

# Universal Multiresolution Source Codes

Michelle Effros, *Member, IEEE*

**Abstract**—A multiresolution source code is a single code giving an embedded source description that can be read at a variety of rates and thereby yields reproductions at a variety of resolutions. The resolution of a source reproduction here refers to the accuracy with which it approximates the original source. Thus, a reproduction with low distortion is a “high-resolution” reproduction while a reproduction with high distortion is a “low-resolution” reproduction. This paper treats the generalization of universal lossy source coding from single-resolution source codes to multiresolution source codes. Results described in this work include new definitions for weakly minimax universal, strongly minimax universal, and weighted universal sequences of fixed- and variable-rate multiresolution source codes that extend the corresponding notions from lossless coding and (single-resolution) quantization to multiresolution quantizers. A variety of universal multiresolution source coding results follow, including necessary and sufficient conditions for the existence of universal multiresolution codes, rate of convergence bounds for universal multiresolution coding performance to the theoretical bound, and a new multiresolution approach to two-stage universal source coding.

**Index Terms**—Embedded, multiuser information theory, progressive transmission, successive refinement, universal source coding.

## I. INTRODUCTION

MULTIRESOLUTION source codes, also known as progressive transmission codes, embedded source codes, and successive refinement codes, are source codes giving embedded data descriptions to allow decoding at a variety of rates (and hence reproductions at a variety of resolutions). For example, Fig. 1 shows a single source description of total length  $\sum_{\ell=1}^L R_{\ell}$  bits per symbol. Reading only the first  $R_1$  bits per symbol yields a source reproduction with per-symbol average distortion  $D_1$ ; reading an additional  $R_2$  bits per symbol, for a total description of  $R_1 + R_2$  bits per symbol, yields a source reproduction with per-symbol average distortion  $D_2$  such that  $D_2 < D_1$ ; and so on.

Multiresolution source coding systems are useful for a wide variety of applications where access to source descriptions at a variety of resolutions enhances system performance. Examples of such applications include wireless communication systems, multiuser systems like the World Wide Web, and database storage and query systems. For example, in a wireless commu-

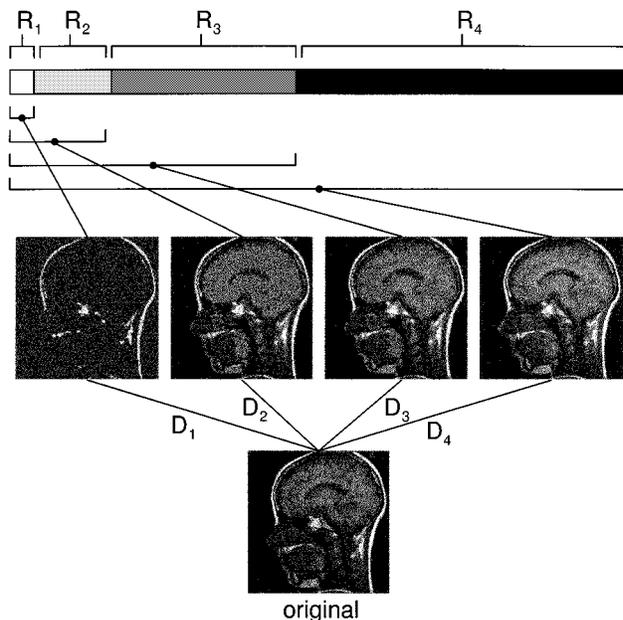


Fig. 1. A four-resolution description of a magnetic resonance brain scan. Decoding the first  $R_1$  bits per symbol of the binary description yields a reproduction with distortion  $D_1$ . Decoding an additional  $R_2$  bits per symbol, for a total rate of  $R_1 + R_2$  bits per symbol, yields a reproduction of distortion  $D_2$ , and so on.

nication system, the rate available for communication may vary as a function of time. Making full use of all of the available rate at any given instant requires either extensive buffering (to spread the rate out evenly across time) or access to a rate-scalable source description (to use the maximum rate available at any particular instant). While the buffering approach guarantees the most consistent quality of performance, it is impractical for many applications. For example, in real-time communications systems carrying voice or video, the introduction of the delays required for buffering is unacceptable. Another family of applications for multiresolution codes falls under the heading of multiuser systems. Multiuser systems are characterized by a variety of users with varying needs and system capabilities accessing the same information source. For example, on the World Wide Web, a single image stored on a website may be accessed by hundreds or thousands of users with vastly different computational capabilities, communications resources, and reproduction quality requirements. For these types of applications, a single-source description that can be decoded at a variety of rates adds greater flexibility to the system, thereby allowing each user to access the same data set with the tradeoff between description length and reproduction quality that is appropriate for that user’s needs. Finally, in large database storage and query systems, multiresolution source codes allow queries to be accomplished at a variety of resolutions so that fast searches may

Manuscript received April 13, 1999; revised June 22, 2000. This work was supported in part by NSF CAREER Award MIP-9501977, NSF Award CCR-9909026, under a grant from the Charles Lee Powell Foundation, and a grant from Caltech’s Lee Center for Advanced Networks. The material in this paper was presented in part at the 1999 IEEE Information Theory Workshop (DECI), Santa Fe, NM, February 1999.

The author is with the Department of Electrical Engineering, MC 136-93, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: effros@caltech.edu).

