained settings. We then show how to optimize λ_1^N given K_1^N and K^c to minimize fMSE. We first analyze the network assuming the chatting graph is acyclic. Later, we will show this condition on the graph is sufficient for the fusion center to successfully infer the codebook used by each sensor and hence recover the quantized values from messages M_1^N. The graph structure provides the schedule of transmission on the chatting channel, i.e. a sensor transmits its chatting messages to its children only when it receives its parents' chatting messages.

Before studying fMSE, we need to extend the definition of the functional sensitivity profile.

Definition 5.1. Let $\mathcal{N}_p(n) \subseteq V$ be the set of parents of Sensor n in the graph $\mathcal{G}^c$ induced by chatting. The nth *conditional functional sensitivity profile*, or conditional sensitivity for short, of computation g given all chatting messages M^c is

$$\gamma_{n|M^c}(x|m) = \left(\mathrm{E}\left[|g_n(X_1^N)|^2 \,\middle|\, X_n = x, M_{i\to n} = m_{i\to n} \text{ for all } i \in \mathcal{N}_p(n)\right]\right)^{1/2}.$$

Notice only messages from parent sensors are relevant to $\gamma_{n|M^c}$. Intuitively, chatting messages reveal information about the parent sensors' quantized values and reshape the functional sensitivity profile appropriately. Depending on the encoding of chatting messages, this may induce don't-care intervals in the conditional sensitivity (where $\gamma_{n|M^c} = 0$).

The distortion's dependence on the number of codeword points and the conditional sensitivity is given in the following theorem:

Theorem 5.1. *Given the source distribution $f_{X_1^N}$, computation g, and point densities $\lambda_1^N(M^c)$ satisfying conditions MF1–3 for every possible realization of M^c, the asymptotic distortion of the conditional expectation decoder* (2.12) *given codeword allocation*

K_1^N and K^c is

$$D_{\text{fmse}}(K_1^N, K^c, \lambda_1^N) \simeq \mathrm{E}_{M^c}\left[\sum_{n=1}^{N} \mathrm{E}_{X_n|M^c}\left[\frac{1}{12K_n^2(m)}\frac{\gamma_{n|M^c}^2(X_n|m)}{\lambda_{n|M^c}^2(X_n|m)} \;\middle|\; M^c = m\right]\right].$$

Proof. Extend the proof of [120, Theorem 17] using the Law of Total Expectation. Note that the chatting codebook is assumed fixed and known to all sensors and the fusion center in this formulation. □

Compared to the DFSQ result, the performance of a chatting network can be substantially more difficult to compute since the conditional sensitivity may be different with each realization of M^c and affects the choice of the point density and codebook size. However, Sensor n's dependence on M^c is through a subset of messages from its parent nodes. In Section 5.6, we will see how structured architectures lead to tractable computations of fMSE. Following the techniques in Chapter 3, the theorem can be expanded to account for infinite-support distributions and a simpler decoder. Some effort is necessary to justify the use of normalized point densities in the infinite-support case, especially in the entropy-constrained setting, but high-resolution theory applies in this case as well.

5.4.1 Don't-Care Intervals

We have already alluded to the fact that chatting can induce don't-care intervals in the conditional sensitivity of certain sensors. In this case, we must properly code for these intervals to ensure the high-resolution assumptions hold, as discussed in Section 5.2.

For fixed-rate coding where $R_n = \log(K_n)$, this means shifting one codeword to the interior of each don't-care interval and applying standard high-resolution analysis over the union of all intervals where $\gamma_n(x) > 0$. The resulting distortion of a chatting network is then given as:

Corollary 5.1. *Assume the source distribution $f_{X_1^N}$, computation g, and point densities $\lambda_1^N(M^c)$ satisfying conditions MF1–3 for every possible realization of M^c, with*

the additional requirement that $\lambda_n(x \,|\, m) = 0$ *whenever* $\gamma_{n|M^c}(x \,|\, m) = 0$. *Let* $\Lambda_n(m)$ *be the number of don't-care intervals in the conditional sensitivity of Sensor* n *when* $M^c = m$. *The asymptotic distortion of such a chatting network where communication links use fixed-rate coding is*

$$D_{\text{fmse}}(R_1^N, K^c, \lambda_1^N) \simeq \mathrm{E}_{M^c}\left[\sum_{n=1}^{N} \mathrm{E}_{X_n|M^c}\left[\frac{1}{12(K_n')^2}\frac{\gamma_{n|M^c}^2(X_n \,|\, m)}{\lambda_{n|M^c}^2(X_n \,|\, m)} \;\middle|\; M^c = m\right]\right],$$

where $K_n' = 2^{R_n} - \Lambda_n(m)$.

In the entropy-constrained setting where $R_n = H(\widehat{X}_n)$, we must code first the event $A_n(m)$ that the source is not in a don't-care interval given the chatting messages, and then code the source realization only if A_n occurs. The resulting distortion of a chatting network is:

Corollary 5.2. *Assume the source distribution* $f_{X_1^N}$, *computation* g, *and point densities* $\lambda_1^N(M^c)$ *satisfying conditions MF1–3 for every possible realization of* M^c, *with the additional requirement that* $\lambda_n(x \,|\, m) = 0$ *whenever* $\gamma_{n|M^c}(x \,|\, m) = 0$. *Let* $A_n(m)$ *be the event that* X_n *is not in a don't-care interval given* $M^c = m$. *The asymptotic distortion of such a chatting network where communication links use entropy coding is*

$$D_{\text{fmse}}(R_1^N, K^c, \lambda_1^N) \simeq \mathrm{E}_{M^c}\left[\sum_{n=1}^{N} \mathrm{E}_{X_n|M^c}\left[\frac{\Pr(A_n(m))}{12} 2^{2h(X_n|A_n(m)) + 2\,\mathrm{E}[\log \lambda_n(X_n)|A_n(m)]} \cdot \frac{\gamma_{n|M^c}^2(X_n \,|\, m)}{\lambda_{n|M^c}^2(X_n \,|\, m)} 2^{-2(R_n(m) - H_B(A_n(m)))/P(A_n(m))} \;\middle|\; M^c = m\right]\right].$$

We will use both corollaries in optimizing the design of $\lambda_1^N(M^c)$ in the remainder of the chapter.

5.4.2 Fixed-Rate Quantization Design

We mirror the method used to determine (2.15) in the DFSQ setup but now allow the sensor to choose from a set of codebooks depending on the incoming messages from

parent sensors. The mapping between chatting messages and codebooks is known to the decoder of the fusion center, and each codebook corresponds to the optimal quantizer for a given conditional sensitivity induced by the incoming message. Let $Z_n(M^c)$ be the union of the don't-care intervals of a particular conditional sensitivity. Then using Corollary 5.1, the asymptotically optimal point density for fixed-rate quantization satisfies

$$\lambda^*_{n,\text{fmse,fr,chat}}(x \,|\, m) \propto \begin{cases} \left(\gamma_{n|M^c}(x \,|\, m) f_{X_n|M^c}(x \,|\, m)\right)^{1/3} \\ \qquad x \notin Z_n(m) \text{ and } f_{X_n|M^c}(x \,|\, m) > 0; \\ 0, \qquad \text{otherwise.} \end{cases} \tag{5.1}$$

Recall that the point density is the derivative of the compressor function $c(x)$ in the compander model. Hence, codewords are placed at the solutions to $c(x) = (k-1)/(K-L)$ for $k \in [1 : (K-L)]$. In addition, one codeword must be placed in each of the L don't-care interval.

5.4.3 Entropy-Constrained Quantization Design

Using Corollary 5.2, the asymptotically optimal point density when entropy coding is combined with scalar quantization has the form

$$\lambda^*_{n,\text{fmse,ec,chat}}(x \,|\, m) \propto \begin{cases} \gamma_{n|M^c}(x \,|\, m), & x \notin Z_n(m) \text{ and } f_{X_n|M^c}(x \,|\, m) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

Note that rate amplification can arise through chatting, and this can allow distortion terms to decay at rates faster than 2^{-2R_n}. However, there is also a penalty from proper coding of don't-care intervals, corresponding to $H_B(P(A_n))$. This loss is negligible in the high-resolution regime but may become important for moderate rates.

5.4.4 Conditions on Chatting Graph

We have observed that chatting can influence optimal design of scalar quantizers through the conditional sensitivity, and that sensors will vary their quantization code-

books depending on the incoming chatting messages from parent sensors. Under the assumption that the fusion center does not have access to M^c, success of compression is contingent on the fusion center identifying the codebook employed by every sensor from the messages M_1^N.

Definition 5.2. A chatting network is *codebook identifiable* if the fusion center can determine the codebooks of $Q_{K_1^N,\lambda_1^N}$ using the messages it receives from each sensor. That is, it can determine $\mathcal{C}_n(M^c)$ from M_1^N for each time instant.

We have argued that a chatting network can successfully communicate its compressed observations if it is codebook identifiable. The following are sufficient conditions on the chatting graph $\mathcal{G}^c$ and messages M^c such that the network is codebook identifiable:

C1. The chatting graph $\mathcal{G}^c$ is a directed acyclic graph (DAG).

C2. The causality in the chatting schedule matches $\mathcal{G}^c$, meaning for every n, Sensor n sends its chatting message after it receives messages from from all parent sensors.

C3. The quantizer at Sensor n is a function of the source joint distribution and all incoming chatting messages from parent sensors in $\mathcal{N}_p(n)$.

C4. At any discrete time, the chatting message transmitted by Sensor n is a function of M_n and incoming chatting messages from parent sensors in $\mathcal{N}_p(n)$.

The sufficiency of these conditions can be seen by construction. Because $\mathcal{G}^c$ is a DAG, there is at least one sensor which is a head node and does not have incoming chatting messages. Therefore, the chatting messages of these sensors are known to the decoder by condition C4. The remaining codebooks and chatting messages can be recovered by the decoder by C3 and C4 provided C2 holds.

When each sensor's quantizer is regular and encoder only operates on the quantized values $\widehat{X}_n$, matching the DFSQ setup, the chatting message can only influence the choice of codebook. In this setting, the above conditions become necessary as well. Alternatively, if sensors can locally fuse messages from parents with their own observation, there may exist other conditions for a network to be codebook identifi-

able.

We now revisit the example at the end of Section 5.3 and see the the graph is a DAG. Assuming the system requires that C3 and C4 hold, having an additional chatting link from Sensor 2 to Sensor 1 is not useful. This illuminates an important design decision in intersensor communication. When there is flexibility in the design of chatting channels we can restrict the topology to ones that form DAGs. Choosing between different graphs is beyond the scope of the thesis but is of future interest.

5.5 Rate Allocation in Chatting Networks

A consequence of chatting is that certain sensors can exploit their neighbors' acquisitions to refine their own. Moreover, a sensor can potentially use this side information to adjust its communication rate in addition to changing its quantization if the network is codebook identifiable. These features of chatting networks suggest intelligent rate allocation across sensors can yield significant performance gains. In addition, a strong motivation for intersensor interaction is that sensors may be geographically closer to each other than a fusion center and hence require less transmit power, or can employ low-rate orthogonal channels that do not interfere with the main communication network. As a result, the cost of communicating a bit may vary in a network.

This section explores proper rate allocation to minimize the total cost of transmission in a chatting network, allowing asymmetry of the information content at each sensor and heterogeneity of the communication links. Consider the distributed network in Figure 1-1. The cost per bit of the communication link and the resource allocation between Sensor n and the fusion center are denoted by α_n and b_n respectively, leading to a communication rate of $R_n = b_n/\alpha_n$ from Sensor n to the fusion center. Similarly, for a chatting link between Sensors i and n, the cost per bit and resource allocation are denoted by $\alpha_{i\to n}$ and $b_{i\to n}$ respectively, corresponding to a chatting rate of $R_{i\to n} = b_{i\to n}/\alpha_{i\to n}$. Consistent with previous notation, we denote the set of costs per chatting bit, resource allocations on chatting links, and chatting

rates by $\alpha^c = \{\alpha_{i\to n}\}_{(i,n)\in\mathcal{E}}$, $b^c = \{b_{i\to n}\}_{(i,n)\in\mathcal{E}}$, and $R^c = \{R_{i\to n}\}_{(i,n)\in\mathcal{E}}$.

Given a total resource budget C, how should the rates be allocated among these links? For simplicity, assume all chatting links employ fixed-rate quantization; this implies that $K_n = 2^{R_n}$ for all $n \in [1:N]$ and $K_{i\to n} = 2^{R_{i\to n}}$ for all $(i,n) \in \mathcal{E}$. The distortion–cost trade-off is then expressed as

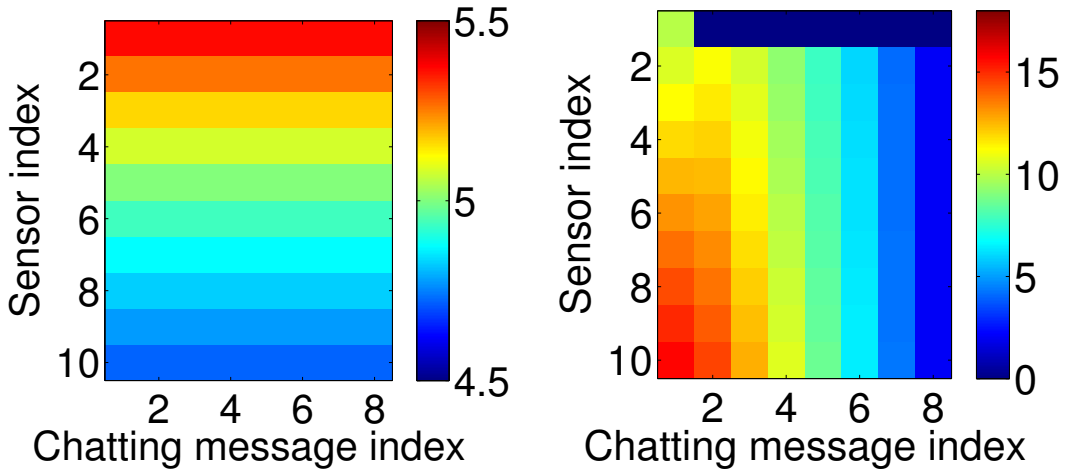
$$D(C) = \inf_{\substack{b_1^N, b^c, \lambda_1^N: \\ \sum_{n=1}^N b_n + \sum_{(i,n)\in\mathcal{E}} b_{i\to n} = C}} D_{\text{fmse}}\left(K_1^N, K^c, \lambda_1^N\right). \tag{5.3}$$

In general, this optimization is extremely difficult to describe analytically since the distortion contribution of each sensor is dependent in a nontrivial way on the conditional sensitivity, which in turn is dependent on the design of the chatting messages. However, the relationship between b_1^N and the overall system distortion is much simpler, as described in Theorem 5.1. Hence, once the chatting allocations b^c is fixed, the optimal b_1^N is easily determined using extensions of traditional rate allocation techniques described in Appendix 5.A. In particular, the optimal b_1^N can be found by applying Lemmas 5.2 and 5.3 with a total cost constraint

$$C' = C - \sum_{(i,n)\in\mathcal{E}} b_{i\to n}.$$

A brute-force search over b^c then provides the best allocation, but this procedure is computationally expensive. More realistically, network constraints may limit the maximum chatting rate, which greatly reduces the search space.

In Figure 5-3, we show optimal communication rates for the network described in Section 5.6. We delay description of the specific network properties and aim only to illustrate how the cost allocations $b_n(m)$ may change depending with sensors or chatting messages. Under fixed-rate coding, b_n varies depending on the chatting graph. In the entropy-constrained setting, the allocation can also vary with the chatting messages, except for Sensor 1. This increased flexibility allows for a wider range of rates, as well as improved performance in many situations.



(a) (b)

Figure 5-3: Cost allocation for a maximum computation network, as described in Section 5.6. In this case, $N = 10$, $C = 5N$, $R_c = 3$, $\alpha_c = 0$, and $\alpha_n = 1$. In the fixed-rate setting **(a)**, the sensors are allowed to have different communication rates but cannot adjust the rate with the received chatting message. In the entropy-constrained setting **(b)**, each sensor except sensor 1 receives chatting messages and can adjust its communication rate appropriately.

