Distributed Functional Scalar Quantization Simplified

John Z. Sun, *Student Member, IEEE*, Vinith Misra, and Vivek K Goyal, *Senior Member, IEEE*

***Abstract*—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. 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 features decoupled communication and computation blocks. Moreover, we extend the DFSQ framework with the simpler decoder to source distributions with unbounded support. Finally, through simulation results, we demonstrate that performance at moderate coding rates is well predicted by the asymptotic analysis, and we give new insight on the rate of convergence.**

***Index Terms*—Asymptotic quantization theory, distributed source coding, functional source coding, data compression, coding for computing.**

I. Introduction

FUNCTIONAL source coding techniques are of great importance in modern distributed systems such as sensor networks and cloud computing architectures because the fidelity of acquired data can greatly impact the accuracy of computations made with that data. In this work, we provide theoretical and empirical results for quantization in distributed systems with communication topologies described by Fig. 1. Here, N memoryless sources produce scalar realizations $X_1^N = (X_1, \ldots, X_N)$ from a joint distribution $f_{X_1^N}$ at each discrete time instant. These measurements are compressed by separate encoders and then sent to a central decoder that approximates a computation on the original data; the computation may be the identity function, meaning that the acquired samples themselves are to be reproduced.

There has been substantial effort to study distributed coding using information theoretic concepts, taking advantage of large block lengths 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 low latency requirements.

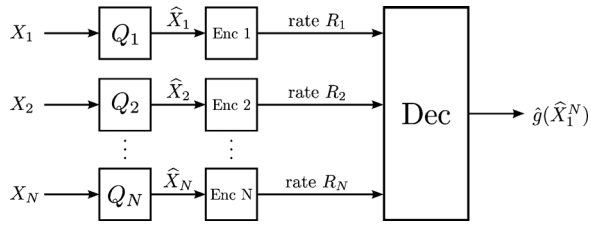


Fig. 1. A distributed computation network, where each of N spatially-separated sources generate a scalar X_n. The scalars are encoded and communicated over rate-limited links to a central decoder without interaction between encoders. The decoder computes an estimate of the function $g(X_1^N) = g(X_1, X_2, \ldots, X_N)$ from the received data using $\hat{g}(\widehat{X}_1^N)$. Each encoder is allowed transmission rate R_n.

Rather than have long blocks, the complementary asymptotic of *high-resolution quantization theory* [1] 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 Fig. 1 is the *distributed functional scalar quantization* (DFSQ) framework [2]. The key message from this previous work is that the design of optimal encoders for systems that perform nonlinear computations can be drastically different from what traditional quantization theory suggests. In recent years, ideas from DFSQ have been applied to compressed sensing [3], compression for media [4], and channel state feedback in wireless networks [5].

Like information theoretic approaches, the existing DFSQ theory relies in principle on a complicated decoder; this is reviewed in Section II-C. The primary contribution of this paper is to study a DFSQ framework that employs a simpler decoder. Remarkably, the same asymptotic performance is obtained with the simpler decoder, so the optimization of quantizer point densities is unchanged. Furthermore, the simplified framework allows a greater decoupling or modularity between communication (source encoding/decoding) and computation aspects of the network.

The analysis presented here uses different assumptions on the source distributions and function than [2]—neither is uniformly more or less restrictive. Unlike in [2], 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 papers contain rather technical conditions, and together they suggest a rather general applicability of DFSQ theory. We begin in Section II by

Manuscript received June 06, 2012; revised December 13, 2012; accepted March 07, 2013. Date of publication April 23, 2013; date of current version June 21, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Prakash Ishwar. This material is based upon work supported by the National Science Foundation under Grants 0643836, 0729069, and 1115159.

J. Z. Sun is with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: johnsun@mit.edu).

V. Misra is with the Department of Electrical Engineering and the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA (e-mail: vinith@stanford.edu).

V. K Goyal is with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: v.goyal@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2259483

reviewing relevant previous work and summarizing the contributions of this paper. In Sections III and IV, we give distortion analysis and optimal quantizer design results. Finally, we provide examples to demonstrate convergence in Section V and conclude in Section VI.

II. Preliminaries

A. Previous Work

The distributed network shown in Fig. 1 is of great interest to the information theory and communications communities, and there exists a variety of results corresponding to different scenarios of interest. We present a short overview of some major works; a more comprehensive review appears in [2].

In the large block length asymptotic, there are many influential and conclusive results. For the case of discrete-valued sources and $g(X_1^N) = X_1^N$, the lossless distributed source coding problem is solved by Slepian and Wolf [6]. In the lossy case, the problem is generally open except in specific situations [7], [8]. The case where $g(X_1^N) = X_1$ and the rate is unconstrained except for R_1 is the well-known source coding with side information problem [9]. For more general computations, the lossless [10]–[12] and lossy [13], [14] cases have both been explored.

There are also results for when the block length is constrained to be very small. We will delay discussion of DFSQ for Section II-C and instead focus on related works. The use of high-resolution for computation has been considered in detection and estimation problems [15]–[18]. In the scalar setting, the scenario where the computation is unknown but is drawn from a set of possibilities has been studied [19]. Finally, there are strong connections between DFSQ and multidimensional companding, a technique used in perceptual coding [20].