Communicated by N. Merhav, Associate Editor for Source Coding.

Publisher Item Identifier S 0018-9448(01)07013-4.

be done on the low-resolution source descriptions while more thorough analysis of especially promising subsets of the data may be done using more accurate source reproductions.

Like traditional (single-resolution) source codes, multiresolution source codes are data dependent. Thus, the optimal multiresolution source code for a particular source guarantees good coding performance on that source, but may achieve poor performance on other sources. Fig. 2 illustrates this point by comparing the performance on a magnetic resonance brain scan of two different families of (locally) optimal multiresolution vector quantizers (MRVQs) [1], [2]. In each case, the training and test sets for the MRVQs do not overlap. In one case, the MRVQs are optimized for a collection of magnetic resonance brain scans, thereby yielding a code that is well matched to the data to be compressed. In the other case, the MRVQs are optimized for a portrait image, yielding a mismatched source code. The performance degradation associated with going from a matched code to a mismatched code in this example is quite severe.

In the interest of designing source-independent multiresolution source codes to achieve good performance across a broad class of possible sources, this work, originally presented in [3], introduces the notion of *universal* multiresolution source coding. Roughly speaking, a sequence of multiresolution source codes is here defined to be universal if it asymptotically achieves the best possible performance on every source in some broad class of possible sources. This performance is achieved without *a priori* information about the source in operation. Thus, universal multiresolution source codes are the natural multiresolution extension of traditional (single-resolution) universal quantizers.

While universal single-resolution codes have been studied quite extensively,<sup>1</sup> the notion of universal multiresolution source coding is entirely new. This work therefore begins, in Section II, with a brief introduction to multiresolution source coding definitions and bounds. Definitions for universal sequences of multiresolution source codes and a proof of their existence follows in Section III. Section IV gives constructive arguments for universal source code design, demonstrating the rates of convergence achievable by universal sequences of multiresolution source codes, and briefly discussing issues relevant to practical multiresolution code design.

## II. PRELIMINARIES AND DEFINITIONS

Consider a stationary random process  $\{X_i\}$ ,  $i = 1, 2, \dots$  with alphabet  $A$ , and let  $\Theta$  be some jointly distributed random variable on alphabet  $\Lambda$ . More precisely, assume that  $(\{X_i\}, \Theta)$  is defined on a standard, measurable space  $(\Omega, \mathcal{B})$ , so that regular conditional probabilities exist and the ergodic decomposition holds [13]. For simplicity, we guarantee this standard space by assuming that  $A$  is a Polish alphabet (complete, separable metric space) and either  $\Lambda$  is Polish or  $\Theta$  is a function of  $\{X_i\}$  [13]. (For example,  $\Theta$  might describe an ergodic mode of  $\{X_i\}$ .) Let  $P$  and  $W$  denote the marginal distribution of  $\{X_i\}$  and  $\Theta$ ,

<sup>1</sup>See, for example, [4]–[12]. In particular, [12] introduces the “quantization interpretation” to universal coding upon which this work’s “multiresolution quantization interpretation” is based.

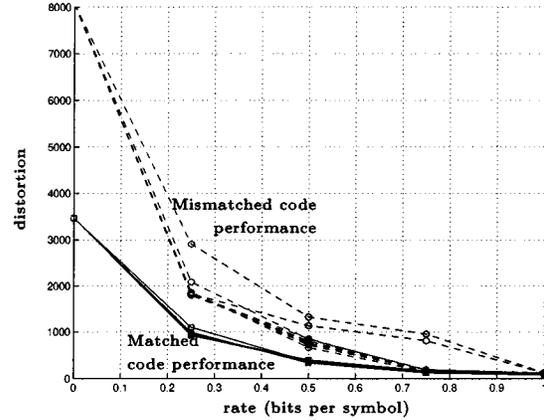


Fig. 2. A demonstration of the data-dependent nature of multiresolution source codes. Each curve gives the rate and distortion performance of a single multiresolution source code on the same brain scan image. The curves differ in the “priorities” placed on the different resolutions and in the training sets used for code design. The performance of codes trained on a training set of portrait images (dashed lines) and brain scans (solid lines) are included. The penalty for mismatch between the training and test sets is quite severe in this example, as shown by the large distortion increase of the solid curves relative to the dashed curves.

respectively, and use  $P_\theta$  to denote the regular conditional distribution of  $\{X_i\}$  given that  $\Theta = \theta$ . For each  $\theta \in \Lambda$ , the distribution  $P_\theta$  is assumed to be stationary but not necessarily ergodic. For any integer  $n \geq 1$ , the addition of a superscript  $n$  to  $P$  or  $P_\theta$ , giving  $P^n$  and  $P_\theta^n$ , denotes the corresponding marginal on an  $n$ -block  $X^n = (X_1, X_2, \dots, X_n)$  of  $\{X_i\}$ .