5.6 Maximum Computation

The results in the previous sections hold generally, and we now build some intuition about chatting by extending the example of Section 5.3. The choice of this computation is not arbitrary; we will show that it allows for a particular chatting architecture that makes it convenient to study large networks. Moreover, this network reveals some surprising insights into the behavior of chatting. The source variables are assumed to independent so that performance gains come from chatting rather than from the dependence that is traditionally exploited in distributed source coding; one could additionally exploit correlations. While this chapter restricts its attention solely to the maximum computation, more examples are discussed in [174].

5.6.1 Problem Model

We consider a network where the fusion center aims to reproduce the maximum of N sources, where each X_n is independent and uniformly distributed on $[0, 1]$. The sensors measuring these sources are allowed to chat in a serial chain, meaning each

sensor has at most one parent and one child (see Figure 5-4). Initially, we will consider the simplest such network with the following assumptions:

1. The chatting is serial, implying a sequence of chatting messages $\{M_{(n-1)\to n}\}_{n=2}^{N}$.
2. Each chatting link is identical and has rate R_c, codebook size $K_c = 2^{R_c}$ and cost α_c.
3. The communication links between sensors and the fusion center are allowed to have different rates. For simplicity, we assume them to be homogeneous and normalize the cost to be $\alpha_n = 1$.
4. The outgoing chatting message at Sensor 1 is the index of a uniformly quantized version of its observation with K_c levels.
5. For $n > 1$, the chatting message from Sensor n is the maximum of the index of Sensor n's own uniformly quantized observation and the chatting message from its parent.

Under this architecture, the chatting messages effectively correspond to a uniformly quantized observation of the maximum of all ancestor nodes:

$$M_{(n-1)\to n} = \mathcal{I}(Q_{K_c,U}(\max(X_1^{n-1}))), \tag{5.4}$$

where $\mathcal{I}$ is the index of the quantization codeword and can takes values $[1 : K_c]$. The simplicity of the chatting message here arises from the permutation-invariance of the maximum function. We will exploit this structure to provide precise characterizations of system performance.

5.6.2 Quantizer Design

Using (2.13), we find the max function has functional sensitivity profile $\gamma_n^2(x) = x^{N-1}$ for all n. Without chatting, each sensor's quantizer would be the same with a point density that is a function of the source distribution and functional sensitivity profile. Moreover, since the cost per bit of transmitting to the fusion center is the same, the solution of the resource allocation problem assigns equal weight to each link. Hence,

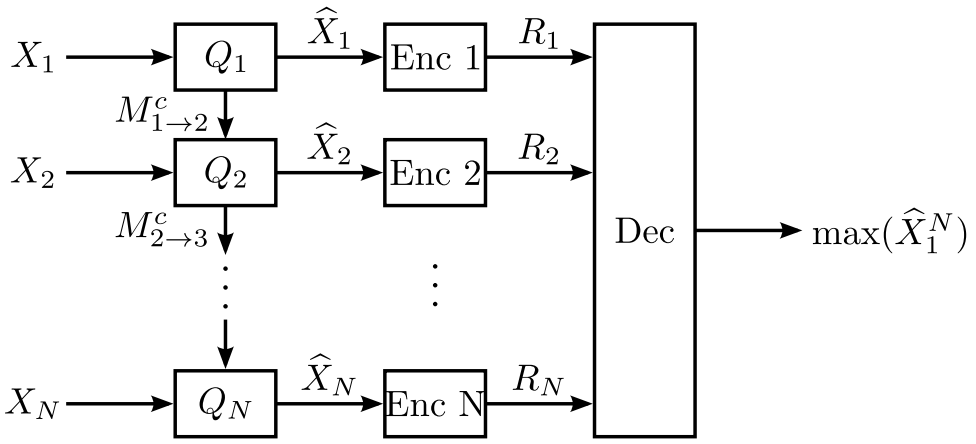


Figure 5-4: A fusion center wishes to determine the maximum of N iid uniform sources and receives messages M_n from each sensor n at rate R_n. The sensors are allowed to chat serially down the network using messages $M_{(n-1)\to n}$ at rate R_c.

minimizing (2.16) yields the optimal fixed-rate distortion–cost trade-off:

$$D_{\max,\mathrm{fr}}(C) \simeq \frac{N}{12}\left(\frac{3}{N+2}\right)^3 2^{C/N}.$$

Similarly, the minimum of (2.19) leads to the optimal entropy-constrained distortion–cost trade-off

$$D_{\max,\mathrm{ec}}(C) \simeq \frac{N}{12} e^{-N+1} 2^{C/N}.$$

These high-resolution expressions provide scaling laws on how the distortion relates to the number of sensors. They require the total cost C increase linearly with N to hold.

With chatting, we first need to determine the conditional sensitivity, which is given below for uniform sources:

Proposition 5.1. *Given $K_c = 2^{R_c}$, the conditional sensitivity corresponding to a received chatting message $M_{(n-1)\to n} = k$ is*

$$\gamma^2_{n\,|\,M_{(n-1)\to n}}(x\,|\,k) = \begin{cases} 0, & x < \frac{k-1}{K_c}; \\ \frac{(K_c x)^{n-1} - (k-1)^{n-1}}{k^{n-1} - (k-1)^{n-1}} x^{N-n}, & \frac{k-1}{K_c} \le x < \frac{k}{K_c}; \\ x^{N-n}, & x \ge \frac{k}{K_c}. \end{cases}$$

Proof. Assuming iid uniform sources on the support $[0, 1]$, the functional sensitivity profile of each sensor in the maximum computation network in Figure 5-4 without chatting is

$$\begin{aligned}\gamma_n^2(x) &= \mathrm{E}[|g_n(X_1^N)|^2 \,|\, X_n = x] \\ &= \Pr\left(\min(X_1^N) = X_n \,|\, X_n = x\right) \\ &= \Pr(X_1 < x) \cdots \Pr(X_{n-1} < x) \Pr(X_{n+1} < x) \cdots \Pr(X_N < x) \\ &= x^{N-1}.\end{aligned}$$

When the chatting graph is a serial chain, Sensor n has some lossy version of the information collected by its ancestor sensors. For the max function, chatting reduces the support of the estimate of $\max(X_1^{n-1})$ by Sensor n. Hence, the message $M_{(n-1)\to n}$ reveals the max of the ancestor sensors is in the range $[s_l, s_u]$. This side information forms three distinct intervals in the conditional sensitivity. First, in the interval $x < s_l$, X_n is assuredly less than $\max(X_1^{n-1})$ and the conditional sensitivity is zero since the information at Sensor n is irrelevant at the fusion center. Second, if $x > s_u$, X_n is greater than $\max(X_1^{n-1})$ and the sensitivity should only depend on the number of descendant sensors, leading to a sensitivity of x^{N-n}. Finally, when $s_l \le x < s_u$, Sensor n must take into consideration both ancestors and descendants, yielding conditional sensitivity

$$\begin{aligned}&\Pr\left(\min(X_1^N) = X_n \,\middle|\, X_n = x, \max(X_1^{n-1}) \in [s_l, s_u]\right) \\ &\qquad = \Pr\left(\max(X_1^{n-1}) < x \,\middle|\, \max(X_1^{n-1}) \in [s_l, s_u]\right) \Pr\left(\max(X_{n+1}^N) < x\right) \\ &\qquad = \frac{x^{n-1} - s_l^{n-1}}{s_u^{n-1} - s_l^{n-1}} x^{N-n}.\end{aligned}$$

More specific to the case when messages correspond to uniform quantization, we define $K_c = 2^{R_c}$ and denote each received message $M_{(n-1)\to n}$ as k_n. Setting $s_l = (k_n - 1)/K_c$ and $s_u = k_n/K_c$ yields Proposition 5.1. □

We have already noted the incident chatting message of Sensor n is a uniformly

quantized observation of $Y_n = \max(X_1^{n-1})$, where $f_Y(y) = (n-1)y^{n-2}$. Hence,

$$\Pr\left(M_{(n-1)\to n} = k\right) = \left(\frac{k}{K_c}\right)^{n-1} - \left(\frac{k-1}{K_c}\right)^{n-1}. \tag{5.5}$$

Below, we give distortion asymptotics for the serial chatting network under both fixed-rate and entropy-constrained quantization.

Fixed-rate case

From Theorem 5.1, the asymptotic total fMSE distortion is

$$\sum_{n=1}^{N} \beta_n 2^{-2R_n},$$

where $\beta_n = \frac{1}{12}\|\gamma^2_{n|M^c}\|_{1/3}$. Because Sensor 1 has no incoming chatting messages, its conditional sensitivity is $\gamma_1^2(x) = x^{N-1}$ and the resulting distortion constant is

$$\beta_1 = \frac{1}{12}\left(\frac{3}{N+2}\right)^3.$$

For other sensors, the distortion contribution is

$$\beta_n = \frac{1}{12}\sum_{k=1}^{K_c} \Pr\left(M_{(n-1)\to n} = k\right) \left\|\gamma^2_{n\,|\,M_{(n-1)\to n}=k}\right\|_{1/3}.$$

For Sensor n with $n > 1$, all incoming messages besides $k = 1$ induce a don't-care interval, so one of the 2^{R_n} codewords is placed exactly at $(k-1)/K$.

We study the trade-off between chatting rate R_c and fMSE for several choices of N and α_c using optimal cost allocation as determined by Lemma 5.2. In Figure 5-5a, we observe that increasing the chatting rate yields improvements in fMSE. As the number of sensors increases, this improvement becomes more pronounced. However, this is contingent on the chatting cost α_c being low. As discussed in Section 5.2, chatting can lead to worse system performance if the cost of chatting is on the same order as the cost of communication given a total resource budget, as demonstrated by

Figure 5-5c. Although the main results of this work are asymptotic, we have asserted the distortion equations are reasonable at finite rates. To demonstrate this, we design real quantizers under the same cost constraint and demonstrate that the resulting performance is comparable to high-resolution approximations of Theorem 5.1. This is observed in Figs. 5-5a and c, which shows the asymptotic prediction of the distortion–rate trade-off is accurate even at 4 bits/sample.

Entropy-constrained case

Generally, the total distortion in the entropy-constrained case is

$$\sum_{n=1}^{N} \mathrm{E}\left[\beta_{n,k} 2^{-2R_{n,k}} \,\middle|\, M_{(n-1)\to n} = k\right],$$

noting each sensor is allowed to vary its communication rate with the chatting messages it receives. Like in the fixed-rate setting, an incoming message k will induce a don't-care interval of $[0, (k-1)/K]$ in the conditional sensitivity. If $A_{n,k}$ is the event that X_n is not in a don't-care interval when receiving message k, then

$$\beta_{n,k} = \frac{1}{12} \Pr\left(M_{(n-1)\to n} = k\right) \, 2^{2h(X_n|A_{n,k}) + 2\,\mathrm{E}[\log \gamma_{n\,|\,M_{(n-1)\to n}}(X_n|k)]}$$

and $R_{n,k} = (R_n - H_B(\Pr(A_{n,k})))/\Pr(A_{n,k})$.

Like in the fixed-rate setting, we study the relationship between the chatting rate R_c and fMSE, this time using the probabilistic allocation optimization of Lemma 5.3 in Appendix 5.A. Due to the extra flexibility of allowing a sensor to vary its communication to the fusion center with the chatting messages it receives, we observe that increasing the chatting rate can improve performance more dramatically than in the fixed-rate case (see Figure 5-5b). Surprisingly, chatting can also lead to inferior performance for some combinations of R_c and N, even when α_c is small. This phenomenon will be discussed in greater detail below. In Figure 5-5d, we compare different choices of α_c to see how performance changes with the chatting rate. Unlike for fixed rate, in the entropy-constrained setting, chatting can be useful even when

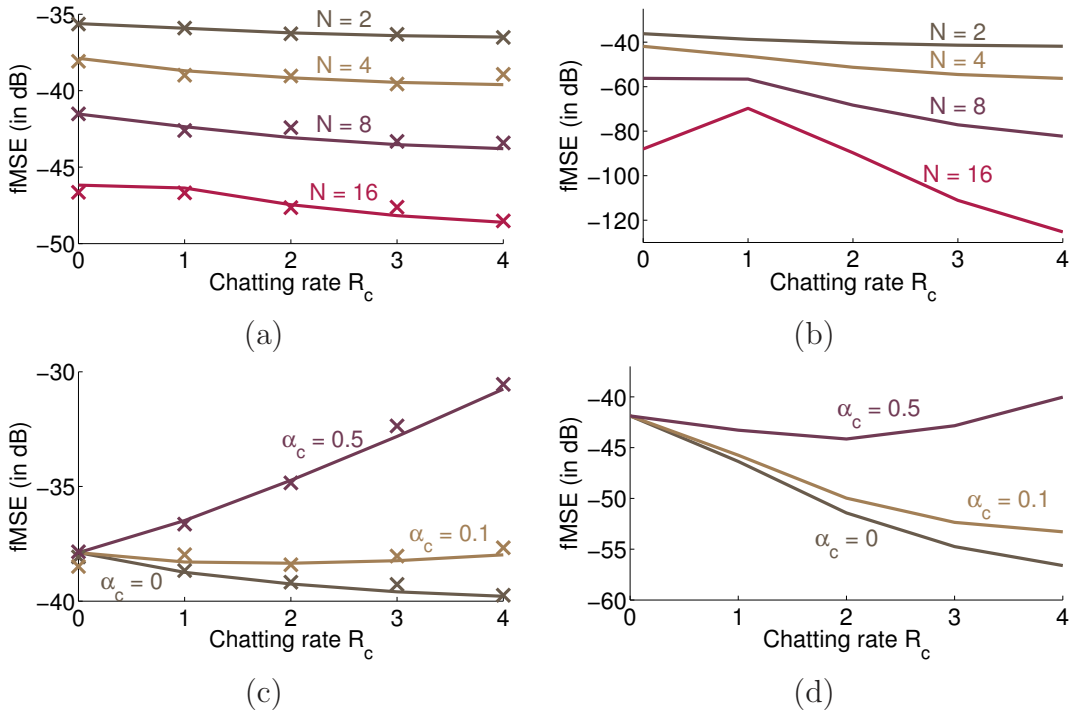


Figure 5-5: Performance of the maximum computation network in both the fixed-rate (left plots) and entropy-constrained (right plots) settings. Plots **(a)** and **(b)** illustrate the trade-off between fMSE and chatting rate for choices of N assuming total cost $C = 4N$ and $\alpha_c = 0.01$. Plots **(c)** and **(d)** illustrate the trade-off between fMSE and chatting rate for choices of α_c assuming $N = 4$ sensors and total cost $C = 4N$. In all cases, the cost of communication is $\alpha_n = 1$. For the fixed-rate setting, we validate the distortion through simulated runs on real quantizers designed using (5.1). We observe that high-resolution theory predicts actual performance at rates as low as 4 bits/sample, as shown by crosses in the fixed-rate plots.

its cost is close to the cost of communication to the fusion center.

5.6.3 Generalizing the Chatting Messages

We have considered the case where a chatting message is the uniform quantization of the maximum of all ancestor nodes, as shown in (5.4). Although simple, this coding of chatting messages is not optimal. Here, we generalize chatting messages to understand how the performance can change with this design choice.

