

Information Rates of Densely Sampled Gaussian Data

David L. Neuhoff
EECS Dept., University of Michigan
neuhoff@eecs.umich.edu

S. Sandeep Pradhan
EECS Dept., University of Michigan
pradhanv@eecs.umich

Abstract—With mean-squared error D as a goal, it is well known that one may approach the rate-distortion function $R(D)$ of a nonbandlimited, continuous-time Gaussian source by sampling at a sufficiently high rate, applying the Karhunen-Loeve transform to sufficiently long blocks, and then independently coding the transform coefficients of each type. Motivated by the question of the efficiency of dense sensor networks for sampling, encoding and reconstructing spatial random fields, this paper studies the following three cases. In the first, we consider a centralized encoding setup with a sample-transform-quantize scheme where the quantization is assumed to be optimal. In the second, we consider a distributed setup, where a spatio-temporal source is sampled and distributively encoded to be reconstructed at a receiver. We show that with ideal distributed lossy coding, dense sensor networks can efficiently sense and convey a field, in contrast to the negative result obtained by Marco et al. for encoders based on time- and space-invariant scalar quantization and ideal Slepian-Wolf distributed lossless coding. In the third, we consider a centralized setup, with a sample-and-transform coding scheme in which ideal coding of coefficients is replaced by coding with some specified family of quantizers. It is shown that when the sampling rate is large, the operational rate-distortion function of such a scheme comes within a finite constant of that of the first case.

I. INTRODUCTION

Suppose a spatially distributed random field is sampled by a network of densely spaced sensors, and suppose these samples must be encoded with some loss for transmission to a common decoder where the field is approximately reconstructed. How should the sensors compress their observations under various constraints on (i) inter-sensor collaboration and (ii) the structure of encoding operation? What happens to the number of encoded bits per unit area (scaling laws), i.e the *encoding rate*, as sensor density increases, given a fixed limit on the decoded reconstruction quality? On the one hand, as density increases, more sensor values must be encoded. On the other hand, adjacent values become more correlated, which might be exploited, perhaps, to prevent the encoding rate from growing. The answers to the above questions have broad implications to, in addition to sensor network applications, the general theory of oversampled quantization.

For concreteness, consider a stationary and ergodic spatio-temporal jointly Gaussian random field $X^{(t)}(s)$, where the spatial variable s is continuous, and, for simplicity, one-dimensional, and the temporal variable t is discrete. Let us suppose that we wish to reconstruct this field with respect to a mean squared error (MSE) distortion criterion. We assume

that the process is correlated in space, but independent and identically distributed (IID) in time. We also assume that the random field is distributed over unbounded space and time. One architecture for encoding (centralized or distributed) this random field that has received attention has the following three components: sampling, linear transformation and quantization. This includes any distributed encoding setup by assuming that the identity transform is used.

The original random field is first sampled spatially by the sensors to obtain a discrete-space, discrete-time random process. At every time t , a transform is applied spatially, and the transformed samples are quantized in temporal blocks and transmitted to a remote joint receiver. These operations are inverted to obtain a joint reconstruction of the entire field at the receiver. There are two performance measures of interest associated with such a system: rate and MSE. The rate measures the number of bits per unit time and unit space required in the quantization operation, and distortion measures the quality of the reconstruction. Assuming a system of this form is optimized, the trade-off between rate and distortion is known as the operational rate-distortion function of this type of system (we will define these quantities formally in the next section). Observe that this rate-distortion function depends on the sampling rate.

For such a system, one can consider four scenarios based on various alternative constraints on the transform and the quantizer: (I) no spatial transform (distributed coding) with only temporal scalar quantization, (II) no spatial transform (distributed coding) with temporal vector quantization, (III) spatial transform (centralized coding) with temporal scalar quantization, and (IV) spatial transform (centralized coding) with temporal vector quantization. We assume that the vector quantization is done with arbitrarily high dimension. Note that one can consider scalar and vector quantization with or without entropy coding in these scenarios.