For any fixed integer  $L \geq 1$ , let  $L$  denote the number of resolutions in a multiresolution source code. For each  $\ell \in \{1, \dots, L\}$ , let  $\hat{A}_{(\ell)}$  and  $\rho_{(\ell)}: A \times \hat{A}_{(\ell)} \rightarrow \mathbb{R}^+$  denote the abstract reproduction alphabet and nonnegative, measurable distortion measure (or *fidelity criterion*), respectively, used in the  $\ell$ th-resolution source reproduction. Assume that for each  $\ell \in \{1, \dots, L\}$  and each  $\theta \in \Lambda$  there exists a reference letter  $y_{(\ell)}^* \in \hat{A}_{(\ell)}$  (possibly dependent on  $\theta$ ) such that  $E_\theta \rho_{(\ell)}(x, y_{(\ell)}^*) < \infty$  for all  $\ell \in \{1, \dots, L\}$ . Further, assume that for each  $\ell \in \{1, \dots, L\}$ ,  $\rho_{(\ell)}(x, y)$  is continuous in  $x \in A$  for all  $y \in \hat{A}_{(\ell)}$ . Typically, the reproduction alphabets and distortion measures are the same for all  $L$  resolutions, in which case  $\hat{A}_{(\ell)} = \hat{A}$  and  $\rho_{(\ell)} = \rho$  for all  $\ell \in \{1, \dots, L\}$ . For any  $x^n \in A^n$ ,  $\ell \in \{1, \dots, L\}$ , and  $y^n \in \hat{A}_{(\ell)}^n$ ,  $\rho_{(\ell)}(x^n, y^n)$  is the additive distortion measure defined by

$$\rho_{(\ell)}(x^n, y^n) = \sum_{i=1}^n \rho_{(\ell)}(x_i, y_i).$$

Define an *L-resolution binary prefix code* to be any prefix code giving  $L$  nested source descriptions, each of which independently meets the prefix condition. For example, consider a set  $\mathcal{S}^L \subseteq \{0, 1\}^*$  of binary strings such that for each  $s^L \in \mathcal{S}^L$ ,  $s^L$  has a unique decomposition into  $L$  fragments; that is,  $s^L = (s_1, s_2, \dots, s_L)$ . For convenience, assume that  $|s_\ell| > 0$  for all  $\ell \in \{1, \dots, L\}$ , where  $|s_\ell|$  denotes the number of bits in the  $\ell$ th fragment.<sup>2</sup> The lengths of the fragments need not be equal. Each string  $s^L \in \mathcal{S}^L$  gives an embedded collection of

<sup>2</sup>An  $L$ -resolution code with a resolution- $\ell$  description of rate 0 can be represented as an  $(L - 1)$ -resolution code.

$L$  binary descriptions  $s^1 = (s_1)$ ,  $s^2 = (s_1, s_2)$ ,  $\dots$ ,  $s^L = (s_1, s_2, \dots, s_L)$ . For each  $\ell \in \{1, \dots, L\}$ , define the set  $\mathcal{S}^\ell$  as

$$\mathcal{S}^\ell = \{s^\ell: \exists s^L \in \mathcal{S}^L \text{ with } s_j = \hat{s}_j \forall 1 \leq j \leq \ell\}.$$

If  $\mathcal{S}^\ell$  is a binary prefix code for each  $\ell \in \{1, \dots, L\}$ , then  $\mathcal{S}^L$  is an  $L$ -resolution binary prefix code.

Given an integer  $n \geq 1$ , let  $Q^{L,n} = g^L \circ f^L$  denote an  $L$ -resolution block quantizer with block length  $n$ , measurable encoder  $f^L: A^n \rightarrow \mathcal{S}^L$ , and measurable decoder

$$g^L: \mathcal{S}^L \rightarrow \hat{A}_{(1)}^n \times \hat{A}_{(2)}^n \times \dots \times \hat{A}_{(L)}^n = \prod_{j=1}^L \hat{A}_{(j)}^n.$$

Thus, the encoder  $f^L$  and decoder  $g^L$  together map the input space  $A^n$  of possible source vectors to the output space  $\prod_{j=1}^L \hat{A}_{(j)}^n$  of  $L$ -resolution source reproductions by way of an  $L$ -resolution binary prefix code  $\mathcal{S}^L$ .

The  $L$ -resolution encoder  $f^L$  and decoder  $g^L$  may be decomposed into  $L$  nested multiresolution encoders  $\{f^1, f^2, \dots, f^L\}$  and decoders  $\{g^1, g^2, \dots, g^L\}$ . For each  $\ell \in \{1, \dots, L\}$  and any  $x^n \in A^n$ , the encoder  $f^\ell: A^n \rightarrow \mathcal{S}^\ell$  and decoder  $g^\ell: \mathcal{S}^\ell \rightarrow \prod_{j=1}^\ell \hat{A}_{(j)}^n$  together map the input space of possible source vectors to the output space  $\prod_{j=1}^\ell \hat{A}_{(j)}^n$  of possible  $\ell$ -resolution reproductions by way of the  $\ell$ -resolution binary prefix code  $\mathcal{S}^\ell$ . The notation  $f_\ell$  is used to describe the  $\ell$ th increment in the binary description given by  $f^L$ , and the notation  $g_\ell$  is used to describe the resolution- $\ell$  source reproduction. Thus, for any  $x^n \in A^n$  and  $\ell \in \{1, \dots, L\}$

$$f^L(x^n) = (s_1, s_2, \dots, s_L) \in \mathcal{S}^L$$

and

$$g^L(s^L) = (y_{(1)}^n, y_{(2)}^n, \dots, y_{(L)}^n) \in \prod_{j=1}^L \hat{A}_{(j)}^n$$

imply  $f_\ell(x^n) = s_\ell$  and  $g_\ell(s^\ell) = y_{(\ell)}^n$ . (Notice that the  $\ell$ th decoder  $g_\ell$  requires only the first  $\ell$  increments  $s^\ell$  of the  $L$ -resolution binary source description  $s^L$ .)