We begin by considering the same network under the restriction that the chatting rate is $R_c = 1$, but allow the single partition boundary p_1 to vary rather than setting it to $1/2$. Currently, we keep the coding consistent for every sensor such that a chatting

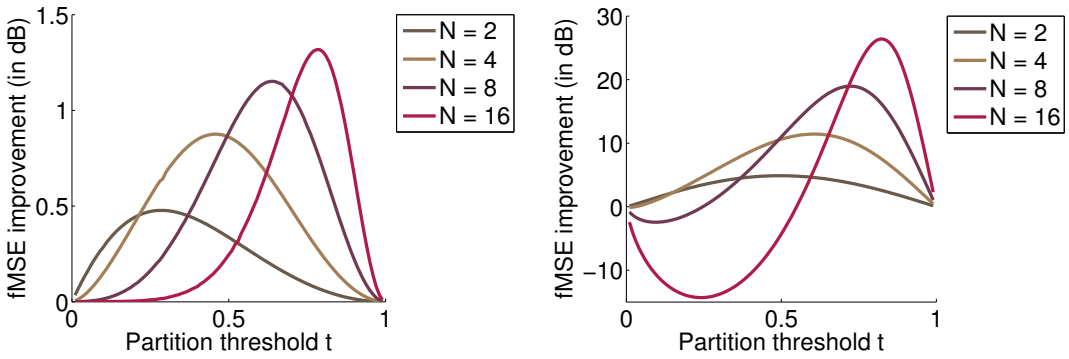
message $k = 1$ implies $\max(X_1^{n-1}) \in [0, p_1]$ and $k = 2$ means $\max(X_1^{n-1}) \in (p_1, 1]$. Distortions for a range of N and p_1 are shown in Figure 5-6.

From these performance results, we see that the choice of p_1 should increase with the size of the network, but precise characterization of the best p_1 is difficult because of the complicated effect the conditional sensitivity has on both the distortion constants and rate allocation. We can recover some of the results of Figure 5-5 by considering $p_1 = 1/2$. It is now evident that this choice of p_1 can be very suboptimal, especially as N becomes large. In fact, we observe that for certain choices of the partition with entropy coding, the distortion with chatting can be larger than from a traditional distributed network even though the chatting cost is zero. This unintuitive fact arises because the system's reliance on the conditional sensitivity is fixed, and the benefits of a don't-care interval are mitigated by creating a more unfavorable conditional sensitivity. We emphasize that this phenomenon disappears as the rate becomes very large.

Since the flexibility in the choice of the chatting encoder's partitions can lead to improved performance when $R_c = 1$, we can expect even more gains when the chatting rate is increased. However, the only method for optimizing the choice of partition boundaries developed currently involve brute-force search using the conditional sensitivity derived in the proof of Proposition 5.1. Another extension that leads to improved performance is to allow chatting encoders to employ different partitions. This more general framework yields strictly improved results, but some of the special structure of the serial chatting network is lost as the chatting message is no longer necessarily the maximum of all ancestor sensors. The added complexity of either of these extensions make their performances difficult to quantify.

5.6.4 Optimizing a Chatting Network

So far, we have formulated a framework allowing low-rate collaboration between sensors in a distributed network. We have introduced several methods to optimize such a network, including nonuniform quantization, rate allocation, and design of chatting messages. Here, we combine these ingredients and see how each one impacts fMSE.



(a) (b)

Figure 5-6: Distortion improvement compared to no chatting in the maximum computation network for the fixed-rate (left plot) and entropy-constrained (right plot) settings when varying the partition boundary p_1. We assume chatting is free, i.e., $\alpha_c = 0$, but the chatting rate is limited to one bit.

We will continue working with the maximum computation network from Figure 5-4 assuming $R_c = 1$, $\alpha_c = 0$, $N = 5$ and $C = 5N$. We further assume the coding of chatting messages is the same for every sensor on the serial chain. We will then consider the following scenarios:

1. A chatting network with $R_n = 5$ for all n and chatting designed by (5.4).
2. A chatting network with rate allocation and chatting designed by (5.4).
3. A chatting network with rate allocation and optimization over chatting messages.

We compare the fMSE of each scenario to the performance of the distributed network without chatting ($R_c = 0$). In the results of Figure 5-7, we see that the simple chatting network with a chatting codebook described in (5.5) provides meaningful performance boost, while additional optimizations such as rate allocation and more general chatting codebooks do not add appreciable benefits. The opposite is true in the entropy-constrained setting, where the addition of the chatting channel is only meaningful when rate allocation and chatting codebook optimizations are considered. However, the potential gains from chatting in the entropy-coded setting is much greater; in the example presented, a 20 dB improvement in fMSE can be seen. We highlight that the current results restrict the communication in the fixed-rate setting to employ fixed-rate quantization. Allowing for entropy-coding on the chatting

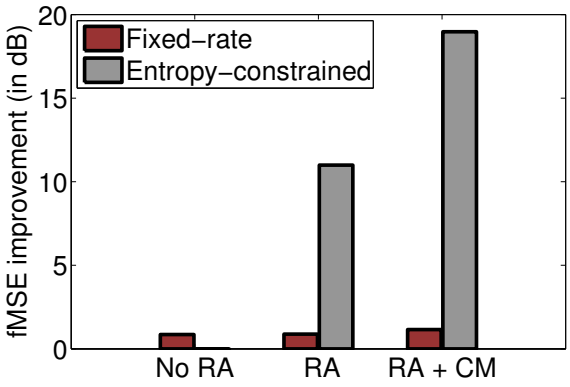


Figure 5-7: Distortion improvement for Scenarios 1–3 over a distributed network without chatting. Both rate allocation (RA) and chatting message optimization (CM) are considered.

channel may lead to even greater compression gain at the expense of increased system complexity.

5.7 Conclusions

In this chapter, we explored how intersensor communication—termed *chatting*—can improve approximation of a function of sensed data in a distributed network constrained to scalar quantization. We have motivated chatting from two directions: providing an analysis technique for distortion performance when block-length limitations make Shannon theory too optimistic, and illustrating the potential gains over simplistic practical designs. There are many opportunities to leverage heterogeneous network design to aid information acquisition using the tools of high-resolution theory, and we provide precise characterizations of distortion performance, quantizer design, and cost allocation to optimize distributed networks. Many challenges remain in analyzing chatting networks. Some future directions that are meaningful include a more systematic understanding of how to design chatting messages and applications where chatting may be feasible and beneficial.

One can consider "sensors" being distributed in time rather than space, with the decoder computing a function of samples from a random process. Connections of this formulation to structured vector quantizers are of independent interest.

5.A Rate Allocation for Distributed Networks

Consider the distributed network in Figure 1-1 without the chatting channel. The cost per bit of the communication link and the cost allocation between Sensor n and the fusion center is denoted by α_n and b_n respectively, leading to a communication rate of $R_n = b_n/\alpha_n$. Below, we solve the cost allocation problem under the assumption that companding quantizers are used and noninteger rates are allowed.

Lemma 5.1. *The optimal solution to*

$$D(C) = \min_{\sum b_n = C, b_n \geq 0} \sum_{n=1}^{N} \beta_n 2^{-2b_n/\alpha_n} \tag{5.6}$$

has cost allocation

$$b_n^* = \max\left(0, \frac{1}{2}\log\frac{\beta_n/\alpha_n}{\tilde{\beta}}\right), \tag{5.7}$$

where $\tilde{\beta}$ is chosen such that $\sum b_n^* = C$.

Proof. This lemma extends the result from [161] or can be derived directly from the KKT conditions. □

Each β_n is calculated using only the functional sensitivity profile γ_n and marginal source pdf f_{X_n}. Although Lemma 5.1 is always true, we emphasize that its effectiveness in predicting the proper cost allocation in a distributed network is only rigorously shown for high cost (i.e. high rate) due to its dependence on (2.14). However, it can be experimentally verified that costs corresponding to moderate communication rates still yield near-optimal allocations.

When the solution of Lemma 5.1 is positive, a closed-form expression exists:

Lemma 5.2. *Assuming each b_n^* in* (5.7) *is positive, it can be expressed as*

$$b_n^* = \frac{\alpha_n}{\tilde{\alpha}}C + \frac{\alpha_n}{2}\log\frac{\beta_n/\alpha_n}{\left(\prod_j (\beta_j/\alpha_j)^{\alpha_j}\right)^{1/\sum\alpha_i}}.$$

Proof. The proof uses Lagrangian optimization. □

If Sensor n is allowed to vary the communication rate depending on the side information $M_{\mathrm{si},n}$ it receives, further gains can be enjoyed. This situation is natural in chatting networks, where the side information is the low-rate messages passed by neighboring sensors. Here, we introduce *probabilistic cost allocation*, yielding a distortion–cost trade-off

$$D(C) = \min_{\substack{\sum \mathrm{E}[b_n(M_{\mathrm{si},n})]=C \\ b_n(m)\geq 0}} \sum_{n=1}^{N} \mathrm{E}\left[\beta_n(M_{\mathrm{si},n}) 2^{-2b_n(M_{\mathrm{si},n})/\alpha_n}\right], \tag{5.8}$$

where the expectation is taken with respect to $M_{\mathrm{si},n}$. Each link will have a cost allocation $b_n(m)$ for every possible message m while satisfying an average cost constraint. An analogous result to Lemma 5.1 can be derived; for the situation where the optimal allocation is positive, it can again be expressed in closed form:

Lemma 5.3. *Assume the side information $M_{\mathrm{si},n}$ received at Sensor n is $m \in \mathcal{M}_n$ and the cost per bit of the communication link may vary with m. Assuming each allocation $b_n^*(m)$ in the solution to (5.8) is positive, it can be expressed as*

$$b_n^*(m) = \frac{\alpha_n(m)}{\tilde{\alpha}} C + \frac{\alpha_n(m)}{2} \log \frac{\beta_n(m)/\alpha_n(m)}{\prod_j \prod_l \left((\beta_j(l)/\alpha_j(l))^{\alpha_j(l)/\tilde{\alpha}} \right)},$$

where $\tilde{\alpha} = \sum_n \sum_m f_{M_{\mathrm{si},n}}(m)\, \alpha_n(m)$.

Here, we extended previous known rate allocation results [54, 161] to account for heterogeneity in distributed networks. Although these results do not account for chatting, we see in Section 5.5 that they become important tools in optimizing performance in such networks.

Chapter 6

Quantization for Relative Error

While the previous chapters explored quantizer performance in networks, the remainder of the thesis is focused on quantization of scalar sources for perceptually relevant distortion measures. In particular, the emphasis is on *relative error*, which is defined as

$$d_{\mathrm{re}}(x, \hat{x}) = \frac{(x - \hat{x})^2}{x^2}. \tag{6.1}$$

Relative error is rarely discussed in the source coding literature but is prominently used in a variety of scientific and engineering contexts, such as coding of speech. The theoretical results of this chapter will then provide the foundation for applying Bayesian quantization theory to understand psychophysical laws for human perception in Chapter 7.

Although we use techniques developed for DFSQ, the motivation of this work is perceptual coding, which also inspired fundamental results on vector quantization for nondifference distortion measures. We begin with discussion on the usefulness of the new theory in Section 6.1 and then describe the main results in Section 6.2. We provide experimental validation in Section 6.3 and generalize the results in Section 6.4. We conclude with some applications that can benefit from this analysis in Section 6.5.

Portions of this work has been published in [173].

6.1 Motivation

In this chapter, we extend methods developed in DFSQ theory to optimize quantizers for a subclass of nondifference distortion measures, with strong emphasis on relative error. Though it may seen trivial to highlight relative error when generalizations are known, we believe that neglect of this distortion measure in the source coding community and its importance to many fields make it worthy of analysis. For example, relative error is often used explicitly in the engineering literature for numerical analysis and fidelity of floating-point representations (e.g. [90]). It is also prominently featured in relating fidelity in many of the physical and social sciences. Finally, it serves as a heuristic justification for using logarithmic quantization, which is so ubiquitous in the compression and circuits literature that it is often simply known as the companding quantizer [35, 54, 125]. Using a logarithmic compressor function before quantization has been used a variety of speech coding standards to reduce the dynamic range. Most prominently, it is used in international standards on telephony compression under the names μ-law and A-law companding [75]. It was also an important component of the SIGSALY secure speech system innovated by Clark *et al.* during World War II.

Here, we formalize this justification by showing logarithmic companding is indeed optimal if expected relative error is the measure of interest and the quantization is entropy constrained. We also demonstrate that this does not hold in the fixed-rate setting, but that optimal quantizers can be very similar depending on the source distribution. Moreover, we present some new applications where relative error may be useful in analyzing signal acquisition.

Relative error falls under a class of nondifference distortion measures called input-weighted quadratic or Itakura–Saito [74], which has been studied using high-resolution theory both in fixed-rate [97] and entropy-constrained [102] settings. Iterative methods exploiting the necessary nearest-neighbor and centroid conditions provide algorithmic results on construction of optimal quantizers [18,98]. Of particular note, Gray

and Karnin determined Lloyd-like conditions for the vector generalization of (6.1):

$$d(x_1^N, \hat{x}_1^N) = \frac{\|x_1^N - \hat{x}_1^N\|^2}{\|x_1^N\|^2},$$

which they called "short term sample quantization-noise-to-signal ratio" [64].

6.2 Optimal Quantization Design

The primary goal of this section is to use DFSQ theory to find asymptotic distortion results for a class of nondifference distortion measures that take the form

$$d_{\mathrm{nd}}(x, \hat{x}) = n(x) \left(m(x) - m(\hat{x})\right)^2,$$

where m is Lipschitz continuous and twice differentiable, and n is non-negative, bounded and piecewise smooth. This is related to Trushkin's distortion measure, where the distortion $d(x, \hat{x})$ has the form $\rho(x, \hat{x} - x)$ [180] and must satisfy certain regularity conditions.

Using high-resolution analysis, we have the following theorem:

Theorem 6.1. *Consider a memoryless source X with probability density f_X that is smooth on a compact subinterval of $(0, \infty)$ and zero elsewhere. The source is quantized using a nonuniform scalar quantizer $Q_{K,\lambda}$ constructed using the compander model and specified by λ and K. Assuming the smoothness conditions of m and n are met, the relative error between the output of the quantizer and the source satisfies*

$$D_{\mathrm{nd}}(K, \lambda) \simeq \frac{1}{12K^2} \mathrm{E}\left[\gamma^2(X)/\lambda^{-2}(X)\right],$$

where $\gamma(x) = |\sqrt{n(x)}m'(x)|$.

Proof. This theorem can be proven by emulating the steps for the derivation of (2.4) (see, e.g., [66]) with the new cost. Presented here is an informal and intuitive argument based on ideas used in DFSQ. The full rigorous proof follows the arguments in Chapter 3.

To compute expected cost for the quantized signal, we use the total expectation theorem to separate the error terms for the quantization cells and use Taylor expansion to rewrite the difference term as

$$m(x) - m(c_k) = m'(c_k)(x - c_k) + \mathcal{O}(|x - c_k|^2)$$

for $x \in P_k$. The residual terms are inconsequential in the limit of large K and the distortion becomes

$$\begin{aligned}
D_{\text{fmse}}(K, \lambda) &= \sum_{k=1}^{K} \mathrm{E}\left[n(X)\left(m(X) - m(c_k)\right)^2 \mid X \in P_k\right] \mathbf{P}(X \in P_k) \\
&\overset{(a)}{\approx} \sum_{k=1}^{K} \mathrm{E}\left[n(X)\left(m'(c_k)(X - c_k)\right)^2 \mid X \in P_k\right] \mathbf{P}(X \in P_k) \\
&\overset{(b)}{\approx} \sum_{k=1}^{K} \mathrm{E}\left[n(c_k)\left(m'(c_k)(X - c_k)\right)^2 \mid X \in P_k\right] \mathbf{P}(X \in P_k) \\
&\overset{(c)}{\approx} \frac{1}{12} \sum_{k=1}^{K} \left(\frac{\sqrt{n(c_k)}m'(c_k)}{K\lambda(c_k)}\right)^2 \mathbf{P}(X \in P_k) \\
&\overset{(d)}{\approx} \frac{1}{12K^2} \mathrm{E}\left[\gamma^2(X)/\lambda^2(X)\right],
\end{aligned}$$

where (a) follows from Taylor expansion; (b) holds when K is large and n is bounded and smooth; (c) uses the high-resolution approximation of $\text{length}(P_i) \approx (K\lambda(c_i))^{-1}$; and (d) follows from setting $\gamma(x) = |\sqrt{n(x)}m'(x)|$ and using the standard high-resolution technique of approximating the expectation using a Riemann sum. □

We now consider *expected relative error* from scalar quantization, which takes the form

$$D_{\text{ere}}(K, \lambda) = \mathrm{E}\left[\frac{|X - Q_{K,\lambda}(X)|^2}{X^2}\right]. \tag{6.2}$$

Relative error corresponds to the squared error scaled by the squared magnitude of the input, making the result invariant to scale. First, we find the distortion performance of a quantizer $Q_{K,\lambda}$ is now simple using Theorem 6.1.