B. High-Resolution Scalar Quantizer Design

A scalar quantizer Q_K is a mapping from the real line to a set of K points $\mathcal{C} = \{c_k\}_{k=1}^K \subset \mathbb{R}$ called the codebook, where $Q_K(x) = c_k$ if $x \in P_k$ and the cells $\{P_k\}_{k=1}^K$ form a partition of $\mathbb{R}$. The quantizer is called *regular* if the partition cells are intervals containing the corresponding codewords. We then assume the codebook entries are indexed from smallest to largest and that $P_k = (p_{k-1}, p_k]$ for each k; this is essentially without loss of generality because the dispositions of the endpoints of the cells are immaterial to performance when the quantizer input is continuous. Regularity implies $p_0 < c_1 \leq p_1 < c_2 \leq \cdots < c_K \leq p_K$, with $p_0 = -\infty$ and $p_K = \infty$. Define the *granular* region as (c_1, c_K) and its complement $(-\infty, c_1] \cup [c_K, \infty)$ as the *overload* region.

Uniform (linear) quantization, where partition cells in the granular region have equal length, is most commonly used in practice, but other quantizer designs can improve reconstruction fidelity. Fig. 2 presents the compander model as a method for generating nonuniform quantizers from a uniform one. In this model, 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 levels in $(0, 1)$, 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 convenient to define a *point density function* as $\lambda(x) = c'(x)$. Because of 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 Fig. 2 can be uniquely specified using a point density function and codebook size. We denote such a quantizer as $Q_{K,\lambda}$. By virtue of this definition, the integral of the point density function over any quantizer cell is $1/K$:

$$\int_{p_k}^{p_{k+1}} \lambda(x)\, dx = \frac{1}{K}, \quad k = 1, 2, \ldots, K. \tag{1}$$

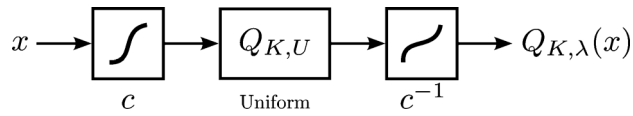


Fig. 2. The compander model for constructing nonuniform scalar quantizers. The compressor function c is defined such that $\lim_{x \to -\infty} c(x) = 0$ and $\lim_{x \to \infty} c(x) = 1$. The notation $Q_{K,U}$ is used to denote the canonical uniform quantizer with K codewords in $(0, 1)$. In this paper, only the partition boundaries are scaled using c; the codewords are defined through midpoint reconstruction (2).

In practice, scalar quantization is rarely, if ever, performed by an explicit companding operation. A slight modification that avoids repeated computation of c^{-1} derives partition boundaries from the compressor function c by applying 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. We assume that the non-extremal reconstruction values are set to the midpoints of the cells, i.e.,

$$c_k = \frac{p_{k-1} + p_k}{2}, \quad k = 2, 3, \ldots, K-1. \tag{2}$$

This is suboptimal in terms of MSE relative to centroid reconstruction, but it has the simplicity of depending only on λ and K, not on the source density. The extremal reconstruction values are fixed to be $c_1 = p_1$ and $c_K = p_{K-1}$. This again is suboptimal but does not depend on the source distribution. We will show later that this suboptimality does not affect asymptotic quantizer performance.

The utility of the compander model is that we can precisely analyze the distortion behavior as K becomes large and use this to optimize λ. Assuming the source is well-modeled as being drawn from a probabilistic distribution, we define the mean-squared error (MSE) distortion as

$$D_{\text{mse}}(K, \lambda) = \mathrm{E}\left[|X - Q_{K,\lambda}(X)|^2\right], \tag{3}$$

where the expectation is with respect to the source density f_X. Under the additional assumption that the tails of f_X decay sufficiently fast,

$$D_{\text{mse}}(K, \lambda) \simeq \frac{1}{12K^2} \mathrm{E}[\lambda^{-2}(X)], \tag{4}$$

where $\simeq$ indicates that the ratio of the two expressions approaches 1 as K increases [21], [22]. Hence, the MSE performance 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 this approximation 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 [23]–[25]:

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

where λ is a valid point density. Experimentally, the approximation is accurate even for moderate K [1], [26]. Since the dependence on K and λ is separated in the limit, calculus techniques can be used to optimize companders.

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 performance and communication rate R, defined to be the expected number of bits per sample. In the simplest case, the codewords are indexed by a simple binary expansion and the communication rate is $R = \log_2(K)$; this is called *fixed-rate* or *codebook-constrained* quantization. Hölder's inequality can be used to show that the optimal point density for fixed-rate is asymptotically

$$\lambda^*_{\mathrm{mse,fr}}(x) \propto f_X^{1/3}(x), \tag{5}$$

and the resulting distortion is asymptotically

$$D^*_{\mathrm{mse,fr}}(R) \simeq \frac{1}{12}\,\|f_X\|_{1/3}\,2^{-2R}, \tag{6}$$

with the notation $\|f\|_p = (\int_{-\infty}^{\infty} f^p(x)\,dx)^{1/p}$ [27]. In general, 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. If the decoding latency is allowed to be large, one can employ block entropy coding and the communication rate approaches $H(Q_{K,\lambda}(X))$:

$$R \simeq h(X) + \log_2 K + \mathrm{E}[\log_2 \lambda(X)]. \tag{7}$$

This particular scenario, called *entropy-constrained* quantization, can be analyzed using Jensen's inequality to show the optimal point density $\lambda^*_{\mathrm{mse,ec}}$ is constant on the support of the input distribution [27]. The optimal quantizer is asymptotically uniform and the resulting distortion is asymptotically

$$D^*_{\mathrm{mse,ec}}(R) \simeq \frac{1}{12}2^{-2(R-h(X))}. \tag{8}$$

Note that block entropy coding suggests that the sources are transmitted in blocks even though the quantization is scalar. As such, (8) is an asymptotic result and serves as a lower bound on practical entropy coders with finite block lengths that match the latency 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 [28]. The compander model used in this paper 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 [29], [30]. 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 under mild conditions it does not change the asymptotic behavior as K and R increase without bound.