The rate-distortion function of systems when constraints are put on both the transform and the quantizer (Case I) was first studied by Marco and Neuhoff in [2]. They showed that under some minimal conditions on the random field, as the sampling rate grows unbounded, the rate-distortion function becomes unbounded for every non-zero MSE distortion. Since the output vector of a scalar quantizer applied componentwise on the sampled continuous-space random process becomes highly correlated as the sampling rate increases, the entropy-

rate of the output of such a scalar quantizer tends to zero. However, the product of entropy-rate and sampling rate, which specifies the rate in bits per unit time per meter, was shown to grow without bound as sampling rate increases [3], [2].

The operational rate-distortion function for a given sampling rate of systems where no constraints are imposed on the transform and the quantizer (Case IV) was first stated (without proof) by Kolmogorov in 1959 [1]. It was stated, under certain assumptions on the underlying random process, that as sampling rate tends to a large number, the asymptotic operational rate-distortion function remains finite for every non-zero MSE distortion, and a parametric expression in terms of the power spectral density was given. Kashyap et al. [4] showed that the asymptotic operational rate-distortion function remains finite in Case II. This problem has also been addressed in different settings in [5], [6].

In [9], [10], we considered Case II and III and obtained a characterization of the corresponding rate-distortion functions. The precise conditions on the underlying random process under which the said characterizations are valid, however, were not given. In the present paper we provide a unified approach to characterizing the rate-distortion functions for Cases II, III and IV by giving precise conditions under which such characterizations are valid. Specifically, we will work under the assumption that the autocorrelation function of the random process is bounded, square integrable ($L^2(\mathbb{R}) \cap L^\infty(\mathbb{R})$) and satisfies certain mild tail conditions. We first obtain the rate-distortion function for Case IV in the $L^2(\mathbb{R})$ setting. This is done with a sample-transform-quantize architecture. The rate-distortion functions for Case IV have been obtained earlier under stricter conditions on the underlying random processes and using continuous-space Karhunen-Loeve transforms. We then treat Case II by providing an upper bound on the asymptotic operational rate-distortion function by exploiting the fact that the random process is distributed over unbounded space and time. We then address Case III and show that as the sampling rate grows unbounded, the asymptotic operational rate-distortion function remains finite for every non-zero distortion. This implies that the catastrophic performance of Case IV is due to lack of either a transform or a vector quantizer. In other words, one may conclude that it is the scalar quantization, rather than distributed coding per se, that caused the negative result of [3], [7].

II. SOURCE MODEL

To keep things simple, we consider a scenario like that of [3], [7], in which a one-dimensional, continuous-space, discrete-time spatially stationary, zero-mean Gaussian random process $X^{(t)}(s)$ is encoded. The superscript denotes the time variable (measured in time units), an integer, and the argument in brackets denotes the space variable (measured in meters), a real number. The process is independent and identically distributed in time, in the sense that $(X^{(1)}, X^{(2)}, \dots)$ is a sequence of independent and identically distributed sample functions. We assume that the spatial autocorrelation function

$\rho_X(s)$ and power spectral density $S_X(\Omega)$ of the process $X^{(t)}(s)$ are known.

Since this process is going to be sampled spatially, and then encoded, we assume that the process satisfies the following three conditions.

- (1) It is assumed that $\rho_X(\tau) \in L^2(\mathbb{R})$, and bounded, where $L^2(\mathbb{R})$ denotes the space of functions defined on the real line \mathbb{R} with finite power and Lebesgue measure. This implies that $S_X(\Omega) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.
- (2) There exist constants $p > 0$, $c_1 > 0$ and $c_2 > 0$ such that

$$\limsup_{|\tau| \rightarrow \infty} \rho_X(\tau) \left(\sqrt{|\tau|} \log^p |\tau| \right) = c_1, \quad (1)$$

- (3) and

$$\limsup_{|\Omega| \rightarrow \infty} S_X(\Omega) (|\Omega| \log^p |\Omega|) = c_2. \quad (2)$$

The first tail condition is used to show that that autocorrelation function of the sampled discrete-space process belongs to $l^2(\mathbb{Z})$. The second tail condition is used to show a well-known fact that a spatially scaled version of the power spectral density of the sampled process approaches the power spectral density of the original process as sampling interval approaches 0. Let $X^{(t)}(0 : \alpha)$ denote the collection of random variables $X^{(t)}(s)$ for $s \in [0, \alpha]$. Let $\Xi(\alpha)$ for $\alpha > 0$ denote the set of all real-valued functions defined over the spatial interval $[0, \alpha]$. All integrals are with respect to Lebesgue measure. For a set $A \subset \mathbb{R}$, $|A|$ denotes its Lebesgue measure.