Given any multiresolution block quantizer  $Q^{L,n}$  the (instantaneous) rate and distortion vectors associated with coding source vector  $x^n$  with quantizer  $Q^{L,n}$  are

$$\begin{aligned} \mathbf{r}(x^n, Q^{L,n}) &= (r_1(x^n, Q^{L,n}), \dots, r_L(x^n, Q^{L,n})) \\ &= (|f_1(x^n)|, \dots, |f_L(x^n)|) \\ \mathbf{d}(x^n, Q^{L,n}) &= (d_1(x^n, Q^{L,n}), \dots, d_L(x^n, Q^{L,n})) \\ &= (\rho_{(1)}(x^n, g_1(f^1(x^n))), \dots, \\ &\quad \rho_{(L)}(x^n, g_L(f^L(x^n))))). \end{aligned}$$

Here

$$\mathbf{r}(x^n, Q^{L,n}) = (r_1, \dots, r_L)$$

and

$$\mathbf{d}(x^n, Q^{L,n}) = (d_1, \dots, d_L)$$

imply, for each  $\ell \in \{1, \dots, L\}$ , that an embedded description of total length  $\sum_{j=1}^\ell r_j$  bits is used in describing  $x^n$  to resolution  $\ell$  with  $Q^{L,n}$  and that the distortion associated with the  $\ell$ th

reproduction is  $d_\ell$ . For any  $\theta \in \Lambda$ , the expected rate and distortion in describing  $n$  symbols from  $P_\theta$  with code  $Q^{L,n}$  are

$$\begin{aligned} \mathbf{R}(P_\theta, Q^{L,n}) &= (R_1(P_\theta, Q^{L,n}), \dots, R_L(P_\theta, Q^{L,n})) \\ &= E_\theta \mathbf{r}(X^n, Q^{L,n}) \\ &= (E_\theta |f_1(X^n)|, \dots, E_\theta |f_L(X^n)|) \\ \mathbf{D}(P_\theta, Q^{L,n}) &= (D_1(P_\theta, Q^{L,n}), \dots, D_L(P_\theta, Q^{L,n})) \\ &= E_\theta \mathbf{d}(X^n, Q^{L,n}) \\ &= (E_\theta \rho_{(1)}(X^n, g_1(f^1(X^n))), \dots, \\ &\quad E_\theta \rho_{(L)}(X^n, g_L(f^L(X^n))))). \end{aligned}$$

Multiresolution block quantizers fall into two subclasses: fixed-rate codes and variable-rate codes. If  $Q^{L,n}$  is a *fixed-rate* multiresolution source code, then  $|\mathcal{S}^L|$  is finite and there exists a collection of constants  $c_1, c_2, \dots, c_L$  such that for each  $s^L \in \mathcal{S}^L$ ,  $|s_\ell| = c_\ell$  for all  $\ell \in \{1, \dots, L\}$ . Note that  $c_\ell$  may vary as a function of  $\ell$ . Thus, a fixed-rate  $L$ -resolution code uses a fixed rate in each resolution but the rate may vary from resolution to resolution. If  $Q^{L,n}$  is a *variable-rate* multiresolution source code, then  $|\mathcal{S}^\ell|$  may be finite or countably infinite, and the length of binary description  $s^\ell \in \mathcal{S}^\ell$  is not determined by the resolution  $\ell$ . In the remainder of this work,  $\mathcal{Q}^{\text{fr}}(L, n)$  and  $\mathcal{Q}^{\text{vr}}(L, n)$  denote the classes of fixed- and variable-rate block-length- $n$ ,  $L$ -resolution source codes, respectively. Clearly,  $\mathcal{Q}^{\text{fr}}(L, n) \subseteq \mathcal{Q}^{\text{vr}}(L, n)$ .

Before appropriate definitions can be given for fixed- and variable-rate universal multiresolution quantizers, it is necessary to understand the optimal performance theoretically achievable in each of these scenarios. The discussion that follows summarizes results from [14] on this topic. Earlier multiresolution source coding bounds for memoryless sources appear in [15]–[17]. Since the optimal performances for fixed- and variable-rate codes differ—and hence the resulting definitions for universal codes differ as well—the two families of codes are treated separately in the subsections that follow. A discussion of the similarities and differences between the two families of results follows.