C. Functional Scalar Quantizer Design

In a distributed network where the encoders employ scalar quantization and the decoder performs a reconstruction using $\hat{g}$ on the quantized data to approximate a desired computation g, optimizing the quantizers for g rather than source fidelity can lead to substantial gains. In [2], distortion performance and quantizer design are discussed for the distributed setting shown in Fig. 1, with g a scalar-valued function. For DFSQ, the cost of interest is functional MSE (fMSE):

$$D_{\mathrm{fmse}}(K_1^N, \lambda_1^N) = \mathrm{E}\left[\left|g(X_1^N) - \hat{g}(Q_{K_1^N,\lambda_1^N}(X_1^N))\right|^2\right], \tag{9}$$

where $\hat{g}$ is chosen to be the joint centroid (JC) reconstruction or minimum functional MSE (fMMSE) estimator

$$\hat{g}_{\mathrm{jc}}(Q_{K_1^N,\lambda_1^N}(x_1^N)) = \mathrm{E}\left[g(X_1^N)\,|\,Q_{K_1^N,\lambda_1^N}(X_1^N) = Q_{K_1^N,\lambda_1^N}(x_1^N)\right], \tag{10}$$

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

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

Note the complexity of computing $\hat{g}_{\mathrm{jc}}$—it requires integrating over an N-dimensional partition cell with knowledge of the joint source density $f_{X_1^N}$. Later in this paper, we avoid this complexity by choosing $\hat{g}$ to be simply the desired computation directly applied to the quantized observations.

Before understanding how a quantizer affects fMSE, it is convenient to define how a computation locally affects distortion.

Definition 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

$$\gamma_n(x) = \left(\mathrm{E}\left[|g_n(X_1^N)|^2\,|\,X_n = x\right]\right)^{1/2}, \tag{11}$$

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 [2] says

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

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 0.

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

$$\mathrm{E}\left[\left(\frac{\gamma_n(X_n)}{\lambda_n(X_n)}\right)^2\right]$$

to be defined and finite for all n.

Following the same recipes to optimize over λ_1^N, the relationship between distortion and communication rate is found. In

both cases, 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

$$\lambda^*_{n,\mathrm{fmse,fr}}(x) \propto \left(\gamma_n(x) f_{X_n}(x)\right)^{1/3}, \tag{13}$$

where f_{X_n} is the marginal distribution of X_n. In the entropy-constrained case, the optimizing point density is asymptotically

$$\lambda^*_{n,\mathrm{fmse,ec}}(x) \propto \gamma_n(x). \tag{14}$$

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

D. Main Contributions of Paper

The central goal of this paper is to develop a more practical method upon the theoretical foundations of [2]. In particular, we provide new insight on how a simplified decoder can be used in lieu of the optimal one in (10). Although the conditional expectations are offline computations, they may be extremely difficult and are computationally infeasible for large N and K. We consider the case when the decoder is restricted to applying the function g explicitly on the quantized measurements. 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. In brief, we derive new conditions on the tail of the source density and computation that allow the distortion to be stably computed. Interestingly, this extends the class of probability densities under which high-resolution analysis techniques have been successfully applied. The generalization to vector-valued g is a more straightforward extension that is included for completeness. We present several examples to illustrate the framework and the convergence to the asymptotics developed in this work.

III. 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 $Q_{K,\lambda}(X)$ rather than compute the joint centroid condition like in [2]. 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 1: Assume f_X, g, and λ satisfy Conditions UF1′–UF6′. Then the fMSE

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

satisfies the following limit:

$$\lim_{K\to\infty} K^2 D_{\mathrm{fmse}}(K,\lambda) = \frac{1}{12}\mathrm{E}\left[\left(\frac{\gamma(X)}{\lambda(X)}\right)^2\right]. \tag{15}$$

Proof: See Appendix A. ∎

Remarks

1) The fMSE in (15) is the same as in (12). 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 (10). Further analysis on this point is given in Section III-C
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 [23]–[25].
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 1 [25]. Both conditions are satisfied in many problems of interest.
4) When g is monotonic, the performance in (15) is as good as quantizing and communicating $g(X)$ [2, Lemma 5]. Otherwise, the use of a regular quantizer results in a distortion penalty, as illustrated in Example 1 of Section V.

5) 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.
6) Although we have assumed f_X, g and λ are "nice" in the sense that they are continuous and positive, the proof of Theorem 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 [2, Section III-F]. A similar argument can also be made for g having a finite number of discontinuities in its first and second derivatives.
7) For the high-resolution assumptions to hold, the point density should be positive where the source distribution is positive. However, a consequence of Theorem 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 [2, Section VII].