III. CENTRALIZED ENCODING WITH VECTOR QUANTIZATION: CASE IV

For some sampling interval τ , at each time t , the process $X^{(t)}(s)$ is spatially sampled at sites $\{k\tau, k \in \mathbb{Z}^+\}$. Let the k th sample at time t be denoted by \bar{X}_{kt} . A block of ML samples of M sites are encoded into bits and are transmitted to a central decoder, which outputs continuous-space reproductions $(\hat{X}^{(1)}, \dots, \hat{X}^{(L)})$ for the $M\tau$ -length segments of the L continuous-space sample functions.

Definition 1: A centralized coding system with parameters (M, L, τ, Θ) is¹ composed of the following mappings: $e : \mathbb{R}^{ML} \rightarrow \{1, 2, \dots, \Theta\}$, for $i = 1, 2, \dots, L$, $f_i : \{1, 2, \dots, \Theta\} \rightarrow \Xi(M\tau)$ where $\hat{X}^{(i)} = f_i[e(\bar{X}_{11}, \bar{X}_{12}, \dots, \bar{X}_{1L}, \bar{X}_{21}, \bar{X}_{22}, \dots, \bar{X}_{2L}, \dots, \bar{X}_{M1}, \bar{X}_{M2}, \dots, \bar{X}_{ML})]$.

Let the rate of the system be given by

$$R = \frac{1}{M\tau L} \log \Theta \text{ bits/meter/time unit,}$$

and the distortion of the system be given by the mean squared error (MSE):

$$\Delta = \frac{1}{ML\tau} \sum_{i=1}^L \int_0^{M\tau} E \left(X^{(i)}(s) - \hat{X}^{(i)}(s) \right)^2 ds \quad (3)$$

The encoder maps ML samples of the process $X^{(t)}(s)$ into bits. The decoder maps the bits into L waveforms, where each is defined over the space interval $[0, M\tau]$.

¹Kolmogorov considered the case where $L = 1$.

Definition 2: A pair (ρ, δ) of non-negative real numbers is said to be τ -achievable for (centralized coding) Case IV if for all $\epsilon > 0$, there exists for all sufficiently large M and L a centralized coding system with parameters (M, L, τ, Θ) such that the rate and distortion of the system satisfy the following constraints: $R \leq \rho + \epsilon$, $\Delta \leq \delta + \epsilon$.

Let $\mathcal{R}_\tau^{CV}(\mathcal{D}) = \inf \rho$ such that (ρ, \mathcal{D}) is τ -achievable. Let $\mathcal{R}^{CV}(\mathcal{D}) = \liminf_{\tau \rightarrow 0} \mathcal{R}_\tau^{CV}(\mathcal{D})$ denote the asymptotic operational rate-distortion function² when sampling rate grows unbounded.

Theorem 1: For a spatio-temporal random process X , with autocorrelation function and power spectral satisfying Conditions 1, 2 and 3, an information theoretic characterization of $\mathcal{R}^{CV}(\mathcal{D})$ is given in the parametric form by

$$\mathcal{R}(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \max \left\{ \frac{1}{2} \log_2 \frac{S_X(\Omega)}{\theta}, 0 \right\} d\Omega \quad (4)$$

$$\mathcal{D}(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \min \{ S_X(\Omega), \theta \} d\Omega, \quad (5)$$

where θ is a free parameter with $0 \leq \theta \leq \text{ess sup}_\Omega S_X(\Omega)$.