Corollary 6.1. *Consider a memoryless source X with probability density f_X that*

is smooth on a compact subinterval of $(0, \infty)$ *and zero elsewhere. The source is quantized using a nonuniform scalar quantizer constructed using the compander model and specified by* λ *and* K*. The relative error between the output of the quantizer and the source satisfies*

$$D_{\text{ere}}(K, \lambda) \simeq \frac{1}{12K^2} \, \mathrm{E}\left[X^{-2}\lambda^{-2}(X)\right].$$

Proof. For relative error, $m(x) = x$ and $n(x) = x^{-2}$. As a result, $\gamma(x) = 1/x$, which can be combined with Theorem 6.1 to yield the result. Moreover, since this expression matches (2.14), we can easily find the optimal point densities using the recipes prescribed in Section 2.3. □

The restriction to memoryless random variables is unnecessary if the quantizer is scalar.

Corollary 6.2. *The result of Corollary 6.1 applies also for stationary discrete-time random processes with distribution* f_X*. If intersample preprocessing is not allowed, the best scalar quantizer for a stationary random process is the same as if the source was memoryless and had the same marginal density.*

With an understanding of the distortion performance with respect to K, we can now consider the source coding problem with scalar quantizers.

Corollary 6.3. *For a given source probability density* f_X *with support contained in* $[a, b]$ *with* $0 < a < b < \infty$*, the optimal point density for fixed-rate quantization is*

$$\lambda^*_{\text{ere,fr}}(x) = \frac{x^{-2/3} f_X^{1/3}(x)}{\int_a^b t^{-2/3} f_X^{1/3}(t)\, dt}, \qquad \text{if } x \in [a, b]; \quad \text{and 0 otherwise.}$$

Corollary 6.4. *For a given source probability density* f_X *with support contained in* $[a, b]$ *with* $0 < a < b < \infty$*, the optimal point density for variable-rate quantization is*

$$\lambda^*_{\text{ere,ec}}(x) = \frac{1/x}{\int_a^b 1/t\, dt}, \qquad \text{if } x \in [a, b]; \quad \text{and 0 otherwise.}$$

The entropy-constrained case is particularly interesting because the equivalent compander is $c(x) = \ln(x)$, meaning the codewords are uniform on a logarithmic

scale over the support of the source. As for the MSE-optimized quantizer, $\lambda^*_{\text{ere,ec}}(x)$ does not depend on the source distribution except in that it spans the same support. In general, the fixed-rate quantizer will not be logarithmic except when $f_X(x) \propto 1/x$ over the support.

6.3 Numerical Results

As expected, if the true cost is ERE, using λ^*_{ere} is better than λ^*_{mse}. In fact, the improvements can be substantial if the source has support that spans several orders of magnitude. However, the measure is ill-posed if the support includes zero, so we currently restrict our attention to sources that take strictly positive values. We loosen this restriction in Section 6.4.

For example, consider X uniformly distributed on $[a, b]$ for $0 < a < b < \infty$. Using (2.7), the fixed-rate quantizer designed for MSE is asymptotically

$$\lambda^*_{\text{mse,fr}}(x) = 1/(b-a), \qquad \text{if } x \in [a,b]; \quad \text{and 0 otherwise.}$$

Using Corollary 6.3, the fixed-rate quantizer designed for relative error is asymptotically

$$\lambda^*_{\text{ere,fr}}(x) = x^{-2/3}/3(b^{1/3} - a^{1/3}), \qquad \text{if } x \in [a,b]; \quad \text{and 0 otherwise.}$$

Applying Corollary 6.1 to both point densities then yields the performance. In particular, the best relative error for the uniform distribution is

$$D^*_{\text{ere,fr}}(R) \simeq \frac{27}{12} \cdot \frac{b^{1/3} - a^{1/3}}{b-a}\, 2^{-2R}.$$

Letting $(a, b) = (1, 10)$, the optimal relative-error quantizer yields a 2.4 dB improvement over the MSE quantizer. Meanwhile, $(a, b) = (1, 1000)$ leads to a performance improvement of 17 dB. We can see this can be arbitrarily large depending on the support of X.

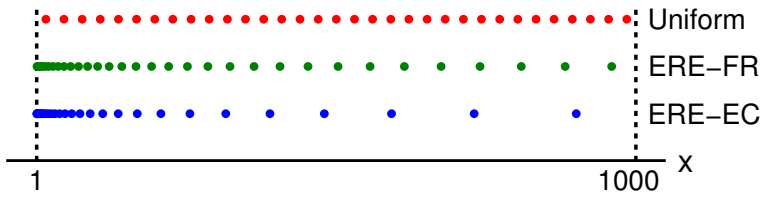


Figure 6-1: Codebooks for fixed-rate and variable-rate quantizers optimized for relative error using a source uniformly distributed on $[1, 1000]$, as well as for a uniform quantizer, which would be optimal if the cost was MSE. The relative-error quantizers have finer quantization for smaller magnitude since the scaled error will be greater. The two variable-rate quantizers only depend on the support of the source, not its distribution.

In the entropy-constrained case, quantizers optimized for both MSE and ERE only depend on the support of the source, and the codewords are uniformly placed on linear and logarithmic scales respectively. Again, we can find the distortion using Corollary 6.1, with the best possible performance for this source being

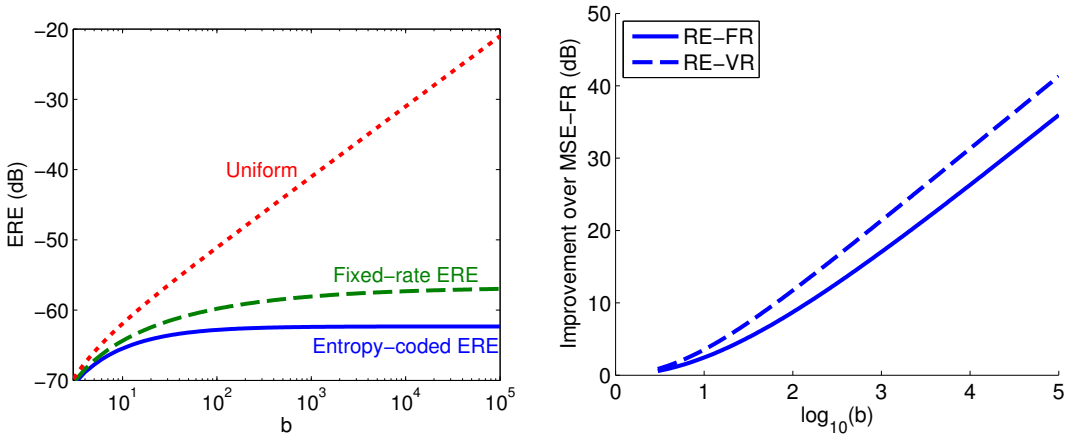
$$D^*_{\text{ere,ec}}(R) \simeq \frac{(b-a)^2}{12} 2^{-2(R-C)},$$

where $C = 1/\ln(2) - (b\log_2 b - a\log_2 a)/(b-a)$. Letting $a = 1$ and $b = 10$ or 1000, using $\lambda^*_{\text{ere,ec}}(x)$ will yield an additional performance gain of 1.1 and 4.3 dB respectively over using $\lambda^*_{\text{ere,fr}}(x)$. These gains are for any rate since all quantizers considered have the same 2^{-2R} decay.

Figure 6-1 shows how the codebooks of the respective types of quantization differ and Figure 6-2 demonstrates the performance trends as the length of the support changes.

6.4 Generalizing ERE

One major limitation of expected relative error is that it is not well-defined when the support of f_X includes zero because the error can grow without bound when X is very small. To combat this problem, we introduce *generalized expected relative error* (gERE) as a way to balance worst-case and expected relative error. Generalized



(a) (b)

Figure 6-2: Performance results of the fixed- and variable-rate quantizers with a uniform source $X \sim \mathcal{U}(1, b)$. The plots provide trends in terms of **(a)** relative error in dB for $R = 6$; and **(b)** performance gain over the uniform quantizer. The performance gain does not depend on rate.

expected relative error, parametrized by ε, is defined as

$$D_{\text{gere}}(K, \lambda, \varepsilon) = \mathrm{E}\left[\frac{|X - Q_{K,\lambda}(X)|^2}{X^2 + \varepsilon}\right]. \tag{6.3}$$

Using the definition of γ in Theorem 6.1, we can find the distortion–rate performance of a scalar quantizer:

Proposition 6.1. *For a memoryless source X with probability density f_X that is smooth on a compact subinterval of $(0, \infty)$ and zero elsewhere, the gERE distortion from applying a companding quantizer $Q_{K,\lambda}$ satisfies*

$$D_{\text{gere}}(K, \lambda, \varepsilon) \simeq \frac{1}{12K^2} \mathrm{E}\left[\lambda^{-2}(X)/(X^2 + \varepsilon)\right].$$

The functional sensitivity profile is

$$\gamma_{\text{gere}}(x, \varepsilon) = |\sqrt{n(x, \varepsilon)} m'(x)| = \frac{1}{\sqrt{x^2 + \varepsilon}}.$$

With $\gamma_{\text{gere}}(x, \varepsilon)$, we can find the asymptotically optimal point densities and the corresponding distortions. An interesting result arises for entropy-constrained quan-

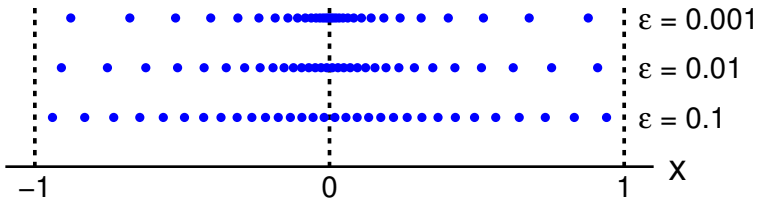


Figure 6-3: Codebooks for variable-rate quantization optimized for generalized relative error when the source has support on $[-1, 1]$.

tization, where $\lambda^*_{\text{gere,ec}}(x, \varepsilon) \propto \gamma_{\text{gere}}(x, \varepsilon)$ yields

$$c^*_{\text{gere}}(x, \varepsilon) \propto \sinh^{-1}(x/\sqrt{\varepsilon}).$$

The effect of ε is demonstrated in Figure 6-3.

We may also generalize the squared error to

$$D_{\text{rere}}(K, \lambda, r) = \mathbf{E}\left[\frac{|X - Q_{K,\lambda}(X)|^r}{X^r}\right], \tag{6.4}$$

defined as the *rth-power expected relative error* (rERE).

Proposition 6.2. *For a memoryless source X with probability density f_X that is smooth on a compact subinterval of $(0, \infty)$ and zero elsewhere, the rERE distortion from applying a companding quantizer $Q_{K,\lambda}$ satisfies*

$$D_{\text{rere}}(K, \lambda, r) \simeq \frac{1}{(r+1)2^r K^r}\, \mathbf{E}[X^{-r}\lambda^{-r}(X)].$$

This result suggests that when entropy coding is allowed, the optimal compressor function is logarithmic regardless of r.

6.5 Applications of ERE

We now discuss some applications where the relative error measure may lead to better performance results. In the quantization literature, relative error is usually a justification for using logarithmic companding in perceptual coding. In this work,

we formalize this intuition and show logarithmic companding is indeed optimal for entropy-constrained quantization using the expected relative error fidelity criterion. We also find that this is not necessarily true for fixed-rate quantization, but the optimal quantizer may have a mapping that is similar depending on the source distribution. In practice, certain parameters for μ-law and A-law companding perform well for speech, and it may be of interest to compare these quantizers to ones optimized for relative error under a realistic speech prior.

Beyond perceptual coding, ERE may have broader implications in biological systems. As a precursor to the fields of psychology and neuroscience, Ernst Weber and Gustav Fechner at different times argued that human perception is logarithmically proportional to many physical stimuli. Examples may be found in touch, vision [166], hearing [124], and numerical cognition [29]. However, the Weber–Fechner law is derived using a differential equation for which there is little neuroscientific evidence. An alternative approach to understanding this phenomenon is to model it as a data acquisition and compression problem. In the next chapter, we tackle this relationship in greater detail.

There are also many engineering applications that perform signal acquisition and may benefit for considering ERE when gain control may be difficult to perform and the signal can have large variations. An example of this is wireless communications with fading channels [181]. If the channel gain is known, the distortion due to quantization can be greatly reduced if the gain is inverted before the ADC. However, when channel state information is not known, designing the quantizer using a relative error metric can yield better results than using a quantizer optimized for squared error.

Chapter 7

Understanding Psychophysical Scales

Scientific inquiries on quantitative relationships between external stimuli and human perception exist under the umbrella of psychophysics [55]. These concepts then motivate lossy compression algorithms that have enormous engineering impact. The results of Chapter 6 and previous work on perceptual coding [97, 102] provide a theoretical foundation for practical algorithms with respect to distortion measures which mimic human perception [78]. In this chapter, we attempt to close the loop by using source coding to develop falsifiable hypotheses on why psychophysical scales are shaped as such. Specifically, we posit two lossy compression models that argue experimental observations in perception are consistent with the sensory being optimal for information acquisition. We begin by motivating the work in Section 7.1. We then introduce two models for human perception in Sections 7.2 and 7.3. To validate the model, we present several theoretical and empirical examples in Section 7.4 and end with a discussion in Section 7.5.

Portions of this work has been published in [177] and [185]. Recently, there has been some follow-up work in the neuroscientific community [144].

7.1 Motivation

Psychophysical scales relate stimulus intensity to perceptual intensity and are central to understanding how the external world is internally represented. For example, the logarithmic scale described by the Weber–Fechner law is consistent with observations in audition, vision, haptics and numerosity [29, 40, 124, 166]. Alternatives to the Weber–Fechner scaling have also been proposed, e.g. a power law [46, 169, 170]. Although there is some debate over the validity of various psychophysical laws for different perceptual modalities [93], many researchers suggest the psychophysical predictions of these models are essentially equivalent [29, 108].

A psychophysical scale is described by an increasing function C such that $P = C(S)$, where S and P are random variables corresponding to stimulus and perceptual intensities respectively. The Weber–Fechner law specifies C as $P \propto \ln(S/s_0)$, where s_0 is the threshold below which a stimulus is not perceived (making P a nonnegative quantity). Thus under the Weber–Fechner law, a multiplicative increase in stimulus intensity leads to an additive increase in perceived intensity.

Several principles have been advanced to explain psychophysical scales, but these are formulated purely at the implementational or algorithmic levels [114] without consideration of computational purpose. In particular, arguments based on the physical chemistry of sensory receptors [22] and based on the informational properties of individual neurons [108] also yielded the Weber–Fechner law, but these arguments did not consider perceptual fidelity. On the other hand, Fechner solved a differential equation inspired by Weber's 'just noticeable difference' (JND) experimental procedure, yielding a logarithmic scale, but did not relate it to neurobiology [40].

Instead, we propose that psychophysical scales arise at the computational level as optimizations under neurobiological constraints (at the implementational and algorithmic levels). Two threads of theoretical work in neuroscience have emerged that attempt to connect physical properties and constraints of the nervous system to psychological and behavioral properties. The first argues that the physical substrate of perception in the nervous system is algorithmically well-matched to the

statistical properties of the natural environment [132] and that therefore operation of the brain is probabilistically efficient, i.e. Bayes-optimal [45, 77, 89]. The second thread argues that internally, brains have remarkable biophysical efficiency when performing information and communication tasks [95] and therefore achieve information theoretically-optimal signal transmission [15]. In both threads, design principles governing the nervous system are said to be similar to those in optimized electronic information systems.