A. Asymptotically Optimal Quantizer Sequences

Since the fMSE of Theorem 1 matches (12), the optimizing quantizers are the same. Using the recipe of Section II-B, we can show the optimal point density for fixed-rate quantization is asymptotically

$$\lambda^*_{\mathrm{fmse,fr}}(x) = \frac{\left(\gamma^2(x) f_X(x)\right)^{1/3}}{\int_{-\infty}^{\infty} \left(\gamma^2(t) f_X(t)\right)^{1/3}\, dt} \tag{16}$$

over the entire support of X, resulting in distortion

$$D^*_{\mathrm{fmse,fr}}(R) \simeq \frac{1}{12} \|\gamma^2 f_X\|_{1/3}\, 2^{-2R}. \tag{17}$$

Meanwhile, optimization in the entropy-constrained case yields

$$\lambda^*_{\mathrm{fmse,ec}}(x) = \frac{\gamma(x)}{\int_{-\infty}^{\infty} \gamma(t)\, dt} \tag{18}$$

over the entire support of X, resulting in distortion

$$D^*_{\mathrm{fmse,ec}}(R) \simeq \frac{1}{12}\, 2^{2h(X)+2\mathrm{E}[\log \gamma(X)]} 2^{-2R}. \tag{19}$$

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 5 of Section V, design

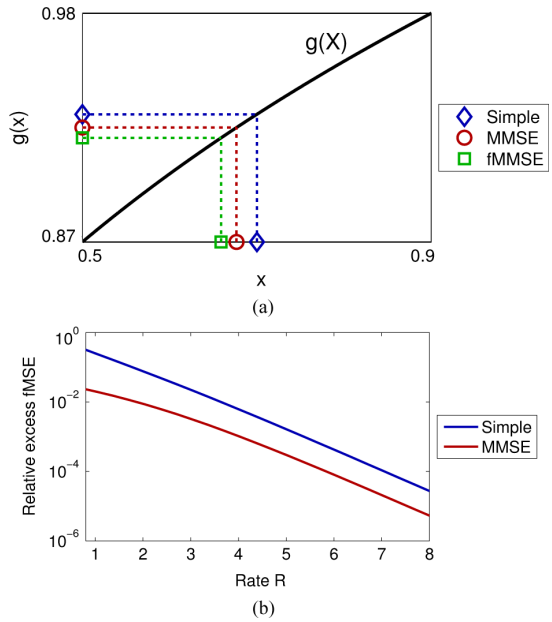


Fig. 3. (a) Codeword placement under simple, MMSE, and fMMSE decoders. The simple decoder performs midpoint reconstruction followed by the application of the computation g. The MMSE decoder applies g to the conditional expectation of X within the cell. Finally, the fMMSE decoder determines (10) for the 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.

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 (18) invalid; for example, a linear computation leads to constant γ. When the source has finite support, the integral in the denominator of (18) 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.

B. 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). 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 (10) (incorporating knowledge of the function as well). The codeword placements of the three decoders are visualized through an example in Fig. 3(a). The asymptotic match of the performance of the simple decoder to the optimal estimator (10) is a main contribution of this paper.

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 [30], and the proof of Theorem 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 [Fig. 3(b)]. Through experimental observation, we see the relative excess fMSE (defined as $(D_{\mathrm{dec}} - D_{\mathrm{opt}})/D_{\mathrm{opt}}$ appears exponential in rate, meaning

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

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.

IV. Multivariate Functional Quantization

We now describe the main result of the paper for the scenario shown in Fig. 1, where N random scalars $(X_1, \ldots, X_N)$ 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 $n \in \{1, \ldots, N\}$, 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 $i, j \in \{1, 2, \ldots, N\}$

MF4′. For any $n \in \{1, \ldots, N\}$,

$$f_{X_n}(x_n)\, |g_n(x_1^N)|^2 / \lambda_n^2(x_n) \tag{20}$$

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

$$f_{X_1^N}(x_1^N) \frac{|g_n(x_1^N)|\, |g_{i,j}(x_1^N)|}{\lambda_i(x_i)\lambda_j(x_j)\lambda_n(x_n)} \tag{21}$$

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

MF5′. for $i, j, m, n \in \{1, \ldots, N\}$,

$$f_{X_1^N}(x_1^N) \frac{|g_{i,j}(x_1^N)|\, |g_{m,n}(x_1^N)|}{\lambda_i(x_i)\lambda_j(x_j)\lambda_m(x_m)\lambda_n(x_n)} \tag{22}$$

is integrable over $\mathbb{R}^N$. For $i, j \in \{1, \ldots, N\}$,

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

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 $n \in \{1, \ldots, N\}$. This condition is a generalization of UF6′ for $m = 0$ applied to (20). Effectively, it bounds the tail contributions of an integral with the integrand being a modified version of (20). We also require similar conditions for (21) and (22), which are analogous to UF6′ for $m = 1$ and $m = 2$ respectively. We omit the exact form here for brevity.

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 1:

Theorem 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

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

satisfies the following limit:

$$\lim_{\kappa \to \infty} \kappa^2 D_{\mathrm{fmse}}(K_1^N, \lambda_1^N) = \sum_{n=1}^N \frac{1}{12 \alpha_n^2} \mathrm{E}\left[\left(\frac{\gamma_n(X_n)}{\lambda_n(X_n)}\right)^2\right]. \tag{23}$$

Proof: See Appendix B ∎

Remarks

1) Like in the univariate case, the simple decoder has performance that is asymptotically equivalent to the more complicated optimal decoder (10).
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 2 of Section V.

3) The strict requirements of $\mathrm{MF1}'$ and $\mathrm{MF3}'$ 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 (23), which would invalidate Theorem 2. However, based on experimental observations, such as in Example 5 of Section V, 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 [2] could potentially be useful in showing this rigorously.
4) Condition $\mathrm{MF5}'$ is known as the asymptotic whiteness property (AWP). For uniform quantization with midpoint reconstruction and nonuniform quantization with centroid reconstruction, it is shown in [31], [32] 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 [31] 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 [33].