Proof Outline: This is known in the literature as the inverse water-pouring formula. This is achieved by the following approach. An (M, L, τ, Θ) transmission system is built by first applying a spatial $M \times M$ Karhunen-Loeve (KL) transform on the sampled data at every time t and obtaining transformed coefficients which are statistically independent within each block of size M . Each type of coefficient is then block-coded across time using vector quantizers of dimension L and an appropriate rate. Each type corresponds roughly to a frequency component. The bits obtained are transmitted to the receiver. At the decoder, the quantized coefficients are recovered from the bits, and the inverse KL transform is applied on the decoded coefficients. Finally the decoder applies a simple sample-and-hold spatial interpolation. The key feature of this approach is that the distortion allocated to each frequency component is held at a constant level, unless the variance of the source corresponding to a frequency component is below this level, given by the parameter θ (see (5)). In the latter case, the corresponding frequency component is not coded. We use an eigenvalue distribution theorem (Theorem 5.4 in [13] for $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$) to get the expressions in terms of $S_X(\cdot)$.

This was first stated without proof in [1]. A proof of this for the case of autoregressive processes was given by [14]. This is proved in [15] under the condition that $S_X(\cdot) \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. This was done using the eigenvalue distribution theorem of Kac-Murdock-Szegö [12].

IV. DISTRIBUTED VECTOR QUANTIZATION: CASE II

A. Problem Statement

The source is sampled spatially by sensors, and they are divided into spatial groups of size M . A block of L samples (in time) at each site is encoded, independently of the samples at other sites, for transmission to a common decoder over a noiseless rate-constrained channel. The decoder, after receiving bits

from a group of M adjacent sites, outputs continuous-space reproductions $(\hat{X}^{(1)}, \dots, \hat{X}^{(L)})$ for the length $M\tau$ segments of the L continuous-space sample functions. Subject to a constraint on MSE, we wish to find an upper bound to the rate (in bits per unit distance) of an ideal distributed lossy source code that operates by encoding spatial samples $X^{(t)}(1/N), \dots, X^{(t)}(M/N)$ (taken at intervals of $\tau = 1/N$) with a distributed code of the kind described in the previous section and then reconstructing an approximate continuous-space sample function.

Definition 3: A distributed coding system with parameters $(M, L, \tau, \Theta_1, \dots, \Theta_M)$ is composed of the following mappings: for $j = 1, 2, \dots, M$

$$e_j : \mathbb{R}^L \rightarrow \{1, 2, \dots, \Theta_j\}, \quad (6)$$

for $i = 1, 2, \dots, L$,

$$f_i : \{1, 2, \dots, \Theta_1\} \times \dots \times \{1, 2, \dots, \Theta_M\} \rightarrow \Xi(M\tau). \quad (7)$$

Let the rate of the system be given by

$$R = \frac{1}{ML\tau} \log [\Theta_1 \dots \Theta_M] \text{ bits/meter/time unit} \quad (8)$$

and the distortion of the system be given by

$$\Delta = \frac{1}{ML\tau} \sum_{i=1}^L \int_0^{M\tau} E \left(X^i(s) - \hat{X}^i(s) \right)^2 ds \quad (9)$$

where $\hat{X}^i = f_i(e_j(X^{(1)}(j\tau), X^{(2)}(j\tau), \dots, X^{(L)}(j\tau)), j = 1, 2, \dots, M)$.

Definition 4: A pair (ρ, δ) of non-negative real numbers is said to be τ -achievable for (distributed coding) Case II if for all $\epsilon > 0$, there exists for all sufficiently large M and L a distributed coding system with parameters $(M, L, \tau, \Theta_1, \dots, \Theta_M)$ such that the rate and distortion of the system satisfy the following conditions $R \leq \rho + \epsilon$, $D \leq \delta + \epsilon$.

Let $\mathcal{R}_\tau^{DV}(\mathcal{D}) = \inf \rho$ such that (ρ, \mathcal{D}) is τ -achievable for Case II. Let $\mathcal{R}^{DV}(\mathcal{D}) = \liminf_{\tau \rightarrow 0} \mathcal{R}_\tau^{DV}(\mathcal{D})$ denote the asymptotic operational rate-distortion function when sampling rate grows unbounded.