Building on these neurobiological insights and therefore adopting the view of optimal information processing and transmission, this chapter provides a mathematical framework for understanding psychophysical scales as Bayes-optimal and information-theoretically optimal representations of time-sampled continuous-valued stimuli. Specifically, for statistical distributions that correspond to natural stimuli, the Weber–Fechner law and related scales minimize the expected relative error (ERE) of representing the stimulus intensity under two models, each motivated by informational limitations. In the first model, each representation of a stimulus takes one of a finite set of values which all have equal communication cost, commonly referred to as fixed-rate quantization in this thesis. We also discuss an analog information system that has asymptotically equivalent behavior. The second model extends the first by allowing for compressed representations, i.e. elements of the set have varying communication cost, and may be more suitable when neural storage or communication has high cost. This is known in the thesis as entropy-constrained quantization.

In measuring quantization error for stimuli that span orders of magnitude, we have already argued in Chapter 6 that ERE, is a suitable measure of accuracy. In this context, we define ERE as

$$D_{\text{ere}}(K, C') = \mathrm{E}\left[\frac{|S - \widehat{S}|^2}{S^2}\right], \tag{7.1}$$

where $\widehat{S}$ is the perceived stimulus intensity given a true stimulus intensity S. As a preamble to the current discussion, we have already seen that the compressor structures optimized for ERE often look logarithmic, and are exactly logarithmic when

optimal coding after quantization.

Another quantization model was recently proposed to show that the spacing of quantization cells are logarithmically spaced when minimizing minimizing *worst-case* relative error [148]. This result is compelling because it does not assume finiteness of representations or any probabilistic model. However, worst-case quantizers generally perform poorly because they do not leverage information about the stimulus distribution and thus have lower information transmission rates. Moreover, while it is intuitive that a worst-case relative error criterion yields a logarithmic scaling, we will show that the logarithmic scaling arises in a Bayesian formulation *only* when the stimulus distribution takes a certain form or when the quantized values are compressed. Hence, the theory presented here is more biologically plausible and provides meaningful insights regarding how the brain might exploit the distributions of natural stimuli.

One may wonder if relative error, either expected or worst-case, has psychological significance. Before making the worst-case assumption, Portugal and Svaiter motivated their work by noting that relative error is prominent in numerical analysis and physics, and hence has significant value in computations that may occur at a cognitive level [148] . Meanwhile, [173] formalized the use of ERE for compression of audio and video, benefiting from decades of psychophysical studies for the purpose of matching data compression to perceptual processes in the brain [78]. Often, perception appears to be sensitive to ratios between stimulus intensities rather than absolute stimulus intensities—the outputs of many perceptual processes appear to be independent of scale—hence relative error is the natural fidelity criterion. The quantization models proposed here can be generalized to account for other distortion measures, and we will briefly discuss how to experimentally test the validity of error criterion assumptions.

7.2 A Bayes-Optimal Model for Limited Perception

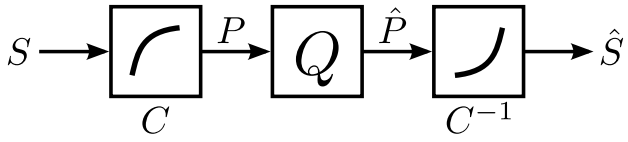
We begin by formulating the first quantization model for perception. Like in earlier work, we abstract away neural circuitry and underlying communication mechanisms at the implementation level. By considering cognitive responses to stimuli, we show that perception fits naturally into a quantization framework.

Recall the psychophysical scale is denoted $P = C(S)$. Since perception lies on a subjective scale, we normalize the minimal and maximal perceptual intensities to be zero and one respectively without loss of generality. For natural stimuli, the space of S is continuous. In most psychophysical studies, the space of P is assumed to have such fine resolution that it can be approximated as continuous. However, assuming that perception is limited, the brain only distinguishes a discrete set of levels. Abstractly, this means that a range of P is mapped to a single representative point $\widehat{P}$ (Figure 7-1). Although many (possibly probabilistic) mappings are possible, a uniform quantizer has many desirable properties such as simplicity of design (set of thresholds) and stability of output. We denote this discrete mapping $Q_{K,U}$, leading to $\widehat{P} = Q_{K,U}(P)$, with output space the set of percepts $\mathcal{P}_K \triangleq \{\widehat{P}\}$ and $K = |\mathcal{P}_K|$. We call $\mathcal{P}_K$ the *perception dictionary*. Using the invertible function C, the equivalent representative stimulus intensities are $\widehat{S} = C^{-1}(\widehat{P})$.

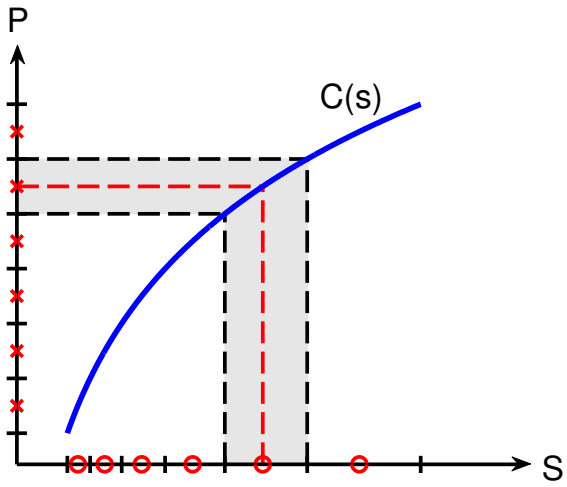
Since the spacings between elements in the perception dictionary, called *perceptual uncertainty*, are equidistant and fixed by K, the only flexibility in this model is C. We have seen this construction before—it is precisely the companding model for nonuniform quantization shown in Figure 2-1. Sticking to our previous notation, the Bayesian optimization is

$$\underset{C(s)}{\operatorname{argmin}}\ D_{\text{ere}}(K, C'). \tag{7.2}$$

For a given stimulus pdf f_S, what choice of C minimizes the ERE between S and $\widehat{S}$? How does the size of the set of $\widehat{P}$ affect this choice? Using the theory developed previously, we can answer these questions precisely. In particular, we use the high-



(a)



(b)

Figure 7-1: **(a)** Block diagram of the quantization model for perception, which is strongly influenced by the companding model (Figure 2-1). Stimulus intensity S is transformed by a nonlinear scaling C, resulting in perceptual intensity P. Since biological constraints limit perception, only a discrete set of K levels $\mathcal{P}_K$ are distinguishable. The corresponding discrete stimulus set has elements $\widehat{S} = C^{-1}(\widehat{P})$. **(b)** The discrete mapping induced by quantization. On the vertical axis, quantization reduces the real line to a discrete set of points (indicated by crosses) called the perception dictionary. Because the scaling C is invertible, the quantization of perception also induces quantization in the stimulus intensity on the horizontal axis (indicated by circles). As a result of quantization, any two stimulus intensities in the gray region are indistinguishable.

resolution result in Corollary 6.3 to argue that as K becomes large for a random stimulus S bounded as $0 < s_0 \leq S \leq s_1 < \infty$, the relationship between C and f_S satisfies

$$\frac{dC(s)}{ds} \propto s^{-2/3} f_S^{1/3}(s) \qquad \text{for } s \in [s_0, s_1], \tag{7.3}$$

with constraints $C(s_0) = 0$ and $C(s_1) = 1$. Although (7.3) is an asymptotic results, we can show that the derivative of C does not charge much with the size of K assuming it is large enough.

Next, we address how the size of the perception dictionary affects ERE. If the set $\{\widehat{P}\}$ has K equally-spaced elements between zero and one, then

$$\widehat{P}_k = \frac{k - 1/2}{K}, \qquad k \in [1 : K].$$

The equivalent stimulus representation is simply $\widehat{S}_k = C^{-1}(\widehat{P}_k)$. As K increases, so does the resolution of perception, leading to a more accurate approximation of stimulus intensity and reducing the squared error factor in ERE. Using Corollary 6.1, ERE falls at the rate of K^2.

It turns out C is heavily dependent on f_S and less dependent on K. Moreover, C is also explicitly dependent on the error measure to be minimized, which is ERE in this analysis. Indeed, the optimal scale $C(s)$, stimulus pdf $f_S(s)$, and error measure are intertwined under quantization theory such that knowledge of any two can predict the third.

That the psychophysical scale C adapts to the statistical properties of the stimulus pdf f_S implies that C should change when f_S changes. Such plasticity would allow individuals to adapt to long-term changes in their perceptual environment. For example, this phenomenon has been observed for perception of sound intensity in individuals after long-term use of auditory prostheses that modify f_S [143, 179].

We briefly mention an analog model for limited perception; a more formal discussion is given in Appendix 7.A. This analog model leads to the same expression for the psychophysical scale that minimizes ERE, making the conclusions we draw from (7.3) not contingent on discretization in the encoding of stimuli. Suppose that rather than

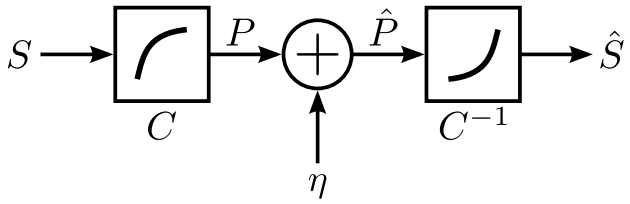


Figure 7-2: Block diagram of an analog model for perception that provides equivalent asymptotic results as the model in Figure 7-1. Stimulus intensity S is transformed by a nonlinear scaling C, resulting in perceptual intensity P. Biological constraints are modeled by limiting the range of C to $[0, 1]$ and by including additive noise η that is independent of S.

quantization, zero-mean additive noise corrupts P, leading to $\widehat{P}$ (Figure 7-2). If this additive noise is independent of the stimulus intensity and has variance of the same order as the perceptual uncertainty, then the Bayes-optimal C has the same form as (7.3) in the limit of high signal-to-noise ratio (SNR). Note that the notion of SNR here is at the cognitive level, which may be high even when the SNRs of single neurons are substantially lower.

7.3 A Bayes-Optimal Model for Limited Perception with Coding

We now formulate the second quantization model, motivated by stimuli for which psychophysical scales are not affected by sensing mechanisms or communication channels from sensory systems. One such example is numerosity, which has been shown to follow the Weber–Fechner law [29,128]; some recent studies of Amazonian tribes suggest that our instinctive number line is logarithmic [30, 145]. As an abstract sensation, numerosity is of particular interest since it does not suffer from physical limitations like resolution or saturation.[1] Note that small numbers may be qualitatively different in how they are perceived [96]. Since numbers can be observed directly with very fine precision, why should numerical perception suffer from the indistinguishability

[1]We must be careful to distinguish the perception of distinct objects and the vocalization of these quantities. This difference has been studied by Izard and Dehaene [76], who employ a simple quantization model.

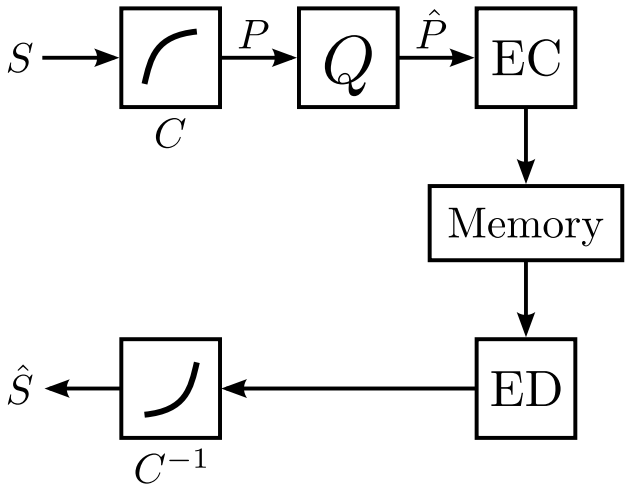


Figure 7-3: Block diagram of the quantization model with coding. An entropy coder (EC) is used to code quantized values $\widehat{P}$ and is then stored into memory. Later, the perception is fetched from memory, decoded using an entropy decoder (ED). The EC and ED steps include no noise, and no information is lost in these steps.

of quantization?

The reason may be *coding*. In the previous model, the quantized values were not represented more efficiently through coding. However, representing more likely values in the perception dictionary with compact representations leads to reduced overall information transmission or storage at the expense of increased computational complexity and delay. Entropy gives the fundamental limits of compressibility through coding [27]. Efficient entropy-based codes have been suggested for transmission of sensory information [37, 48], for short-term memory [16], and in the context of learning [8], where compressed representations may help meet memory capacity constraints [186].

In variable-rate quantization, values from the discrete set $\mathcal{P}$ (or equivalently, the set $\{\widehat{S}\}$) are entropy-coded based on the probabilities of occurrence of the entries in the set (Figure 7-3). As the result of coding, the best choice of C is no longer the same as in the previous model. The Bayesian optimization is now

$$\underset{C(s)}{\operatorname{argmin}}\ D_{\text{ere}}(K, C') \qquad \text{such that } H(\widehat{S}) < R, \tag{7.4}$$

where R is the communication/storage rate and H is the entropy function. For a

random stimulus S bounded as $0 < s_0 \leq S \leq s_1 < \infty$, the optimal scale for ERE has the property

$$\frac{dC(s)}{ds} \propto 1/s \qquad \text{for } s \in [s_0, s_1], \tag{7.5}$$

with additional normalization constraints $C(s_0) = 0$ and $C(s_1) = 1$ (see Corollary 6.4). Unlike in (7.3), there is no dependence on the pdf f_S in (7.5). The scaling that satisfies the above conditions is

$$C(s) = \frac{\ln(s/s_0)}{\ln(s_1/s_0)} \qquad \text{if } s \in [s_0, s_1], \tag{7.6}$$

for any pdf f_S. In fact, the scale is only dependent on the endpoint values of the distribution. For unbounded random variables, an analogous result holds under mild conditions on the decay of the tail of the pdf. Hence, quantization followed by efficient coding leads naturally to the logarithmic relationship in the Weber–Fechner law. This holds for all well-behaved stimulus distributions and the resolution can be tuned easily by the level of compression.

We suggest that coding is an essential part of perceiving numbers, which need not have distributions of a particular form. Furthermore, since the optimal psychophysical scales for entropy-coded representations do not depend on the statistical properties of the source, there is task-independence and constancy [48].

7.4 Examples

In this section, we connect the proposed Bayesian quantization models to the Weber–Fechner law. This relationship is clear in the second model since the logarithmic scale is optimal for all stimulus distributions, as shown in (7.6). However, in the first model, the optimized scale depends explicitly on the stimulus distribution through (7.3).

It is not obvious that the Weber–Fechner law will be realized using the first model. However, the sensations corresponding to many natural phenomena have statistical distributions that obey a power law over a range of intensities that are of behavioral interest [87, 110, 158, 209], and we will demonstrate that such distributions do yield a

logarithmic scale. As a case study, we present empirical evidence that the computed scalings of natural sounds are well approximated by the Weber–Fechner law.

7.4.1 Power-Law Stimulus Distributions

We begin by developing psychophysical scales for the Pareto distribution. Given parameters $\alpha > 0$ and $s_0 > 0$, the Pareto pdf is

$$f_S(s) = \frac{\alpha}{s_0}\left(\frac{s}{s_0}\right)^{-\alpha-1} \qquad \text{if } s \geq s_0; \quad \text{and 0 otherwise,} \tag{7.7}$$

where s_0 corresponds to the lowest perceivable stimulus intensity. The pdf decays with an exponent of $-(\alpha + 1)$ and intensity is not upper-bounded, i.e. there is no upper threshold to perception.