A. 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

$$\lambda^*_{n,\mathrm{fmse,fr}}(x) = \frac{\left(\gamma_n^2(x) f_{X_n}(x)\right)^{1/3}}{\int_{-\infty}^{\infty} \left(\gamma_n^2(t) f_{X_n}(t)\right)^{1/3}\, dt} \tag{24}$$

over the support of X_n, with fMSE

$$D^*_{\mathrm{fmse,fr}}(R_1^N) \simeq \frac{1}{12} \sum_{n=1}^{N} \|\gamma_n^2 f_{X_n}\|_{1/3}\, 2^{-2R_n}. \tag{25}$$

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

$$\lambda^*_{n,\mathrm{fmse,ec}}(x) = \frac{\gamma_n(x)}{\int_{-\infty}^{\infty} \gamma_n(t)\, dt} \tag{26}$$

over the support of X_n, leading to a fMSE of

$$D^*_{\mathrm{fmse,ec}}(R_1^N) \simeq \frac{1}{12} \sum_{n=1}^{N} 2^{2h(X_n)+2\mathrm{E}[\log \gamma(X_n)]} 2^{-2R_n}. \tag{27}$$

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 [2, Lemma 4].

As in the univariate case, this optimization arrives with the caveat that conditions $\mathrm{MF4}'$–$\mathrm{MF7}'$ 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 III-B, the requirements can often be too strict.

B. Vector-Valued Functions

In Theorem 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:

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

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 2, we can find the limit of the weighted fMSE assuming each entry of the vector-valued function satisfies $\mathrm{MF1}'$–$\mathrm{MF7}'$.

Corollary 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:

$$\lim_{\kappa\to\infty} \kappa^2 D_{\mathrm{fmse}}(K_1^N, \lambda_1^N, \beta_1^M) = \sum_{n=1}^{N} \frac{1}{12\alpha_n^2} \mathrm{E}\left[\left(\frac{\gamma_n(X_n, \beta_1^M)}{\lambda_n(X_n)}\right)^2\right], \tag{28}$$

where the *combined functional sensitivity profile* is

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

The point densities given in (24) and (26) are again optimal under this new definition of γ_n.

V. 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.

A. Examples for Univariate Functional Quantization

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

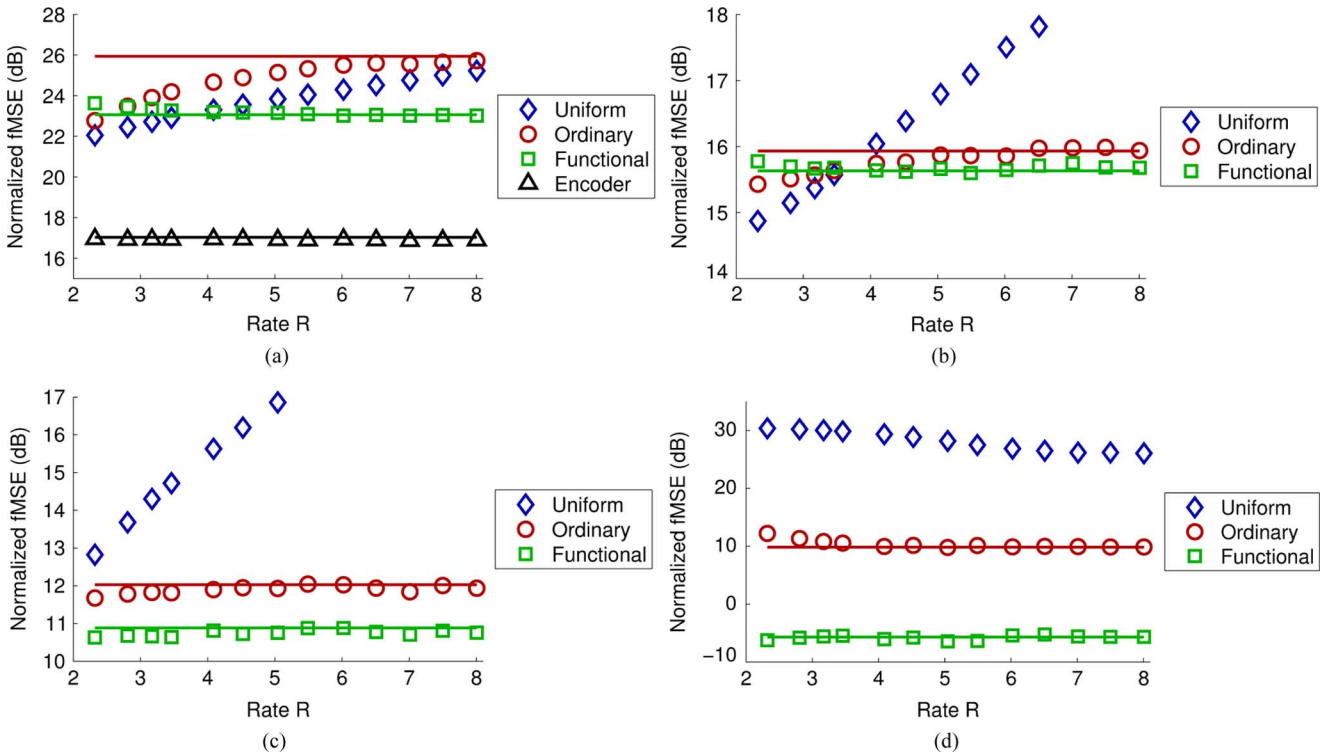


Fig. 4. Empirical and theoretical performance for the ordinary and functional quantizers for: (a) a scalar Gaussian source and $g(x) = x^2$; (b) jointly Gaussian sources with correlation coefficient 0.5 and $g(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$; (c) exponential sources with parameter $\lambda = 1$ and $g(x_1, x_2) = x_1/(1 + x_2)$; 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 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}.