B. Main Results

Let the i th sensor transmit at rate R_i bits per temporal sample³. Then, (see (8)) the rate of the system in bits per meter per time unit is given by $\frac{1}{M} \sum_{i=1}^M R_i$ multiplied by the sampling rate N (spatial samples or sensors per meter). The decoder first reconstructs the L samples observed at each sensor. Let D_i denote the average distortion of L -length block at the i th sensor for $i = 1, 2, \dots, M$. For doing this we employ a sequence (indexed by L) of Berger-Tung source codes that achieve the tuple $(R_1, R_2, \dots, R_M, D_1, \dots, D_M)$. Then we let M becomes large, and then N becomes large.

Theorem 2: An upper bound on the asymptotic optimal rate-distortion function $\mathcal{R}^{DV}(\mathcal{D})$ in the setting of distributed

²C stands for centralized and V stands for vector quantization.

³The i th sensor encodes a block of L temporal samples with LR_i bits.

vector quantization is given by the parametric formula: for any $\theta \in [0, \infty)$

$$\mathcal{R}^{BT}(\theta) \triangleq \frac{1}{4\pi} \int_{-\infty}^{\infty} \log \left(\frac{S(\Omega)}{\theta} + 1 \right) d\Omega \quad (10)$$

$$\mathcal{D}^{BT}(\theta) \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{S(\Omega)}{\frac{S(\Omega)}{\theta} + 1} \right) d\Omega, \quad (11)$$

where the autocorrelation function and the power spectral density of the underlying spatio-temporal random process satisfy Conditions 1, 2 and 3.

Proof: The proof is skipped due to lack of space.

V. SCALAR QUANTIZATION WITH TRANSFORMS

A. Problem Statement

We will use the source model considered in the previous sections. The transformed coefficients are scalar quantized and sent to a central decoder. The scalar quantized coefficients of a given type may be coded losslessly across time. For example if one uses entropy-constrained scalar quantization, then to achieve a specified rate of the quantizer, several quantized coefficients coming from a scalar quantizer at different time instants have to be coded jointly. In contrast, if one uses a fixed-rate scalar quantizer, then coding of scalar quantized coefficients across time may not be necessary.

The encoder has two components. The first component involves an $M \times M$ orthogonal linear transform $\mathcal{T}^{N,M}$ that maps $\mathcal{R}^M \rightarrow \mathcal{R}^M$. The encoder applies this transform to the set of M samples given by $(X^{(t)}(1/N), \dots, X^{(t)}(M/N))$ to obtain M random variables $(Z^{(t)}(1), \dots, Z^{(t)}(M))$. The second component involves scalar quantization with a structure described in the sequel. The decoder first reconstructs an approximation (inverse of quantization) to the M -length vector $(Z^{(t)}(1), \dots, Z^{(t)}(M))$, applies the inverse transform, and then uses sample-and-hold to get an approximation to the continuous-time sample function. Since for every t , the encoder and decoder perform an identical set of operations, in the following we drop the superscript that captures the dependence of random variables on t .

Definition 5: A collection \mathcal{Q} of scalar quantizers is called a *family* if for every $0 < D < 1$, there exists a scalar quantizer Q in the collection such that the mean squared error of Q when operating on a unit variance discrete-time IID Gaussian process is at most D .

The rate of a quantizer Q in \mathcal{Q} is given by the average number of bits per sample produced by the quantizer when operating on a unit variance IID Gaussian source.

Definition 6: With a family of quantizers \mathcal{Q} , one can associate a function $R_{\mathcal{Q}}^*(D)$, called the *operational rate-distortion function*, which specifies the minimum rate of all the quantizers in \mathcal{Q} with distortion D or less when operating on a unit variance IID Gaussian source.

Definition 7: A family \mathcal{Q} of scalar quantizers is said to be *good* if it satisfies the following conditions. The rate-distortion function $R_{\mathcal{Q}}^*(D)$ of the family is finite except at $D = 0$ and is

zero for $D \geq 1$. Let $\tilde{R}_{\mathcal{Q}}(D)$ denote the lower convex envelope of $R_{\mathcal{Q}}^*(D)$.