#### A. Fixed-Rate Multiresolution Quantizers

By definition of fixed-rate coding, the rate vector achieved by a fixed-rate multiresolution quantizer does not vary as a function of the data being coded. Thus, optimization of fixed-rate multiresolution quantizers is here performed for a fixed target rate vector. Given a rate vector  $R^L$ , let  $\mathcal{Q}^{\text{fr}, R^L}(L, n) \subseteq \mathcal{Q}^{\text{fr}}(L, n)$  denote the class of fixed-rate, block-length- $n$   $L$ -resolution quantizers achieving (per-symbol) rates  $(1/n)r_\ell(x^n, Q^{L,n}) \leq R_\ell$  for all  $\ell \in \{1, \dots, L\}$  and let  $\mathcal{D}^{\text{fr}}(R^L, P_\theta)$  denote the set of distortion vectors achievable (to arbitrary accuracy) on  $P_\theta$  by some  $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$  for any  $n \geq 1$ . More formally,  $\mathcal{D}^{\text{fr}}(R^L, P_\theta)$  is defined as

$$\mathcal{D}^{\text{fr}}(R^L, P_\theta) = \overline{\left\{ D^L: \exists n \geq 1, Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n) \text{ with } \frac{1}{n} D_\ell(P_\theta, Q^{L,n}) \leq D_\ell \forall \ell \in \{1, \dots, L\} \right\}}$$

where for any set  $F \subseteq \mathbb{R}^L$ ,  $\bar{F}$  denotes the closure (with respect to the Euclidean norm) of  $F$ . Since  $\mathcal{D}^{\text{fr}}(R^L, P_\theta)$  is convex for stationary sources [14, Lemma 2] and closed by definition, it is entirely characterized by its support functional [18, p. 135]  $\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta)$ , here called the *operational fixed-rate distortion-rate function*. For any vector of Lagrangian parameters  $\alpha^L = (\alpha_1, \alpha_2, \dots, \alpha_L)$  such that  $\alpha_\ell \geq 0$  for all  $\ell \in \{1, \dots, L\}$  and  $0 < \sum_{\ell=1}^L \alpha_\ell < \infty$

$$\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) = \inf_{d^L \in \mathcal{D}^{\text{fr}}(P_\theta, R^L)} \sum_{\ell=1}^L \alpha_\ell d_\ell. \quad (1)$$

The vector  $\alpha^L$  may be interpreted as the planar direction or “slope” of a tangential hyperplane supporting the space of achievable  $D^L$ -vectors at a single point. Given a source  $P_\theta$ , the support functional  $\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta)$  describes the optimal performance theoretically achievable (the OPTA) by a quantizer of any dimension.

For any  $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$ , let

$$\begin{aligned} j^{\text{op}}(\alpha^L, x^n, Q^{L,n}) &= \sum_{\ell=1}^L \alpha_\ell d_\ell(x^n, Q^{L,n}), \\ J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}) &= E_\theta j^{\text{op}}(\alpha^L, X^n, Q^{L,n}) \\ &= \sum_{\ell=1}^L \alpha_\ell D_\ell(P_\theta, Q^{L,n}) \end{aligned}$$

denote the instantaneous and expected Lagrangians, respectively, associated with coding  $x^n$  with code  $Q^{L,n}$ . Then, given the assumed stationarity of  $P_\theta$ , an alternative to the definition of  $\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta)$  given in (1) is

$$\begin{aligned} \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) &= \inf_n \hat{J}_n^{\text{fr}}(\alpha^L, R^L, P_\theta) \\ &= \lim_{n \rightarrow \infty} \hat{J}_n^{\text{fr}}(\alpha^L, R^L, P_\theta) \end{aligned} \quad (2)$$

[14, Lemma 14]. Function  $\hat{J}_n^{\text{fr}}(\alpha^L, R^L, P_\theta)$  describes the optimal performance theoretically achievable by any quantizer  $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$  (the  $n$ th-order OPTA), here called the  *$n$ th-order fixed-rate operational distortion-rate function* and given by

$$\hat{J}_n^{\text{fr}}(\alpha^L, R^L, P_\theta) = \inf_{Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)} \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}). \quad (3)$$

The “distortion-rate function” for source  $P_\theta$  and rate vector  $R^L$  is defined as

$$J(\alpha^L, R^L, P_\theta) = \inf_n J_n(\alpha^L, R^L, P_\theta) \quad (4)$$

where

$$\begin{aligned} J_n(\alpha^L, R^L, P_\theta) &= \inf_q \left\{ \frac{1}{n} \sum_{\ell=1}^L \alpha_\ell E_{P_\theta^n} \rho_\ell(X^n, Y_\ell^n): \right. \\ &\quad \left. \frac{1}{n} I_{P_\theta^n}^{q^n}(X^n; \mathbf{Y}_\ell | \mathbf{Y}^{\ell-1}) \leq R_\ell, \right. \\ &\quad \left. \ell \in \{1, \dots, L\} \right\}. \end{aligned}$$

Here  $I_{P_\theta^n}^{q^n}(X^n; \mathbf{Y}_\ell | \mathbf{Y}^{\ell-1})$  denotes, for some fixed “test channel”  $q^n$  from  $A^n$  to  $\prod_{j=1}^L \hat{A}_{(j)}^n$ , a conditional mutual information between  $X^n \in A^n$  and  $\mathbf{Y}_\ell \in \hat{A}_{(\ell)}^n$  given

$$\mathbf{Y}^{\ell-1} \in \prod_{j=1}^{\ell-1} \hat{A}_{(j)}^n.$$

If  $A$  is Polish and  $P_\theta$  is ergodic, then (under the earlier described constraints on the distortion measures  $\{\rho_{(j)}\}_{j=1}^L$ )

$$\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) = J(\alpha^L, R^L, P_\theta) \quad (5)$$