Example 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 II-B, and optimal functional quantizers given in Section III-C, 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 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 Fig. 4(a). The distortion-minimizing uniform quantizer has a granular region that depends on R, which was explored in [34]. 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 III-A. In the fixed-rate case, the empirical performance approaches the distortion limit described by Theorem 1. The convergence is fast and the asymptotic results predict practical quantizer performance at rates as low as 4 bits/sample.

B. Examples for Multivariate Functional Quantization

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

Example 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 1. The distortion is also the same, except now scaled by N.

Example 3: We now consider a more interesting extension of Example 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

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

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

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

In Fig. 4(b), 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 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 (11), 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 Fig. 4(c), we experimentally verify that sequences of real quantizers approach the predicted distortion-rate trade-off.

Example 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 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:

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

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

In Fig. 4(d), we experimentally verify that the asymptotic predictions are precise. This serves as evidence that MF3′ may be loosened.

VI. Conclusion

In this paper, 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 paper 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.

Appendix A
Proof of Theorem 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

$$|R_1(x,c_k)| \le \frac{1}{2}\max_{\xi\in[c_k,p_k]}|g''(\xi)|\,(p_k-c_k)^2. \tag{29}$$

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 $Q_{K,\lambda}$:

$$\tilde{g}(x) = \sup_{t\in P_k}|g''(t)| \text{ if } x \in P_k,\ k = 1,\ldots,K. \tag{30}$$

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

$$|R_1(x,c_k)| \le \frac{1}{8}\tilde{g}(x)I^2(x). \tag{31}$$

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

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

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

$$\int_{p_{K-1}}^{\infty}|g(x)-g(p_{K-1})|^2 f_X(x)\,dx, \tag{32}$$

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

$$D_{\mathrm{fmse}}(K,\lambda) \simeq \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} |g(x) - g(c_k)|^2 f_X(x)\,dx, \quad (33)$$

where we recall $\simeq$ indicates that the ratio of the two expressions approaches 1 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 (33) using Taylor's theorem yields:

$$\begin{aligned}
&K^2 D_{\mathrm{fmse}}(K,\lambda)\\
&\simeq K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} |g'(c_k)(x-c_k) + R_1(x,c_k)|^2 f_X(x)\,dx\\
&\le \underbrace{K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} |g'(c_k)|^2\,|x-c_k|^2 f_X(x)\,dx}_{A} \quad (34)\\
&\quad + \underbrace{K^2 \sum_{k=1}^{K-2} 2\int_{p_k}^{p_{k+1}} |R_1(x,c_k)|\,|g'(c_k)|\,|x-c_K| f_X(x)\,dx}_{B} \quad (35)\\
&\quad + \underbrace{K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} R_1(x,c_k)^2 f_X(x)\,dx}_{C}. \quad (36)
\end{aligned}$$

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

$$\begin{aligned}
&\lim_{K\to\infty} K^2 \sum_{k=1}^{K-2} \int_{p_k}^{p_{k+1}} |g'(c_k)|^2\,|x-c_k|^2 f_X(x)\,dx\\
&\overset{(a)}{=} \lim_{K\to\infty} K^2 \int_{p_1}^{p_{K-1}} |g'(x)|^2\,|x - Q_{K,\lambda}(x)|^2 f_X(x)\,dx\\
&\overset{(b)}{=} \lim_{K\to\infty} K^2 \int_{\mathbb{R}} |g'(x)|^2\,|x - Q_{K,\lambda}(x)|^2 f_X(x)\,dx\\
&\overset{(c)}{=} \frac{1}{12}\int_{\mathbb{R}} \left(\frac{g'(x)}{\lambda(x)}\right)^2 f_X(x)\,dx, \quad (37)
\end{aligned}$$

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 [25], which is given in Theorem 3 in Appendix C. Conditions UF2′, UF4′ and UF6′ for $m = 0$ are used here. Noting that $\gamma(x) = |g'(x)|$ gives (15).

The higher-order error terms become negligible with increasing K using the bound reviewed in (29):

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

where (a) follows from bounding $R_1(x,c_k)$ using (31); (b) from $p_1 \to -\infty$ and $p_{K-1} \to \infty$; (c) from a similar extension of Theorem 3 (see Appendix D), using UF2′ and UF6′ for $m = 1$; and (d) from UF4′ for $m = 1$. Compared to (37), there is an extra $1/K$ factor arising from the second-order Taylor error, which drives term B to 0. 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.

Appendix B
Proof of Theorem 2

We parallel the proof of Theorem 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

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

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

$$\left|R_1(x_1^N, a_1^N)\right| \le \sum_{i=1}^{N}\sum_{j=1}^{N} |x_i - a_i|\,|x_j - a_j|\,|g_{i,j}(x_1^N)|, \quad (38)$$

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

$$\kappa^2 D_{\mathrm{fmse}}(K_1^N, \lambda_1^N) \le A + 2B + C,$$

where

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

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

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

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

$$\sum_{t\in\mathcal{T}_N} \sum_{n=1}^N \int_{x_1^N\in t} |g_n\left((c_s)_1^N\right)|^2\,|x_n - c_{t,n}|^2 f_{X_1^N}(x_1^N)\, dx_1^N. \tag{39}$$

We note that these distortion contributions are equivalent to those in the univariate case and can apply the derivations in Theorem 1. Using Conditions MF2′, MF3′ and MF7′, (39) approaches the integral expression

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

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 (11), 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 (39) in the granular region; therefore, they do not contribute to the asymptotic distortion. In Remark 3 of Section IV-A, 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 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 $Q_{K,\lambda}$:

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

where t is an N-dimensional cell in $\mathcal{T}_N$. We can then bound (38):

$$\left|R_1(x_1^N, a_1^N)\right| \leq \frac{1}{4}\sum_{i=1}^N \sum_{j=1}^N I_j(x_i) I_i(x_j) \tilde{g}_{i,j}(x_1^N). \tag{40}$$

We now consider B:

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

where (a) follows from bounding $R_1(x_1^N, c_t)$ using (30) and the fact that the limits of integration converge to $\mathbb{R}^N$; and (b) from a generalization of the proof by Linder [25], 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 0, 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.