Using (7.3), the derivative of C is

$$\frac{dC(s)}{ds} = \frac{\alpha}{3s_0}\left(\frac{s}{s_0}\right)^{-\alpha/3-1} \qquad \text{if } s \geq s_0; \quad \text{and 0 otherwise.}$$

The psychophysical scale that results from the above expression and satisfies the boundary conditions is

$$C(s) = 1 - \left(\frac{s}{s_0}\right)^{-\alpha/3} \qquad \text{if } s \geq s_0\ . \tag{7.8}$$

In general, the psychophysical scales generated by (7.8) are concave on a logarithmic scale and hence are inconsistent with the Weber–Fechner law. However, a bounded pdf is more practical because there are usually lower and upper limits to what is perceivable. With parameters α, s_0 and s_1, the bounded power-law distribution is

$$f_S(s) \propto s^{-\alpha-1} \qquad \text{if } s \in [s_0, s_1]; \quad \text{and 0 otherwise,}$$

normalized to have unit integral. Here, s_0 and s_1 are the lower and upper thresholds of perception, yielding a more psychophysically reasonable model. Note that α is no longer restricted to be positive as opposed to in (7.7). Repeating the same analysis

as above, the derivative of C is

$$\frac{dC(s)}{ds} \propto s^{-\alpha/3-1} \qquad \text{if } s \in [s_0, s_1]; \quad \text{and 0 otherwise.} \tag{7.9}$$

For the special case of $\alpha = 0$, or equivalently an exponent of -1 in the decay of the pdf, (7.9) simplifies to

$$C(s) = \frac{\ln(s/s_0)}{\ln(s_1/s_0)} \qquad \text{if } s \in [s_0, s_1],$$

which is precisely the Weber–Fechner law. For other choices of α, the scaling is

$$C(s) = \frac{s^{-\alpha/3} - s_0^{-\alpha/3}}{s_1^{-\alpha/3} - s_0^{-\alpha/3}} \qquad \text{if } s \in [s_0, s_1] \ ,$$

providing a generalization to logarithmic scaling accounting for a large class of scales. Figure 7-4 demonstrates how three choices of α affect the Bayes-optimal C.

Thus, there is an intimate match between the Weber–Fechner law and a bounded power-law distribution. Indeed, such a distribution with $\alpha = 0$ matches precisely with logarithmic scaling. However, other exponents yield minor deviations which may also be observed experimentally.

7.4.2 Natural Sounds

The above results predict experimentally falsifiable psychophysical scales based on power-law stimulus distributions. In general, while natural stimuli may not be easily identified as exactly power-law distributed, many are approximately power-law over relevant ranges. One such example is the intensity of speech, which is often modeled as Gaussian-distributed on a dB scale (lognormal). Indeed, lognormal and power-law distributions are often empirically indistinguishable [122]. We test datasets of animal vocalizations and human speech and find the optimal psychophysical scale to be well-approximated by a logarithmic relationship where the intensity is most probable.

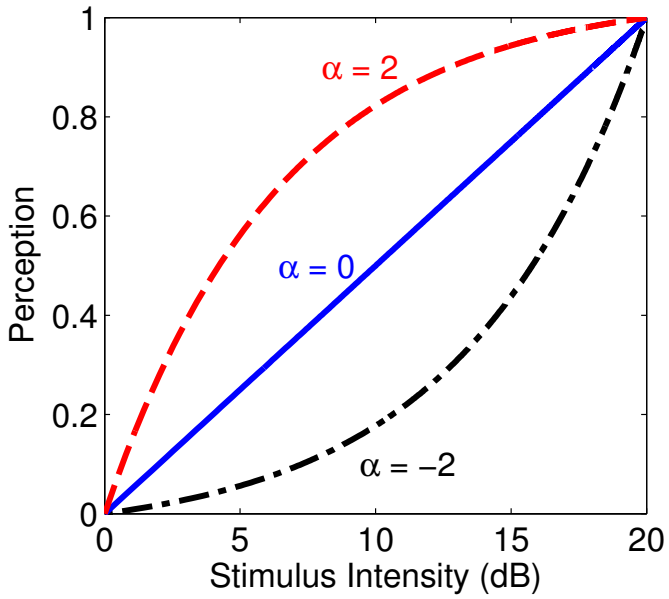


Figure 7-4: Sample scaling C for bounded power-law densities with three choices for α. Perception intensity on the vertical axis is normalized to lie between 0 and 1, corresponding to the smallest and largest perceivable stimulus intensities, respectively. Logarithmic scaling results when $\alpha = 0$.

Animal vocalizations and human speech comprise complex harmonic and transient components that convey behaviorally-relevant meaning. For example, many animals vocalize to convey information related to mating rituals, predator warnings, or the locations of food sources. The individuals that best process these sounds may be those that are most likely to survive and reproduce. In this way, the auditory system may have evolved to optimally process natural sounds like vocalizations in order to efficiently extract relevant acoustic cues. One such cue is the perceived intensity of a sound, or its loudness. In particular, the normal human ear perceives sound intensities with roughly 1 dB JND across nearly the entire range of perceivable levels [39], with only slight variation near the extremes, which is consistent with the logarithmic relationship in the Weber–Fechner law [44, 187].

To test our model, we employ two datasets comprising animal vocalizations and human speech sounds. The animal vocalization data (DS1) includes 55 rain forest mammals (33 minutes) taken from commercially available CDs [36]. The human speech data (DS2) corresponds to a male speaker reciting a corpus of 280 English

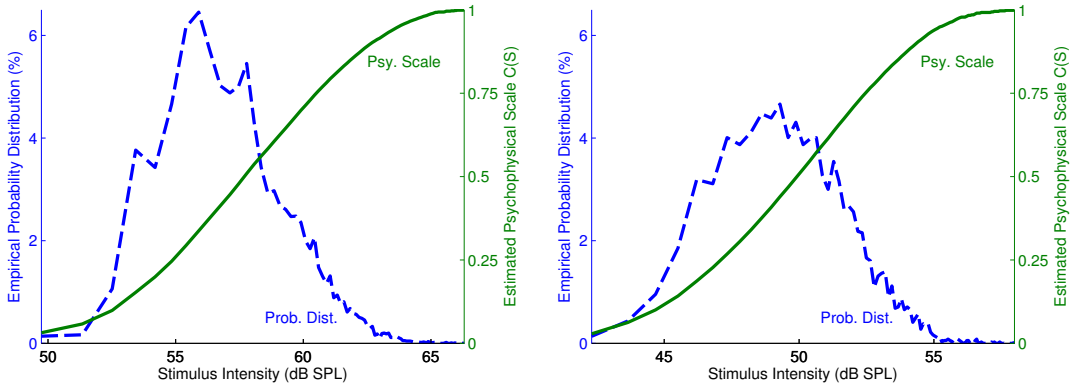


Figure 7-5: Empirical distribution of stimulus intensities (blue dashed line, left axis) and corresponding psychophysical scaling function, based on (7.3) (green solid line, right axis) for **(a)** mammalian vocalizations (DS1) and **(b)** human speech (DS2). The near linearity of the scaling functions in log scale indicates an approximation to the Weber–Fechner law.

sentences (8 minutes) [191].

Silence intervals, defined as intervals of 50 ms in which the signal did not exceed 10% of the maximum magnitude of the recording, were removed from the recordings. The resulting sound files were broken into successive intervals of 100 ms, and the root mean square (rms) was computed for each interval. The empirical sound level distributions of the rms values were used to compute $C(s)$.

For both DS1 and DS2, the predicted psychophysical scales are well-approximated by a straight line where the intensity is most probable (Figure 7-5); since the horizontal axis is logarithmic, this indicates a logarithmic relationship. Moreover, the deviation from a logarithmic scaling is most prominent at the extremes of the stimulus pdf, where experimental studies also demonstrate breakdown of the Weber–Fechner law [4, 118].

We also varied several parameters that affect the empirical distribution. These parameters included the thresholds and interval lengths used to calculate the silence intervals, lengths of the interval over which the rms values were computed, and histogram bin widths. To account for gain control mechanisms, we also used an rms gain parameter to horizontally shift the empirical distribution. We found that the scales induced by these parameter changes did not vary greatly and had similar goodness-

Table 7.1: Goodness-of-fit of linear regression of $C(s)$ over the "most likely" range of the empirical distribution using DS1. We vary the parameters that affect the empirical distribution and find that the model is robust to these changes. Using the parameters set in the left columns of the table, a linear regression is performed over a range that covers the desired percentage of the empirical pdf around the median and the goodness-of-fit (R^2) value is found. We see that the R^2 is close to one, especially where the stimulus intensity is most probable. As a point of comparison on the shape of the curves, we also present R^2 results for the erf function, a well-known sigmoid function that is closely related to the Gaussian distribution function. The erf is close to linear around its median but becomes more nonlinear as more of its range is considered, much like the empirical $C(s)$. The scaling law is more linear than the erf in all regimes.

Parameters					**Distribution Coverage**		
Silence Threshold	**Silence Length**	**RMS Length**	**RMS Gain**	**Histogram Bins**	**50%**	**90%**	**100%**
0.1	50 ms	100 ms	1	100	0.9998	0.9988	0.9635
0.01	50 ms	100 ms	1	100	0.9991	0.9921	0.9744
0.3	50 ms	100 ms	1	100	0.9995	0.9970	0.9658
0.1	25 ms	100 ms	1	100	0.9996	0.9988	0.9699
0.1	100 ms	100 ms	1	100	0.9998	0.9992	0.9641
0.1	50 ms	50 ms	1	100	0.9996	0.9987	0.9569
0.1	50 ms	200 ms	1	100	0.9999	0.9989	0.9629
0.1	50 ms	100 ms	10^3	100	0.9998	0.9988	0.9635
0.1	50 ms	100 ms	10^{-3}	100	0.9998	0.9988	0.9635
0.1	50 ms	100 ms	1	250	0.9998	0.9988	0.9590
0.1	50 ms	100 ms	1	1000	0.9997	0.9987	0.9438
				Erf function	0.9960	0.9867	0.9114

of-fit characteristics on a linear regression (Table 7.1).

To summarize, sound intensity perception scales determined from animal vocalization data and our optimality principles are consistent with the basic Weber–Fechner law.

7.5 Discussion

Through quantization frameworks for perception, we have determined that scaling laws observed in psychophysical studies are Bayes-optimal for expected relative error. Sensations that are measured by sensory mechanisms in the periphery and ones that

are abstract are both considered. Although they have different costs, both situations have optimized scalings that include the Weber–Fechner law. Moreover, other scaling laws may be consistent under the first model, and this theory provides an experimentally testable hypothesis for them.

7.5.1 Key Assumptions

There are several assumptions that anchor this framework. The first and most fundamental assumption is that the acquisition of information in the stimulus is Bayes-optimal at a computational level [114]. This is well-motivated since numerous studies suggest that there exist neural mechanisms for adaptation to the stimulus distribution as well as numerous feedback and feedforward channels [94, 191].

The second assumption is that perception is algorithmically discrete and the mapping is simply a deterministic function of the stimulus intensity, as illustrated in Figure 1. This framework is inspired by engineering systems and information-theoretic results on Bayes-optimal data acquisition. Although it is debatable whether this type of discretization occurs in individual neurons, quantization is plausible at the cognitive level. Moreover, we have described how equivalent asymptotic scalings occur with an analog coding scheme. In the context of this framework, the adaptation mechanisms discussed above better estimate the distribution shape and thresholds to precisely predict $C(s)$.

A third assumption is that the accuracy measure to minimize is ERE. We have motivated this choice through related research in numerical computation and perceptual coding, as well as the general perceptual principle of scale-invariance. However, ERE may be too simple to fully capture the goals of information acquisition and may provide only a crude first-order approximation. Using Proposition 6.2, we see that the psychophysical implications of this framework are robust within a large class of error criteria. More generally, the true cost may be a combination of several error measures weighted differently depending on behavioral context. In this case, our framework is still useful since the error measure, stimulus distribution and optimizing psychophysical scale are interrelated such that the observation of any two predicts

the third. Therefore one can predict the error measure optimized by the system using the experimentally observed stimulus distribution and psychophysical scale.

Finally, although the analysis and optimality results presented apply to any stimulus signal that is stationary in time (see Corollary 6.2), the scalar system model does not exploit the memory or correlation in time of natural stimuli. Our system model takes acquisition to be independent across intensity samples. Indeed, one could code over blocks of samples or apply a causal filter to improve compression, which may change the optimal choice of $C(s)$ [66]. However, such coding will increase signal acquisition delay. For external stimuli such as sound or vision, speed of perception may be more important than improved compression. For this reason, the models presented here are psychologically reasonable under latency constraints.

7.5.2 Applicability of Theory

We have show that the Weber–Fechner law arises as the Bayes-optimal mapping in two quantization frameworks. In the first model, when stimulus intensity is simply discretized, Weber–Fechner is intimately tied to stimulus distributions that decay as power-law functions. In the second, when discrete percepts are coded, the Weber–Fechner law becomes more general and is optimal for *all* statistical distributions. These results are dependent on several assumptions which are well-supported by neuroscientific and information scientific principles. This work points out at least three psychophysical ideas: the importance of stimulus statistics in interpreting scaling laws, the necessity of adaptation to stimulus distributions in neural circuitry, and the possibility of information-theoretically optimal acquisition structures at a cognitive level.

7.5.3 Future Directions

An interesting question is whether either of the two quantization models proposed here can be applied to the stimulus class of the other. For example, can coding occur for stimuli sensed by the periphery? Alternatively, many numerical quantities,

such as salaries or populations have been proposed to approximately follow Pareto distributions. Could numerosity scaling laws be logarithmic without the need for coding? We feel affirmative answers to these questions are less plausible than what is proposed here but they remain topics of worthwhile exploration.

We have argued that the framework provides falsifiable hypotheses for understanding psychophysical laws. It is therefore interesting to consider the scenarios where the Weber–Fechner law does not hold. One such scenario is vision, where the perception of contrast follows the Weber–Fechner law for high luminance, but that relationship breaks down in low light [142, 157]. It is of interest to study if the theory presented here can predict this breakdown and match the observed scaling for night vision.

7.A Mathematical Methods for Analog Model

In Section 7.2, we introduced a quantization model (QM) and determined the psychophysical scale that minimizes ERE for a given stimulus distribution. Here, we summarize an analog model (AM) based on suppression of additive noise and relate the optimization of the scale in the two models. The proofs for the results in QM have already been discussed in the thesis. We will focus solely on AM here (Figure 7-2).

Consider the error arising from AM when η is bounded noise independent of the stimulus and has variance $\sigma^2(\eta)$. The noisy stimulus intensities take the form $\widehat{S} = C^{-1}(C(S) + \eta)$. Because $C(s)$ is continuous and strictly monotonic (hence differentiable), we use Taylor's theorem to describe $\widehat{S}$ through a linear approximation. Taylor's theorem states a function $g(x)$ that is $n+1$ times continuously differentiable on a closed interval $[a, x]$ takes the form

$$g(x) = g(a) + \left\{ \sum_{k=1}^{n} \frac{g^{(k)}(a)}{k!} (x-a)^k \right\} + R_n(x, a),$$

with a Taylor remainder term

$$R_n(x, a) = \frac{g^{(n+1)}(\xi)}{(n+1)!} (x-a)^{n+1}$$

for some $\xi \in [a, x]$. Therefore, $\widehat{S} \approx W(C(S)) + W'(C(S))\eta$ for small η, where $W(s) = C^{-1}(s)$. By the definition of W, $W(C(s)) = s$ and $W'(C(s)) = 1/C'(s)$, so we can simplify the above equation to $\widehat{S} \approx S + \eta/\lambda(S)$.

Using (6.2) while noting the expectation is taken with respect to both S and η, the ERE in the AM is

$$\text{ERE} \approx \mathbf{E}[S^{-2}\lambda^{-2}(S)]\,\sigma^2(\eta),$$

with approximation error corresponding to the first-order Taylor remainder term. Since the remainder decays as η^2 while the ERE decays as η, we can be more mathematically precise about the behavior of the ERE.