by [14, Theorem 2]. If  $P_\theta$  is not ergodic, then let  $\{P_x\}$  denote the ergodic components of  $P_\theta$ . The ergodic component  $P_x$  is equal to the stationary measure induced by the limiting relative frequencies

$$P_x(F) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_F(x_{i+1}, x_{i+2}, \dots)$$

on events  $F$  in a countable generating field for the  $\sigma$ -field on which  $P$  and  $P_\theta$  for each  $\theta \in \Lambda$  are all defined, and hence the ergodic decomposition is the same for all  $P_\theta$ ,  $\theta \in \Lambda$ . Then, under the same conditions as used in (5)

$$\begin{aligned} \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) &= \int \hat{J}^{\text{fr}}(\alpha^L, R^L, P_x) dP_\theta(x) \\ &= \int J(\alpha^L, R^L, P_x) dP_\theta(x) \\ &=: \bar{J}(\alpha^L, R^L, P_\theta) \end{aligned} \quad (6)$$

by [14, Theorem 5]. Notice that if  $P_\theta$  is ergodic, then

$$\bar{J}(\alpha^L, R^L, P_\theta) = J(\alpha^L, R^L, P_\theta).$$

Thus, (5) and (6) combine to yield a single result. For any stationary ergodic or stationary nonergodic source  $P_\theta$

$$\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) = \bar{J}(\alpha^L, R^L, P_\theta) \quad (7)$$

provided that:  $A$  is Polish; for each  $\ell \in \{1, \dots, L\}$  there exists a reference letter  $y_{(\ell)}^* \in \hat{A}_{(\ell)}$  such that  $E_\theta \rho_\ell(x, y_{(\ell)}^*) < \infty$ ; and for each  $\ell \in \{1, \dots, L\}$ ,  $\rho_\ell(x, y)$  is continuous in  $x \in A$  for all  $y \in \hat{A}_{(\ell)}$ .

### B. Variable-Rate Multiresolution Quantizers

Let  $\mathcal{L}^{\text{vr}}(P_\theta)$  denote the space of rate-distortion vectors  $(R^L, D^L)$  achievable (to within arbitrary accuracy) through variable-rate  $L$ -resolution source coding. More formally

$$\begin{aligned} \mathcal{L}^{\text{vr}}(P_\theta) &= \overline{\left\{ (R^L, D^L): \exists n \geq 1, Q^{L,n} \in \mathcal{Q}^{\text{vr}}(L, n) \text{ with} \right.} \\ &\quad \left. \frac{1}{n} R_\ell(P_\theta, Q^{L,n}) \leq R_\ell, \frac{1}{n} D_\ell(P_\theta, Q^{L,n}) \leq D_\ell, \right.} \\ &\quad \left. \forall \ell \in \{1, \dots, L\} \right\}. \end{aligned}$$

Notice that  $\mathcal{L}^{\text{vr}}(P_\theta) \subset \mathbb{R}^{2L}$  describes the space of rate-distortion vectors  $(R^L, D^L)$  achievable by *any* variable-rate code,

while the set  $\mathcal{D}^{\text{fr}}(P_\theta, R^L) \subseteq \mathbb{R}^L$ , discussed earlier, describes only the subset of distortion vectors  $D^L$  achievable by fixed-rate codes with rate vectors less than  $R^L$ .

If  $P_\theta$  is stationary, then the set  $\mathcal{L}^{\text{vr}}(P_\theta)$  is convex [14, Lemma 2] and closed, and is thus entirely characterized by its support functional [18, p. 135]  $\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta)$ , here called the *operational variable-rate distortion-rate function*, where

$$\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) = \inf_{(r^L, d^L) \in \mathcal{L}^{\text{vr}}(P_\theta)} \sum_{\ell=1}^L [\alpha_\ell d_\ell + \beta_\ell r_\ell]. \quad (8)$$

For any

$$\alpha^L = (\alpha_1, \alpha_2, \dots, \alpha_L) \quad \text{and} \quad \beta^L = (\beta_1, \beta_2, \dots, \beta_L)$$

such that  $\alpha_\ell, \beta_\ell \geq 0$  for all  $\ell \in \{1, \dots, L\}$  and

$$0 < \sum_{\ell=1}^L (\alpha_\ell + \beta_\ell) < \infty$$

$(\alpha^L, \beta^L)$  may be interpreted as the planar direction or ‘‘slope’’ of a tangential hyperplane supporting the space of  $(R^L, D^L)$  vectors achievable through variable-rate coding. The support functional  $\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta)$  describes the optimal performance theoretically achievable on source  $P_\theta$  by an  $L$ -resolution quantizer of any dimension (the OPTA).