Appendix C
Weighted Distortion of Companding Quantizers

In this section, we prove a modest extension to Linder's rigorous results [25] 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:

$$\int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2\, w(x)\, f_X(x)\, dx. \tag{41}$$

Linder's result for MSE relies 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 1.

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 2: 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

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

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

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

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 [25], due to the different placement of codewords in the extremal quantization cells.

Definition 4: We define as an approximation to s a function $\hat{s}_K : (0,1) \to (0,\infty)$:

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

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 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 [35, Section 4.4] and $s_K \to s$ as $K \to \infty$ to show

$$\lim_{K\to\infty} \int_0^{\varepsilon} s_K^2(y)\, w(c^{-1}(y))\, p(y)\, dy = \int_0^{\varepsilon} s^2(y)\, w(c^{-1}(y))\, p(y)\, dy. \tag{42}$$

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

$$\lim_{K\to\infty} \int_{1-\varepsilon}^{1} s_K^2(y)\, w(c^{-1}(y))\, p(y)\, dy = \int_{1-\varepsilon}^{1} s^2(y)\, w(c^{-1}(y))\, p(y)\, dy. \tag{43}$$

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

$$\lim_{K\to\infty} \int_{\varepsilon}^{1-\varepsilon} s_K^2(y)\, \gamma^2(c^{-1}(y))\, p(y)\, dy = \int_{\varepsilon}^{1-\varepsilon} s^2(y)\, \gamma^2(c^{-1}(y))\, p(y)\, dy. \tag{44}$$

Combining (42)–(44) proves the lemma. ∎

Next we relate quantization error and s_K:

Lemma 2: For large K and $x \in I_{k,K}$, $k = 1, \ldots, K$,

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

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: 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 2, we find $s_K(y)$ dominates $\hat{s}_K(y)$, i.e., $\hat{s}_K(y) \le s_K(y)$ for $y \in (0,1)$. Using Lemma 1, we see $s_K^2(c(x))h(x)$ is Lebesgue integrable. Combining the General Dominated Convergence Theorem [35, 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: 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

$$\lim_{K\to\infty} K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2 \, h(x)\, dx = \lim_{K\to\infty} K^2 \int_{\mathbb{R}} |x - Q_{K,\lambda}(x)|^2 \, \tilde{h}_K(x)\, dx. \tag{45}$$

To prove (45), we note

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

where the last inequality holds only for large K since $A_K(h)$ approaches 1 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 $|h(x) - \tilde{h}_K(x)| \le f(x)$, along with Lemma 1 to show (45).

To complete the proof of the theorem, we combine Lemma 3 and (45).

Appendix D
Generalizing Theorem 3

We also need a Linder-style proof to bound the higher-order distortion terms (35) and (36). Here, we provide only a brief sketch on how to extend Theorem 3. Consider the integral

$$K^2 \int_{\mathbb{R}} I^2(x)\, w(x)\, |x - Q_{K,\lambda}(x)|\, f_X(x)\, dx, \tag{46}$$

where $I(x) = \mu(I_{k,K})$ if $x \in I_{k,K}$. We can rewrite (46) as

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

where the first line uses the definition of $\hat{s}_K$ and the second uses Lemma 2. Ensuring that the right-hand side is integrable is sufficient to show that (46) becomes negligible as K becomes large. The success of convergence with K depends on a condition analogous to LC4.

Acknowledgment

The authors would like to thank the anonymous reviewers and the Associate Editor P. Ishwar for the many helpful comments that led to improved rigor of the results and clarity of the presentation.

References

[1] D. L. Neuhoff, "The other asymptotic theory of lossy source coding," in *Coding and Quantization, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, R. Calderbank, G. D. Forney, Jr., and N. Moayeri, Eds. Providence, RI, USA: Amer. Math. Soc., 1993, vol. 14, pp. 55–65.

[2] V. Misra, V. K Goyal, and L. R. Varshney, "Distributed scalar quantization for computing: High-resolution analysis and extensions," *IEEE Trans. Inf. Theory*, vol. 57, pp. 5298–5325, Aug. 2011.

[3] J. Z. Sun and V. K Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun.–Jul. 2009, pp. 6–10.

[4] J. Z. Sun and V. K Goyal, "Scalar quantization for relative error," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, USA, Mar. 2011, pp. 293–302.

[5] M. Pugh and B. D. Rao, "Distributed quantization of order statistics with applications to CSI feedback," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, USA, Mar. 2011, pp. 323–332.

[6] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 471–480, Jul. 1973.

[7] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-terminal source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1938–1961, May 2008.

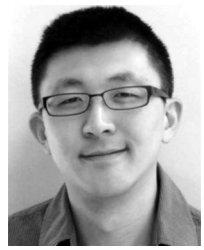
[8] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inf. Theory*, vol. 45, pp. 106–117, Jan. 1999.

[9] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.

[10] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, pp. 903–917, Mar. 2001.

[11] 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. Inf. Theory*, vol. IT-33, pp. 69–76, Jan. 1987.