Theorem 7.1. *Consider a stimulus intensity S following a stationary random process with probability density f_S that is smooth and positive on $\mathbb{R}$ (or a compact interval). The intensity is scaled through function C and then perturbed by independent additive noise η that is bounded and has variance $\sigma^2(\eta)$. The expected relative error between the stimulus and its noisy version satisfies*

$$\lim_{\sigma^2(\eta)\to 0} \text{ERE} \cdot \sigma^2(\eta) = \mathbf{E}[S^{-2}\lambda^{-2}(S)]. \tag{7.10}$$

Noting that the right-hand sides of (7.10) and the distortion result in Corollary 6.1 match, the choice of point density λ that minimizes ERE is the same for both. Hence, QM and AM have equivalent asymptotic psychophysical scales.

Chapter 8

Conclusion

With the advent of effective distributed storage and cloud computing systems, along with the proliferation of sensor networks, digital cameras and monitoring systems, data has been growing at a fast rate. This changing paradigm leads to shifts in the meaning of information in source coding contexts. Rather than consuming data sequentially and entirely, like in a video stream, more applications desire fidelity in statistics or other nonlinear low-dimensional mappings of the data. The information is now application-defined and domain-specific. Moreover, as networks grow and the nodes become diverse, ignorance to system parameters such as power consumption, synchronicity and latency at the source coding layer can lead to extremely suboptimal designs.

As the applications of distributed systems have multiplied and become more varied, many techniques that have been previously successful in bounding the performance of compression for point-to-point systems are no longer sufficient in analyzing network problems. Not only do many such methods, e.g. Shannon theory, become intractable quickly with the size and complexity of the network, but they can lead to misleading design principles. Hence, there is an opportunity for new mathematical frameworks and tools to understand information acquisition and compression in a network.

The main contribution of this thesis is the study of compression taking advantage of domain-specific knowledge, such as accounting for the functional demands of mod-

ern systems so that knowledge of the overall goals can be pushed into the acquisition block. One such framework is distributed functional scalar quantization (DFSQ), which was introduced in Chapter 2. We extended the space of problems for which DFSQ theory is meaningful in Chapter 3 and used the high-resolution asymptotic to provide insight into the behavior of the system when sensors are allowed to collaborate in Chapter 5. We also studied Shannon-theoretic formulations of distributed source coding in Chapter 4 and provided fundamental limits when a computation is desired at the fusion center in the regime where the allowed distortion is small.

A secondary goal of the thesis was to use source coding theory to help explain neuroscientific phenomena such as human perception. In Chapter 7, we borrowed theory developed in Chapter 6 to understand psychophysical laws as being Bayes-optimal for expected relative error under communication limitations in neural channels. Consequently, we argue that human perception is informationally efficient for signal acquisition.

We believe this thesis has promoted several general ideas that can spur further research:

1. The use of high-resolution quantization theory to study source coding problems that better represents real-world systems and provide sharp and meaningful performance guarantees that are analyzable.
2. The utility of intersensor collaboration to alleviate communication loads in a network, taking advantage of heterogeneous channels that may exist.
3. The brain is optimal in a Bayesian sense, and can robustly exploit the distributions of stimuli to better acquire and process information.

Below, we highlight particular problems where investigations may yield important contributions.

High-resolution quantization theory

Although high-resolution quantization theory (HRQT) has traditionally received less attention from the source coding community than the large-blocklength asymptotic, the needs of practical systems may lead to renewed interest. DFSQ is one promis-

ing model to study networks using HRQT, but there still remains many interesting problems to be considered. Of particular interest are generalization of existing results to vector quantizers or scalar quantizers with memory. We have discussed that transform and predictive coding are predominant in applications and standards, but folding in a functional aspect is nontrivial. Further extensions in this direction may have high impact in applications where temporal correlation is strong, e.g. video or speech. Another interesting network is the cascade network, where sensors pass information to each other in serial. This setting was recently explored in [121], but more general formulations are still unexplored.

Applying DFSQ to real-world systems

A goal of the thesis is to influence design of compression protocols for sensor networks. We have highlighted the constructive nature of the results and fast convergence of finite-rate quantizers to the high-resolution prediction. However, precise characterization of the suboptimality of the high-resolution estimate for finite codebook size is still unknown. It is of great interest to understand the loss from companding designs to what is achieved using iterative methods such as the Lloyd–Max algorithm. Even if precise bounds may be difficult to develop, computational studies may still be useful. One possible type of analysis is to start with quantizers proposed by HRQT and quantifying how different the quantizers become after running Lloyd–Max. Extensions to the network setting can then be devised by combining the DFSQ results with [43].

Another direction of future work is on the design of analog-to-digital converters (ADCs) to incorporate companding directly. In practice, ADC design is primarily focused on making the quantizer as linear as possible while getting the fastest sampling rate possible. If the eventual goal of the application is nonuniform quantizers, tunable ADCs may have high value. This is especially true in low-power sensors.

Neuroscientific inquiry

In the thesis, we present a simple Bayesian model that is harmonious with psychophysical behavior. This adds to the body of work that suggests that humans brains are very effective at processing, communicating and acquiring information. The relationship between neuroscience and information theory has a long and rich history [13, 31, 81], but it is uncertain whether the theory is explanatory, especially since ingredients such as long block lengths and maximum-likelihood estimation are infeasible.[1] We believe that quantization theory has the potential of being more useful in neuroscience because its asymptotics are simpler to satisfy or approximate, and because the regularity conditions it requires match real-world signals well.

[1]The danger of imposing mathematical beauty to problems where they may not exist was eloquently discussed by Shannon himself [163].

Bibliography

[1] A. Aaron, S. Rane, and B. Girod, "Wyner–Ziv video coding with hash-based motion compensation at the receiver," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Singapore, Oct. 2004, pp. 837–841.

[2] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.

[3] E. Akyol and K. Rose, "On constrained randomized quantization," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Apr. 2012, pp. 72–81.

[4] W. H. Atkinson, "A general equation for sensory magnitude," *Perception & Psychophysics*, vol. 31, no. 1, pp. 26–40, Jan. 1982.

[5] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4905–4918, Oct. 2008.

[6] V. Back and R. Seiler, "Analysis of optimal high resolution and fixed rate scalar quantization," *IEEE Trans. Inform. Theory*, vol. 55, no. 4, pp. 1683–1691, Apr. 2009.

[7] R. G. Baraniuk, "More is less: Signal processing and the data deluge," *Science*, vol. 331, no. 6018, pp. 717–719, Feb. 2011.

[8] H. B. Barlow, "Unsupervised learning," *Neural Computation*, vol. 1, no. 3, pp. 295–311, Fall 1989.

[9] G. R. Benitz and J. A. Bucklew, "Asymptotically optimal quantizers for detection of i.i.d. data," *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 316–325, Mar. 1989.

[10] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 446–472, Jul. 1948.

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

[12] T. Berger and J. D. Gibson, "Lossy source coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2693–2723, Oct. 1998.

[13] T. Berger and W. B. Levy, "A mathematical theory of energy efficient neural computation and communication," *IEEE Trans. Inform. Theory*, vol. 56, no. 2, pp. 852–874, Feb. 10.

[14] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[15] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nat. Neurosci.*, vol. 2, no. 11, pp. 947–957, Nov. 1999.

[16] T. F. Brady, T. Konkle, and G. A. Alvarez, "Compression in visual working memory: Using statistical regularities to form more efficient memory representations," *J. Exp. Psychol. Gen.*, vol. 138, no. 4, pp. 487–502, Nov. 2009.

[17] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with rth power distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 1, pp. 239–247, Mar. 1982.

[18] A. Buzo, A. H. Gray, Jr., R. M. Gray, and J. D. Markel, "Speech coding based upon vector quantization," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-28, pp. 562–574, Oct. 1980.

[19] S. Cambanis and N. L. Gerr, "A simple class of asymptotically optimal quantizers," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 5, pp. 664–676, Sep. 1983.

[20] J. Chen, D.-K. He, A. Jagmohan, and L. A. Lastras-Montaño, "On universal variable-rate Slepian–Wolf coding," in *Proc. IEEE Int. Conf. Commun.*, Beijing, China, May 2008, pp. 1426–1430.

[21] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 31–42, Jan. 1989.

[22] F. W. Cope, "Derivation of the Weber-Fechner law and the Loewenstein equation as the steady-state response of an Elovich solid state biological system," *Bull. Math. Biol.*, vol. 38, no. 2, pp. 111–118, Mar. 1976.

[23] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. Cambridge, MA: MIT Press, 2009.

[24] T. Courtade and R. Wesel, "Efficient universal recovery in broadcast networks," in *Proc. 48th Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2010, pp. 1542–1549.

[25] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," arXiv:1110.3069v3 [cs.IT]., Jul. 2012.

[26] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.

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

[28] E. Csiszár and J. Kőrner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[29] S. Dehaene, "The neural basis of the Weber–Fechner law: a logarithmic mental number line," *Trends in Cogn. Sci.*, vol. 7, no. 4, pp. 145–147, 2003.

[30] S. Dehaene, V. Izard, E. Spelke, and P. Pica, "Log or linear? Distinct intuitions of the number scale in western and amazonian indigene cultures," *Science*, vol. 320, no. 5880, pp. 1217–1220, May 2008.

[31] A. G. Dimitrov, A. A. Lazar, and J. D. Victor, "Information theory in neuroscience," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 1–5, feb 2011.

[32] V. Doshi, D. Shah, and M. Médard, "Source coding with distortion through graph coloring," in *Proc. IEEE Int. Symp. Inform. Theory*, Jun. 2007, pp. 1501–1505.

[33] S. C. Draper, "Universal incremental Slepian–Wolf coding," in *Proc. 42nd Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sep. 2004.

[34] A. El-Gamal and Y. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[35] N. Elia and S. K. Mitter, "Stabilization of linear systems with limited information," *IEEE Trans. Autom. Control*, vol. 46, no. 9, pp. 1384–1400, Sep. 2001.

[36] L. H. Emmons, B. M. Whitney, and D. L. Ross Jr., *Sounds of Neotropical Rainforest Mammals*. Ithaca, NY: Cornell Laboratory of Ornithology, 1997.

[37] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck, "Efficiency and ambiguity in an adaptive neural code," *Nature*, vol. 412, no. 6849, pp. 787–792, Aug. 2001.

[38] J. Fang and H. Li, "Distributed consensus with quantized data via sequence averaging," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 944–948, Feb. 2010.

[39] R. R. Fay, *Hearing in Vertebrates: a Psychophysics Databook*. Hill-Fay Associates, 1988.

[40] G. T. Fechner, *Elemente der Psychophysik*. Leipzig: Breitkopf & Härtel, 1860.

[41] S. Feizi and M. Medard, "On network functional compression," arXiv:1011.5496v2 [cs.IT]., Nov. 2010.

[42] H. Feng, M. Effros, and S. A. Savari, "Functional source coding for networks with receiver side information," in *Proc. 42nd Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2004, pp. 1419–1427.

[43] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1584–2004, Aug. 2004.

[44] M. Florentine, S. Buus, and C. R. Mason, "Level discrimination as a function of level for tones from 0.25 to 16 kHz," *J. Acoust. Soc. Am.*, vol. 81, no. 5, pp. 1528–1541, May 1987.

[45] K. Friston, "The free-energy principle: a unified brain theory?" *Nat. Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, Feb. 2010.

[46] G. S. Fullerton and J. M. Cattell, *On the Perception of Small Differences*. Philadelphia, PA: UPenn Press, 1892.

[47] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[48] D. Ganguli and E. P. Simoncelli, "Implicit encoding of prior probabilities in optimal neural populations," in *Adv. Neural Inform. Proc. Sys.*, Vancouver, Canada, Dec. 2010, pp. 658–666.

[49] J. Garcia-Frias and Y. Zhao, "Near-Shannon/Slepian–Wolf performance for unknown correlated sources over AWGN channels," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 555–559, Apr. 2005.

[50] W. R. Gardner and B. D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 5, pp. 367–381, Sep. 1995.

[51] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen–Loève transform," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.

[52] A. Gersho, "Quantization," *IEEE Commun. Soc. Mag.*, vol. 15, no. 5, pp. 16–29, Sep. 1977.

[53] ——, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 4, pp. 373–380, Jul. 1979.

[54] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer Acad. Pub., 1992.

[55] G. A. Gescheider, *Psychophysics: The Fundamentals*, 3rd ed. Psychology Press, 1997.

[56] H. Gish and J. P. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, no. 5, pp. 676–683, Sep. 1968.

[57] T. J. Goblick and J. L. Holsinger, "Analog source digitalization A comparison of theory and practice," *IEEE Trans. Inform. Theory*, vol. IT-13, no. 2, pp. 323–326, Apr. 1967.

[58] V. K. Goyal, "Beyond traditional transform coding," Ph.D. dissertation, Univ. California, Berkeley, 1998, published as Univ. California, Berkeley, Electron. Res. Lab. Memo. No. UCB/ERL M99/2, Jan. 1999.

[59] ——, "High-rate transform coding: How high is high, and does it matter?" in *Proc. IEEE Int. Symp. Inform. Theory*, Sorrento, Italy, Jun. 2000, p. 207.

[60] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Berlin: Springer, 2000.

[61] R. M. Gray, "Quantization noise spectra," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1220–1244, Nov. 1990.

[62] R. M. Gray, S. Boyd, and T. D. Lookabaugh, "Low rate distributed quantization of noisy observations," in *Proc. 23rd Annu. Allerton Conf. Commun. Control Comput.*, Oct. 1985.

[63] R. M. Gray and A. H. Gray, Jr., "Asymptotically optimal quantizers," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 1, pp. 143–144, Feb. 1977.

[64] R. M. Gray and E. D. Karnin, "Multiple local optima in vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 256–261, Mar. 1982.

[65] R. M. Gray, T. Linder, and J. Li, "A lagrangian formulation of Zadors entropy-constrained quantization theorem," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 695–707, Mar. 2002.

[66] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[67] R. M. Gray and T. G. Stockham, Jr., "Dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.

[68] R. Gupta and A. O. Hero, III, "High-rate vector quantization for detection," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1951–1969, Aug. 2003.

[69] A. György, T. Linder, P. A. Chou, and B. J. Betts, "Do optimal entropy-constrained quantizers have a finite or infinite number of codewords?" *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 3031–3037, Nov. 2003.

[70] T. S. Han and K. Kobayashi, "A dichotomy of functions $F(X, Y)$ of correlated sources (X, Y) from the viewpoint of the achievable rate region," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 1, pp. 69–76, Jan. 1987.

[71] J. J. Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 3, pp. 289–296, Sep. 1963.

[72] D. Hui and D. L. Neuhoff, "Asymptotic analysis of optimal fixed-rate uniform scalar quantization," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 957–977, Mar. 2001.

[73] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *Proc. IEEE Data Compression Conf.*, Snowbird, Utah, Mar. 2011, pp. 53–62.

[74] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Proc. 6th Int. Cong. Acoust.*, 1968, pp. C17–C20.

[75] ITU-T, "Pulse code modulation (PCM) of voice frequencies," G.711, 1988.

[76] V. Izard and S. Dehaene, "Calibrating the mental number line," *Cognition*, vol. 106, no. 3, pp. 1221–1247, Mar. 2008.

[77] R. A. Jacobs and J. K. Kruschke, "Bayesian learning theory applied to human cognition," *Wiley Interdisciplinary Rev. Cogn. Sci.*, vol. 2, no. 1, pp. 8–21, Jan./Feb. 2011.

[78] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385–1422, Oct. 1993.

[79] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood-Cliffs, NJ: Prentice-Hall, 1984.

[80] D. Jimenez, L. Wang, and Y. Wang, "White noise hypothesis for uniform quantization errors," *SIAM J. Math. Anal.*, vol. 38, no. 6, pp. 2042–2056, 2007.

[81] D. H. Johnson, "Dialogue concerning neural coding and information theory," 2003.

[82] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing de-quantization with applications to compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6270–6281, Dec. 2012.

[83] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, Mar. 2010.

[84] A. Kashyap, T. Başar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, no. 7, pp. 1192–1203, Jul. 2007.

[85] A. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 735–740, Nov. 1985.

[86] A. H. Kaspi and T. Berger, "Rate–distortion for correlated sources with partially separated encoders," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 6, pp. 828–840, Nov. 1982.