For any code  $Q^{L,n} \in \mathcal{Q}^{\text{vr}}(L, n)$ , let

$$\begin{aligned} & j^{\text{op}}(\alpha^L, \beta^L, x^n, Q^{L,n}) \\ &= \sum_{\ell=1}^L [\alpha_\ell d_\ell(x^n, Q^{L,n}) + \beta_\ell r_\ell(x^n, Q^{L,n})] \\ & J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \\ &= E_\theta j^{\text{op}}(\alpha^L, \beta^L, X^n, Q^{L,n}) \\ &= \sum_{\ell=1}^L [\alpha_\ell D_\ell(P_\theta, Q^{L,n}) + \beta_\ell R_\ell(P_\theta, Q^{L,n})] \end{aligned}$$

denote the instantaneous and expected Lagrangians, respectively, associated with coding  $x^n$  with variable-rate code  $Q^{L,n}$ . The equation for  $\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta)$  given in (8) may alternatively be given as

$$\begin{aligned} \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) &= \inf_n \hat{J}_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta) \\ &= \lim_{n \rightarrow \infty} \hat{J}_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta) \end{aligned} \quad (9)$$

[14, Lemma 14]. The function  $\hat{J}_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta)$  describes the optimal performance theoretically achievable by a block-length- $n$ , variable-rate  $L$ -resolution quantizer on source  $P_\theta$  (the  $n$ th-order OPTA), here called the ‘‘ $n$ th-order variable-rate operational distortion-rate function’’ and given by

$$\hat{J}_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta) = \inf_{Q^{L,n} \in \mathcal{Q}^{\text{vr}}(L,n)} \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}). \quad (10)$$

For variable-rate codes, the ‘‘distortion-rate function’’ for source  $P_\theta$  is given by

$$J(\alpha^L, \beta^L, P_\theta) = \inf_n J_n(\alpha^L, \beta^L, P_\theta) \quad (11)$$

where

$$\begin{aligned} J_n(\alpha^L, \beta^L, P_\theta) &= \inf_q \left\{ \frac{1}{n} \sum_{\ell=1}^L \left( \alpha_\ell E_{P_\theta^n} \rho_\ell \left( X^n, Y_{(\ell)}^n \right) \right. \right. \\ &\quad \left. \left. + \beta_\ell I_{P_\theta^n} \left( X^n; \mathbf{Y}_\ell | \mathbf{Y}^{\ell-1} \right) \right) \right\}. \end{aligned}$$

If source  $P_\theta$  is ergodic, then

$$\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) = J(\alpha^L, \beta^L, P_\theta)$$

[14, Theorem 3]. If  $P_\theta$  is not ergodic, then, under the conditions of (5)

$$\begin{aligned} \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) &= \int \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_x) dP_\theta(x) \\ &= \int J(\alpha^L, \beta^L, P_x) dP_\theta(x) \\ &= J(\alpha^L, \beta^L, P_\theta) \end{aligned}$$

since both  $\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta)$  and  $J(\alpha^L, \beta^L, P_\theta)$  admit the ergodic decomposition [14, Theorems 7 and 8]. Thus, once again, the results can be summarized in a single equation. For any stationary-ergodic or stationary-nonergodic source  $P_\theta$

$$\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) = J(\alpha^L, \beta^L, P_\theta) \quad (12)$$

provided that:  $A$  is Polish; for each  $\ell \in \{1, \dots, L\}$  there exists a reference letter  $y_{(\ell)}^* \in \hat{A}_{(\ell)}$  such that  $E_\theta \rho_\ell(x, y_{(\ell)}^*) < \infty$ ; and for each  $\ell \in \{1, \dots, L\}$ ,  $\rho_\ell(x, y)$  is continuous in  $x \in A$  for all  $y \in \hat{A}_{(\ell)}$ .

### III. UNIVERSAL MULTIREOLUTION QUANTIZERS

#### A. Definitions

For any  $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$ , define the *fixed-rate- $R^L$  redundancy* of  $Q^{L,n}$  as

$$\begin{aligned} \Delta^{\text{fr}}(\alpha^L, R^L, P_\theta, Q^{L,n}) \\ &= \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}) - \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta). \end{aligned}$$

Likewise, for any  $Q^{L,n} \in \mathcal{Q}^{\text{vr}}(L, n)$ , define the *variable-rate redundancy* as

$$\begin{aligned} \Delta^{\text{vr}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \\ &= \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) - \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta). \end{aligned}$$

By (2), (3), (9), and (10), both redundancies are nonnegative. Given a sequence of codes  $\{Q^{L,n}\}_{n=1}^\infty$  such that

$$Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n), \quad \text{for all } n \geq 1$$

$\{Q^{L,n}\}$  is a universal sequence of fixed-rate multiresolution quantizers if

$$\Delta^{\text{fr}}(\alpha^L, R^L, P_\theta, Q^{L,n}) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Similarly, given a sequence of codes  $\{Q^{L,n}\}_{n=1}^\infty$  such that

$$Q^{L,n} \in \mathcal{Q}^{\text{vr}}(L, n), \quad \text{for all } n \geq 1$$

$\{Q^{L,n}\}$  is a universal sequence of variable-rate multiresolution quantizers if

$$\Delta^{\text{vr}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

More specifically, a sequence of codes is *weakly minimax universal* if the convergence is pointwise in  $\theta$ , *strongly minimax universal* if the convergence is uniform over  $\theta \in \Lambda$ , and *weighted universal* if the convergence is in expectation with respect to  $W$ .