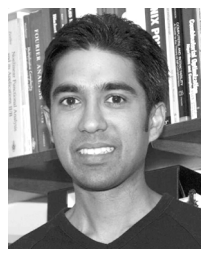
[12] V. Doshi, D. S. M. Médard, and S. Jaggi, "Distributed functional compression through graph coloring," in *Proc. IEEE Data Compress. Conf.*, Snowbird, UT, USA, Mar. 2007, pp. 93–102.
[13] H. Yamamoto, "Wyner–ziv theory for a general function of the correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 803–807, Sep. 1982.
[14] 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.
[15] H. V. Poor, "High-rate vector quantization for detection," *IEEE Trans. Inf. Theory*, vol. 34, pp. 960–972, Sep. 1988.
[16] G. R. Benitz and J. A. Bucklew, "Asymptotically optimal quantizers for detection of i.i.d. data," *IEEE Trans. Inf. Theory*, vol. 35, pp. 316–325, Mar. 1989.
[17] R. Gupta and A. O. Hero, "III high-rate vector quantization for detection," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1951–1969, Aug. 2003.
[18] S. Marano, V. Matta, and P. Willett, "Asymptotic design of quantizers for decentralized MMSE estimation," *IEEE Trans. Signal Process.*, vol. 55, pp. 5485–5496, Nov. 2007.
[19] J. A. Bucklew, "Multidimensional digitization of data followed by a mapping," *IEEE Trans. Inf. Theory*, vol. IT-30, pp. 107–110, Jan. 1984.
[20] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: Multidimensional companding," *IEEE Trans. Inf. Theory*, vol. 45, pp. 548–561, Mar. 1999.
[21] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, Jul. 1948.
[22] 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.
[23] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with r th power distortion measures," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 239–247, Mar. 1982.
[24] S. Cambanis and N. L. Gerr, "A simple class of asymptotically optimal quantizers," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 664–676, Sep. 1983.
[25] T. Linder, "On asymptotically optimal companding quantization," *Prob. Control Inf. Theory*, vol. 20, no. 6, pp. 475–484, 1991.
[26] V. K Goyal, "High-rate transform coding: How high is high, and does it matter?," in *Proc. IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, Jun. 2000, pp. 207–207.
[27] R. M. Gray and A. H. Gray, Jr, "Asymptotically optimal quantizers," *IEEE Trans. Inf. Theory*, vol. IT-23, pp. 143–144, Feb. 1977.
[28] 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. Inf. Theory*, vol. 49, pp. 3031–3037, Nov. 2003.
[29] H. Gish and J. P. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 676–683, Sep. 1968.
[30] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.
[31] H. Viswanathan and R. Zamir, "On the whiteness of high-resolution quantization errors," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2029–2038, Jul. 2001.
[32] 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.
[33] D. Marco and D. L. Neuhoff, "The validity of the additive noise model for uniform scalar quantization," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1739–1755, May 2005.
[34] D. Hui and D. L. Neuhoff, "Asymptotic analysis of optimal fixed-rate uniform scalar quantization," *IEEE Trans. Inf. Theory*, vol. 47, pp. 957–977, Mar. 2001.
[35] H. L. Royden and P. M. Fitzpatrick, *Real Analysis*, 4th. ed. Boston, MA, USA: Prentice-Hall, 2010.



John Z. Sun (S'08) received the B.S. degree (with honors) in electrical and computer engineering *(summa cum laude)* from Cornell University, Ithaca, NY, USA, in 2007. He received the M.S. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2009. Currently, he is pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, MIT. Mr. Sun was awarded the Student Best Paper Award at the IEEE Data Compression Conference in 2011 and was the recipient of the Claude E. Shannon Research Assistantship in 2011–2012. His research interests include signal processing, information theory, and statistical inference.

Vinith Misra received the S.B. and M.Eng. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2008, where his thesis was awarded the David Adler Memorial M.Eng. prize. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Stanford University.

He is a Stanford Graduate Fellow and a recipient of the National Defense Science and Engineering Graduate Fellowship. His research interests include information theory and signal processing, with applications to communication systems and statistical learning.



Vivek K Goyal (S'92–M'98–SM'03) received the B.S. degree in mathematics and the B.S.E. degree in electrical engineering from the University of Iowa, where he received the John Briggs Memorial Award for the top undergraduate across all colleges. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, CA, USA, where he received the Eliahu Jury Award for outstanding achievement in systems, communications, control, or signal processing. He was a Member of Technical Staff in the Mathematics of Communications Research Department of Bell Laboratories, Lucent Technologies, from 1998–2001 and a Senior Research Engineer at Digital Fountain, Inc., from 2001–2003. He has been with the Massachusetts Institute of Technology since 2004. His research interests include computational imaging, sampling, quantization, and source coding theory.

Dr. Goyal is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, Eta Kappa Nu and SIAM. He was awarded the 2002 IEEE Signal Processing Society Magazine Award and an NSF CAREER Award. As a research supervisor, he is the co-author of papers that have won Student Best Paper awards at the IEEE Data Compression Conference in 2006 and 2011 and the IEEE Sensor Array and Multichannel Signal Processing Workshop in 2012. He served on the IEEE Signal Processing Society's Image and Multiple Dimensional Signal Processing Technical Committee from 2003–2009. He currently serves on the Steering Committee of the IEEE Transactions on Multimedia, the Editorial Board of Foundations and Trends in Signal Processing, and the Scientific Advisory Board of the Banff International Research Station for Mathematical Innovation and Discovery. He is a Technical Program Committee Co-chair of IEEE ICIP 2016 and a permanent Conference Co-chair of the SPIE Wavelets and Sparsity Conference Series.