Definition 8: Given a random process $X(s)$ and a good family of scalar quantizers, we can now define the operational rate-distortion function $\mathcal{R}_{\mathcal{Q}}(M, N, \mathcal{D})$ of transform coding with family \mathcal{Q} as the minimum rate in bits/meter required to attain distortion \mathcal{D} or less using a decoder that applies the inverse transform on the quantized coefficients and then a subsequent sample-and-hold operation, where the minimization is over $T^{N,M}$ and target distortions $\{D_i^{N,M}\}_{i=1}^M$. Further, we define the asymptotic operational rate-distortion

$$\mathcal{R}_{\mathcal{Q}}(\mathcal{D}) = \liminf_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \mathcal{R}_{\mathcal{Q}}(M, N, \mathcal{D}).$$

B. Main Result

We pursue an upper bound to $\mathcal{R}_{\mathcal{Q}}(D)$ by letting $\mathcal{T}^{N,M}$ be the Karhunen-Loeve transform corresponding to the random vector $(X(1/N), \dots, X(M/N))$. This results in the collection of independent random variables $\{Z(i)\}_{i=1}^M$, where the variance of $Z(i)$ is equal to the i th eigenvalue $\lambda_i^{N,M}$ of the covariance matrix of the random vector $(X(1/N), \dots, X(M/N))$. This results in the following average rate (in bits per sample) and average distortion pair:

$$R^{N,M} = \frac{1}{M} \sum_{i=1}^M R_{\mathcal{Q}}^* \left(\frac{D_i^{N,M}}{\lambda_i^{N,M}} \right), \quad D^{N,M} = \frac{1}{M} \sum_{i=1}^M D_i^{N,M}.$$

By optimizing this pair over the allocation of distortions $D_i^{N,M}$ to different components Y_i , with a slight abuse of notation, we can get a compact parametric solution to the following function:

$$R^{N,M}(D) \triangleq \min_{\{D_i^{N,M}: \frac{1}{M} \sum_{i=1}^M D_i^{N,M} = D\}} \left[\frac{1}{M} \sum_{i=1}^M R_{\mathcal{Q}}^* \left(\frac{D_i^{N,M}}{\lambda_i^{N,M}} \right) \right].$$

Theorem 3: This rate-distortion function $R^{N,M}(D)$ is characterized as follows:

$$R^{N,M}(\beta, \theta) = \frac{1}{M} \sum_{i=1}^M \tilde{R}_{\mathcal{Q}}(\psi_{\beta}(\theta \lambda_i)),$$

$$D^{N,M}(\beta, \theta) = \frac{1}{M} \sum_{i=1}^M \lambda_i \psi_{\beta}(\theta \lambda_i),$$

for $\beta \in [0, 1]$ and $\theta \in (-\infty, 0]$, where

$$\psi(\theta) = \{d \in (0, 1] : \tilde{R}_{\mathcal{Q}}^-(d) \leq \theta \leq \tilde{R}_{\mathcal{Q}}^+(d) \text{ and } d < 1,$$

$$\text{or } \tilde{R}_{\mathcal{Q}}^-(d) \leq \theta \text{ and } d = 1\},$$

$\tilde{R}_{\mathcal{Q}}^-(d)$ and $\tilde{R}_{\mathcal{Q}}^+(d)$ are left-sided and right-sided derivative, respectively, of $\tilde{R}_{\mathcal{Q}}(d)$, $\psi_l(\theta)$ and $\psi_u(\theta)$ are the lower and upper end points, respectively, of $\psi(\theta)$, and $\psi_{\beta}(\theta) = (1 - \beta)\psi_l(\theta) + \beta\psi_u(\theta)$.

Now using the eigenvalue distribution theorem in $L^2(\mathbb{T})$ [13, Theorem 5.1], as $M \rightarrow \infty$ (following sub-additivity of the rate-distortion function), we get the following achievable rate-distortion pair:

$$R_{\mathcal{Q}}^{N,\infty}(\beta, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{R}_{\mathcal{Q}}(\psi_{\beta}(\theta\Phi_N(w))) dw \quad (12)$$

$$D_{\mathcal{Q}}^{N,\infty}(\beta, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_N(w)\psi_{\beta}(\theta\Phi_N(w)) dw, \quad (13)$$

for $\beta \in [0, 1]$ and $\theta \in (-\infty, 0]$, where, as before, $\Phi_N(w)$ denotes the discrete-space power spectral density of the process that is obtained by sampling the original process with sampling interval $1/N$. We make additional assumptions on the family \mathcal{Q} in terms of conditions on the lower-convex envelope $\tilde{R}_{\mathcal{Q}}(\cdot)$ of the rate-distortion function $R_{\mathcal{Q}}^*(\cdot)$.