[87] C. T. Kello *et al.*, "Scaling laws in cognitive sciences," *Trends in Cogn. Sci.*, vol. 14, no. 5, pp. 223–232, may 2010.

[88] Y. H. Kim and A. Ortega, "Maximum a posteriori MAP-based algorithm for distributed source localization using quantized acoustic sensor readings," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, Toulouse, France, May 2006, pp. 1520–6149.

[89] D. C. Knill and A. Pouget, "The Bayesian brain: the role of uncertainty in neural coding and computation," *Trends Neurosci.*, vol. 27, no. 12, pp. 712–719, Dec. 2004.

[90] E. Kofman, "Relative error control in quantization based integration," in *XII Reunión Trabajo Proc. Info. Control*, Oct. 2007.

[91] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Proc. 6th Prague Conf. Info. Theory*, 1973, pp. 411–425.

[92] V. Kostina and S. Verdu, "Fixed-length lossy compression in the finite block-length regime," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[93] L. E. Krueger, "Reconciling Fechner and Stevens: Toward a unified psychophysical law," *Behav. Brain Sci.*, vol. 12, no. 2, pp. 251–267, Jun. 1989.

[94] V. A. F. Lamme, H. Supèr, and H. Spekreijse, "Feedforward, horizontal, and feedback processing in the visual cortex," *Curr. Opin. Neurobiol.*, vol. 8, no. 4, pp. 529–535, Aug. 1998.

[95] S. B. Laughlin and T. J. Sejnowski, "Communication in neuronal networks," *Science*, vol. 301, no. 5641, pp. 1870–1874, Sep. 2003.

[96] M. Le Corre and S. Carey, "One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles," *Cognition*, vol. 105, no. 2, pp. 395–438, Nov. 2007.

[97] J. Li, N. Chaddha, and R. M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," *IEEE Trans. Inform. Theory*, vol. 45, no. 4, pp. 1082–1091, May 1999.

[98] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, no. 1, pp. 84–95, Jan. 1980.

[99] T. Linder, "On asymptotically optimal companding quantization," *Prob. Contr. Inform. Theory*, vol. 20, no. 6, pp. 475–484, 1991.

[100] T. Linder and R. Zamir, "On the asymptotic tightness of the Shannon lower bound," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 2026–2031, Nov. 1994.

[101] ——, "High-resolution source coding for non-difference distortion measures: The rate–distortion function," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 533–547, Mar. 1999.

[102] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: Multidimensional companding," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 548–561, Mar. 1999.

[103] Y. N. Linkov, "Evaluations of epsilon entropy of random variables for small epsilon," *Prob. Info. Transm.*, vol. 1, pp. 12–18, 1965.

[104] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982, originally an unpublished Bell Telephone Laboratories tech. memo., July 31, 1957.

[105] T. D. Lookabaugh and R. M. Gray, "High resolution quantization theory and the vector quantization advantage," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1020–1033, Sep. 1989.

[106] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6180–6195, Sep. 2011.

[107] N. Ma, P. Ishwar, and P. Gupta, "Interactive source coding for function computation in collocated networks," *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4289–4305, Jul. 2012.

[108] D. M. MacKay, "Psychophysics of perceived intensity: A theoretical basis for Fechner's and Stevens' laws," *Science*, vol. 139, no. 3560, pp. 1213–1216, Mar. 1963.

[109] G. Maierbacher and J. Barros, "Low-complexity coding and source-optimized clustering for large-scale sensor networks," *ACM Trans. Sensor Networks*, vol. 5, no. 3, p. Article 24, May 2009.

[110] B. B. Mandelbrot, *The Fractal Geometry of Nature*. W. H. Freeman, 1982.

[111] S. Marano, V. Matta, and P. Willett, "Asymptotic design of quantizers for decentralized MMSE estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5485–5496, Nov. 2007.

[112] D. Marco and M. Effros, "On lossless coding with coded side information," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3284–3296, Jul. 2009.

[113] D. Marco and D. L. Neuhoff, "The validity of the additive noise model for uniform scalar quantization," *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1739–1755, May 2005.

[114] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY: W. H. Freeman, 1982.

[115] E. Martinian, G. W. Wornell, and R. Zamir, "Source coding with distortion side information," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4638–4665, Oct. 2008.

[116] T. Matsuta, T. Uyematsu, and R. Matsumoto, "Universal Slepian–Wolf source codes using low-density parity-check matrices," in *Proc. IEEE Int. Symp. Inform. Theory*, Austin, TX, Jun. 2010, pp. 186–190.

[117] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, no. 1, pp. 7–12, Mar. 1960.

[118] R. L. McBride, "Category scales of sweetness are consistent with sweetness-matching data," *Perception & Psychophysics*, vol. 34, no. 2, pp. 175–179, Mar. 1983.

[119] V. Misra, V. K. Goyal, and L. R. Varshney, "High-resolution functional quantization," in *Proc. IEEE Data Compression Conf.*, Snowbird, Utah, Mar. 2008, pp. 113–122.

[120] ——, "Distributed scalar quantization for computing: High-resolution analysis and extensions," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 5298–5325, Aug. 2011.

[121] V. Misra and K. Viswanathan, "Sequential functional quantization," Jan. 2013, in preparation.

[122] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Math.*, vol. 1, no. 2, pp. 226–251, 2004.

[123] P. W. Moo and D. L. Neuhoff, "Optimal compressor functions for multidimensional companding," in *Proc. IEEE Int. Symp. Inform. Theory*, Ulm, Germany, Jun. 1997, p. 515.

[124] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. San Diego, CA: Academic Press, 1992.

[125] G. Motta, F. Rizzo, and J. A. Storer, *Hyperspectral Data Compression*. New York, NY: Springer, 2006.

[126] A. Nedić, A. Olshevksy, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. Automat. Control*, vol. 54, no. 11, pp. 2506–2517, Nov. 2009.

[127] D. L. Neuhoff, "The other asymptotic theory of lossy source coding," in *Coding and Quantization*, ser. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, R. Calderbank, G. D. Forney, Jr., and N. Moayeri, Eds. American Mathematical Society, 1993, vol. 14, pp. 55–65.

[128] A. Nieder and E. K. Miller, "Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex," *Neuron*, vol. 37, no. 1, pp. 149–157, Jan. 2003.

[129] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[130] S. Nitinawarat and P. Narayan, "Perfect omniscience, perfect secrecy, and Steiner tree packing," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6490–6500, Dec. 2010.

[131] B. M. Oliver, J. Pierce, and C. E. Shannon, "The philosophy of PCM," *Proc. IRE*, vol. 36, pp. 1324–1331, Nov. 1948.

[132] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Curr. Opin. Neurobiol.*, vol. 14, no. 4, pp. 481–487, Aug. 2004.

[133] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.

[134] ——, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[135] ——, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.

[136] Y. Oohama and T. S. Han, "Universal coding for the Slepian–Wolf data compression system and the strong converse theorem," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1908–1919, Nov. 1994.

[137] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[138] A. Orlitsky, "Worst-case interactive communication. I: Two messages are almost optimal," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 1111–1126, Sep. 1990.

[139] ——, "Worst-case interactive communication. II. Two messages are not optimal," *IEEE Trans. Inform. Theory*, vol. 37, no. 4, pp. 995–1005, Jul. 1991.

[140] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[141] P. F. Panter and W. Dite, "Quantizing distortion in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, vol. 39, pp. 44–48, Jan. 1951.

[142] E. Peli, J. Yang, R. Goldstein, and A. Reeves, "Effect of luminance on suprathreshold constrast perception," *J. Opt. Soc. Amer.*, vol. 8, no. 8, pp. 1352–1359, Aug. 1991.

[143] B. Philibert, L. Collet, J.-F. Vesson, and E. Veuillet, "Intensity-related performances are modified by long-term hearing aid use: a functional plasticity?" *Hearing Research*, vol. 165, no. 1-2, pp. 142–151, Mar. 2002.

[144] S. T. Piantadosi, "Approximate number from first principles," 2012.

[145] P. Pica, C. Lerner, V. Izard, and S. Dehaene, "Exact and approximate arithmetic in an amazonian indigene group," *Science*, vol. 306, no. 5695, pp. 499–503, Oct. 2004.

[146] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[147] H. V. Poor, "High-rate vector quantization for detection," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 960–972, Sep. 1988.

[148] R. D. Portugal and B. F. Svaiter, "Weber-Fechner law and the optimality of the logarithmic scale," *Minds & Mach.*, vol. 21, no. 1, pp. 73–81, Feb. 2011.

[149] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE Int. Symp. Inform. Theory*, Chicago, IL, Jul. 2004, p. 117.

[150] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Sig. Process. Mag.*, vol. 19, no. 2, pp. 51–60, Mar. 2002.

[151] M. Pugh and B. D. Rao, "Distributed quantization of order statistics with applications to csi feedback," in *Proc. IEEE Data Compression Conf.*, Snowbird, Utah, Mar. 2011, pp. 323–332.

[152] R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," in *Proc. 40th Allerton Conf. on Commun., Control, and Computing*, Allerton, IL, Oct. 2002.

[153] D. Rebollo-Monedero and B. Girod, "Network distributed quantization," in *Proc. Inform. Theory Workshop*, Tahoe City, CA, Sep. 2007, pp. 426–431.

[154] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2003, pp. 13–22.

[155] J. B. Rhim and V. K. Goyal, "Keep ballots secret: On the futility of social learning in decision making by voting," arXiv:1212.5855v1 [cs.IT]., Dec. 2012.

[156] ——, "Social teaching: Being informative vs. being right in sequential decision making," arXiv:1212.6592v1 [cs.IT]., Dec. 2012.

[157] F. Rieke and M. E. Rudd, "The challenges natural images pose for visual adaptation," *Neuron*, vol. 64, no. 5, pp. 605–616, Dec. 2009.

[158] D. Rocchesso and F. Fontana, *The Sounding Object*. Firenze: Mondo Estremo, 2003.

[159] H. L. Royden and P. M. Fitzpatrick, *Real Analysis*, 4th ed. Boston, MA: Prentice-Hall, 2010.

[160] M. Sefidgaran and A. Tchamkerten, "On cooperation in multi-terminal computation and rate distortion," in *Proc. IEEE Int. Symp. Inform. Theory*, Jul. 2012, pp. 771–775.

[161] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 2, pp. 162–169, Mar. 1976.

[162] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948, continued 27:623–656, October 1948.

[163] ——, "The bandwagon," *IRE Trans. Inform. Theory*, vol. IT-2, no. 1, p. 3, Mar. 1956.

[164] ——, "Coding theorems for a discrete source with a fidelity criterion," *IRE Int. Conv. Rec., part 4*, vol. 7, pp. 142–163, 1959, reprinted with changes in *Information and Decision Processes*, ed. R. E. Machol, McGraw-Hill, New York, 1960, pp. 93–126.

[165] G. Shen and A. Ortega, "Transform-based distributed data gathering," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3802–3815, Jul. 2010.

[166] J. Shen, "On the foundations of vision modeling: I. Weber's law and Weberized TV restoration," *Physica D: Nonlinear Phenomena*, vol. 175, no. 3–4, pp. 241–251, Apr. 2003.

[167] S. F. Simon, "On suboptimal multidimensional companding," in *Proc. IEEE Data Compression Conf.*, Snowbird, Utah, Mar. 1998, pp. 438–447.

[168] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[169] S. S. Stevens, "On the psychophysical law," *Psych. Rev.*, vol. 64, no. 3, pp. 153–181, May 1957.

[170] ——, "To honor Fechner and repeal his law," *Science*, vol. 133, no. 3446, pp. 80–86, Jan. 1961.

[171] M. Studeny and J. Vejnarova, "The multiinformation function as a tool for measuring stochastic dependence," *Learning in Graphical Models*, pp. 261–297, 1998.

[172] J. Z. Sun and V. K. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, Jun.–Jul. 2009, pp. 6–10.

[173] ——, "Scalar quantization for relative error," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2011, pp. 293–302.

[174] ——, "Chatting in distributed quantization networks," in *Proc. 50th Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Oct. 2012, pp. 2045–2052.

[175] ——, "Intersensor communication in distributed quantization networks," *IEEE Trans. Commun.*, 2013, in review.

[176] J. Z. Sun, V. Misra, and V. K. Goyal, "Distributed functional scalar quantization simplified," *IEEE Trans. Signal Process.*, 2013, accepted.

[177] J. Z. Sun, G. I. Wang, V. K. Goyal, and L. R. Varshney, "A framework for Bayesian optimality of psychophysical laws," *J. Math. Psych.*, vol. 56, no. 6, pp. 495–501, Dec. 2012.

[178] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," arXiv:1201.3901v2 [cs.IT]., Feb. 2012.

[179] H. Thai-Van, B. Philibert, E. Veuillet, and L. Collet, "Assessment of auditory plasticity using psychoacoustic and electrophysiological measurements," *Audiological Medicine*, vol. 7, no. 1, pp. 55–66, 2009.

[180] A. V. Trushkin, "Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 187–198, Mar. 1982.

[181] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[182] S. Tung, "Multiterminal source coding," Ph.D. dissertation, Cornell University, 1978.

[183] D. Varodayan, "Adaptive distributed source coding," Ph.D. dissertation, Stanford Univ., Mar. 2010.

[184] K. R. Varshney and L. R. Varshney, "Quantization of prior probabilities for hypothesis testing," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4553–4562, Oct. 2008.

[185] L. R. Varshney and J. Z. Sun, "Why do we perceive logarithmically?" *Significance*, vol. 10, no. 1, pp. 28–31, Feb. 2013.

[186] L. R. Varshney, P. J. Sjöström, and D. B. Chklovskii, "Optimal information storage in noisy synapses under resource constraints," *Neuron*, vol. 52, no. 3, pp. 409–423, Nov. 2006.

[187] N. F. Viemeister, "Auditory intensity discrimination at high frequencies in the presence of noise," *Science*, vol. 221, no. 4616, pp. 1206–1208, Sep. 1983.

[188] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.

[189] H. Viswanathan and R. Zamir, "On the whiteness of high-resolution quantization errors," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 2029–2038, Jul. 2001.

[190] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-terminal source-coding problem," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[191] B. Wen, G. I. Wang, I. Dean, and B. Delgutte, "Dynamic range adaptation to sound level statistics in the auditory nerve," *J. Neurosci.*, vol. 29, no. 44, pp. 13 797–13 808, Nov. 2009.

[192] B. Widrow and I. Kollár, *Quantization Noise*. Cambridge, MA: Cambridge, 2008.

[193] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-16, no. 4, pp. 406–411, Jul. 1970.

[194] A. D. Wyner, "Recent results in Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 1, pp. 2–10, Jan. 1974.

[195] ——, "The rate-distortion function for source coding with side information at the decoder – II: General sources," *Inform. and Control*, vol. 38, no. 1, pp. 60–80, Jul. 1978.

[196] A. D. Wyner and J. Ziv, "The rate–distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[197] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Sig. Process. Mag.*, vol. 21, no. 5, pp. 80–94, May 2004.

[198] Y. Yamada, S. Tazaki, and R. M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 1, pp. 6–14, Jan. 1980.

[199] H. Yamamoto, "Wyner–Ziv theory for a general function of the correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 5, pp. 803–807, Sep. 1982.

[200] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *Trans. IECE Japan*, vol. E63, no. 10, pp. 700–706, Oct. 1980.

[201] Y. Yang, Y. Zhang, and Z. Xiong, "On the sum-rate loss of quadratic Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 5588–5614, Sep. 2011.

[202] M. E. Yildiz and A. Scaglione, "Coding with side information for rate-constrained consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3753–3764, Aug. 2008.

[203] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Comm. Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[204] P. L. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," Ph.D. dissertation, Stanford Univ., 1963.

[205] ——, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 139–148, Mar. 1982.

[206] R. Zamir, "The rate loss in the Wyner–Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.

[207] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 106–117, Jan. 1999.

[208] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[209] G. K. Zipf, *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949.