We restrict our attention to good families of quantizers \mathcal{Q} such that $\tilde{R}_{\mathcal{Q}}(\cdot)$ satisfies the following two conditions

(4)

$$\tilde{R}_{\mathcal{Q}}(d) = \begin{cases} h(d) & 0 \leq d \leq d_0 \\ \frac{h(d_0)}{(1-d_0)}(1-d) & d_0 < d \leq 1 \end{cases}, \quad (14)$$

where $h(d)$ is defined on $[0, d_0]$, and convex and twice differentiable over the interval $(0, d_0]$ ⁴, and $h(0) = \infty$.

(5) There exists a constant $c_3 < 0$ such that for all $d \in (0, d_0]$,

$$\frac{h'(d)}{h''(d)} \geq c_3. \quad (15)$$

It turns out that we require an additional condition on the power spectral density $S_X(\cdot)$, which is given in the following

(6) If $S_X(\cdot)$ has infinite support, then $S_X(\cdot)$ does not have a flat spot, i.e., for every constant $\alpha > 0$,

$$|\{\Omega : S_X(\Omega) = \alpha\}| = 0. \quad (16)$$

Clearly, $h'(d)$ exists for all $d \in (0, d_0)$. Let $g(\cdot)$ denote the inverse of this function. Let $\theta_0 = \lim_{d \uparrow d_0} h'(d)$. Note that since $\tilde{R}_{\mathcal{Q}}(\cdot)$ is convex (by definition), we have that $\theta_0 \leq -\frac{h(d_0)}{(1-d_0)}$. Moreover, since $\tilde{R}_{\mathcal{Q}}(\cdot)$ always lies above the Shannon rate-distortion function, the slope of $\tilde{R}_{\mathcal{Q}}(\cdot)$ at $d = 1$ must be no greater than the corresponding slope of the latter, i.e.,

$$\theta_0 \leq -\frac{h(d_0)}{(1-d_0)} \leq -\frac{\log e}{2} < 0. \quad (17)$$

Under the above conditions, the inverse $\psi_{\beta}(\cdot)$ of the first derivative of $\tilde{R}_{\mathcal{Q}}(\cdot)$ can be written as for all $\beta \in [0, 1]$, and $\gamma \in (-\infty, 0]$,

$$\psi_{\beta}(\theta) = \begin{cases} g(\theta) & \theta < \theta_0 \\ d_0 & \theta_0 \leq \theta < -\frac{h(d_0)}{(1-d_0)} \\ (1-\beta)d_0 + \beta & \theta = -\frac{h(d_0)}{(1-d_0)} \\ 1 & -\frac{h(d_0)}{(1-d_0)} < \theta \leq 0. \end{cases}$$

Note that $\psi_{\beta}(\theta)$ is continuous at $\theta = \theta_0$. The following is the main result of this section.

Theorem 4: For a spatio-temporal random process X satisfying Condition 1, 2, 3 and 6, and a good family of quantizers \mathcal{Q} , with $\tilde{R}_{\mathcal{Q}}$ satisfying Condition 4 and 5, an upper bound on

⁴ $h'(d_0)$ is defined as $\lim_{d \uparrow d_0} h'(d)$, and similarly, $h''(d_0)$ is defined.

the asymptotic operational rate-distortion function $\mathcal{R}_{\mathcal{Q}}(\mathcal{D})$ is given by the following: for $\gamma \in (-\infty, 0]$,

$$\mathcal{R}_{\mathcal{Q}}^{KT}(\gamma) \triangleq \frac{1}{2\pi} \int_{\{\Omega: S_X(\Omega) > \frac{\theta_0}{\gamma}\}} h(g(\gamma S_X(\Omega))) d\Omega + \frac{h(d_0)}{2\pi} \left| \left\{ \Omega : \frac{\theta_0}{\gamma} > S_X(\Omega) > -\frac{h(d_0)}{\gamma(1-d_0)} \right\} \right| \text{ bits/meter} \quad (18)$$

$$\mathcal{D}_{\mathcal{Q}}^{KT}(\gamma) \triangleq \frac{1}{2\pi} \int_{\{\Omega: S_X(\Omega) > \frac{\theta_0}{\gamma}\}} S_X(\Omega) g(\gamma S_X(\Omega)) d\Omega + \frac{1}{2\pi} \int_{\{\Omega: \frac{\theta_0}{\gamma} > S_X(\Omega) > -\frac{h(d_0)}{\gamma(1-d_0)}\}} S_X(\Omega) d_0 d\Omega + \frac{1}{2\pi} \int_{\{\Omega: -\frac{h(d_0)}{\gamma(1-d_0)} > S_X(\Omega)\}} S_X(\Omega) d\Omega. \quad (19)$$

Proof: The proof of skipped due to lack of space.

REFERENCES

- [1] N. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IEEE Transactions on Information Theory*, vol. 2, pp. 102–108, 1956.
- [2] D. Marco and D. L. Neuhoff, "Entropy of quantized data at high sampling rates," *Proc. IEEE International Symposium on Information Theory (ISIT)*, September Adelaide, Australia, 2005.
- [3] D. Marco, E. J. Duarte-Melo, M. Liu, and D. L. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data," *Proc. Information processing in sensor networks (IPSN)*, 2003.
- [4] A. Kashyap, L. A. Lastras-Montano, C. Xia, and Z. Liu, "Distributed source coding in dense sensor networks," *Proceedings of the Data Compression Conference (DCC)*, April, Snowbird, UT, 2005.
- [5] A. Kumar, P. Ishwar, and K. Ramchandran, "High-resolution distributed sampling of bandlimited fields with low-precision sensors," *IEEE Trans. on Inform. Theory*, vol. 57, no. 1, pp. 269–275, January 2011.
- [6] E. Masry and P. Ishwar, "Field estimation from randomly located, binary noisy sensors," *IEEE Trans. on Inform. Theory*, vol. 55, no. 11, pp. 5197–5210, November 2009.
- [7] D. Marco and D. L. Neuhoff, "Entropy of quantized data at high sampling rates," *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, September Adelaide, Australia, 2005.
- [8] —, "Low-resolution scalar quantization for Gaussian sources and squared error," *IEEE Trans. Inform. Theory*, vol. IT-52, pp. 1689–1697, April 2006.
- [9] D. L. Neuhoff and S. S. Pradhan, "An upper bound to the rate of ideal distributed source coding of densely sampled data," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, 2006.
- [10] D. L. Neuhoff and S. S. Pradhan, "Centralized and distributed lossy source coding of densely sampled Gaussian data with and without transforms," *Proc. IEEE Int. Symp. on Infor. Theory (ISIT)*, 2007.
- [11] T. Berger, "Multiterminal Source Coding," *Information Theory Approach to Communication, (CISM Courses and Lecture Notes No. 229)*, G. Longo, Ed., Wien and New York: Springer-Verlag, 1977.
- [12] U. Grenander and G. Szegö, *Toeplitz forms and their applications*. Berkeley, CA: University of California Press, 1958.
- [13] E. Tyrtshnikov, "A unifying approach to some old and new theorems on distribution and clustering," *Linear Algebra and its Applications*, vol. 232, pp. 1–43, January 1996.
- [14] R. Gray, "On unbounded toelptiz matrices and nonstationary time series with an application to information theory," *Information and control*, vol. 24, pp. 181–196, 1974.
- [15] T. Berger, *Rate distortion theory: A mathematical basis for data compression*. Englewood Cliffs: Prentice Hall, 1971.
- [16] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.
- [17] R. Zamir, "Rate loss in the Wyner-Ziv problem," *IEEE Transactions on Information Theory*, vol. 42, pp. 2073–2084, November 1996.
- [18] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Press, 1992.