### B. Discussion and Proof of Existence

It is interesting to notice that the definitions for the distortion-rate function given in (4) and (11) differ. In particular, for fixed-rate coding, the surface was parameterized by  $(\alpha^L, R^L)$ , while for variable-rate coding the surface was parameterized by  $(\alpha^L, \beta^L)$ . It is therefore important to check that the set of  $(R^L, D^L)$  points described by these two functionals are the same. This check is easily accomplished as

$$J(\alpha^L, R^L, P_\theta) = \inf_{\beta^L} \left[ J(\alpha^L, \beta^L, P_\theta) - \sum_{\ell=1}^L \beta_\ell R_\ell \right]$$

$$J(\alpha^L, \beta^L, P_\theta) = \inf_{R^L} \left[ J(\alpha^L, R^L, P_\theta) + \sum_{\ell=1}^L \beta_\ell R_\ell \right]$$

which together imply that the spaces described by the two functionals are identical. It is interesting to note, however, that while  $J(\alpha^L, \beta^L, P_\theta)$  meets the ergodic decomposition,  $J(\alpha^L, R^L, P_\theta)$  does not [14]. Roughly speaking, the lower convex hull of the set of achievable  $(R^L, D^L)$  vectors for a stationary nonergodic source is equal to the weighted combination of the lower convex hulls of achievable  $(R^L, D^L)$  vectors for that source's stationary ergodic components, where the combination is taken at points of equal "slope" rather than points of equal rate. This parallels the findings for single-resolution codes given in [19].

As a result of the above discrepancies, the optimal performance theoretically achievable by block multiresolution source coding in general differs for fixed- and variable-rate source codes. In particular, if  $P_\theta$  is nonergodic, then

$$\begin{aligned} & \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) \\ &= \int \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_x) dP_\theta(x) \\ &= \int J(\alpha^L, \beta^L, P_x) dP_\theta(x) \\ &= \int \inf_{R^L} \left[ J(\alpha^L, R^L, P_x) + \sum_{\ell=1}^L \beta_\ell R_\ell \right] dP_\theta(x) \\ &= \int \inf_{R^L} \left[ \hat{J}^{\text{fr}}(\alpha^L, R^L, P_x) + \sum_{\ell=1}^L \beta_\ell R_\ell \right] dP_\theta(x) \\ &\leq \inf_{R^L} \left[ \int \hat{J}^{\text{fr}}(\alpha^L, R^L, P_x) dP_\theta(x) + \sum_{\ell=1}^L \beta_\ell R_\ell \right] \\ &= \inf_{R^L} \left[ \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) + \sum_{\ell=1}^L \beta_\ell R_\ell \right] \end{aligned}$$

and, thus, the space of distortion-rate vectors achievable through fixed-rate coding is a subset of the space of distortion-rate vectors achievable through variable-rate coding, as shown in [14]. Thus, a sequence of fixed-rate codes that is "optimal" for source  $P_\theta$  in the sense that it asymptotically achieves the fixed-rate OPTA  $\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta)$  is not necessarily optimal for source  $P_\theta$  from a variable-rate coding perspective. As a result, while all fixed-rate multiresolution quantizers are also variable-rate multiresolution quantizers, a *universal* sequence of fixed-rate multiresolution quantizers is not necessarily a *universal* sequence of variable-rate multiresolution quantizers. (Notice that the same observation holds also in the single-resolution case.)

The relationship between weakly minimax universal, strongly minimax universal, and weighted universal sequences of *single-resolution* source codes is discussed at some length in [12]. The conclusion of that argument, which applies equally well to sequences of *multiresolution* codes, is that from a practical coding perspective, the differences between the three types of universality seem rather minor. For example, the existence of a weighted universal sequence of codes implies, for any  $\epsilon > 0$ , the existence of a strongly minimax universal sequence of codes on some set  $\Lambda_\epsilon \subseteq \Lambda$  with  $W(\Lambda_\epsilon) > 1 - \epsilon$ . Thus, the existence of the "weakest" form of universal code in some sense implies the existence of the "strongest" form of universal code. The remainder of this work includes results on all three forms of universality but emphasizes the weighted universal case.

The first step in the study of universal sequences of multiresolution quantizers is a proof of their existence. The following theorem gives a necessary and sufficient condition for the existence of weighted universal sequences of multiresolution quantizers under the assumption that  $P_\theta$  describes a stationary, ergodic source for each  $\theta \in \Lambda$ . The proof of Theorem 1 appears in the Appendix.

*Theorem 1:* Assume that  $\{X_i\}$  is stationary and ergodic for each  $\theta \in \Lambda$ . Weighted universal codes exist for  $(\{X_i\}, \Theta)$  if and only if the OPTA has an ergodic decomposition.

Thus, under the conditions of Theorem 1, weighted universal sequences of multiresolution quantizers exist for the fixed-rate case if and only if

$$\hat{J}^{\text{fr}}(\alpha^L, R^L, P) = \int \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) dW(\theta)$$

and, for the variable-rate case, if and only if

$$\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P) = \int \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) dW(\theta).$$

Both of these ergodic decompositions are satisfied under the following three conditions:  $A$  is Polish; for each  $\ell \in \{1, \dots, L\}$  there exists a reference letter  $y_{(\ell)}^* \in \hat{A}_{(\ell)}$  such that  $E_\theta \rho(x, y_{(\ell)}^*) < \infty$ ; and for each  $\ell \in \{1, \dots, L\}$ ,  $\rho_{(\ell)}(x, y)$  is continuous in  $x \in A$  for all  $y \in \hat{A}_{(\ell)}$ . Theorem 1 justifies the differences between the parameterizations of the fixed- and variable-rate redundancies. In particular, fixed-rate codes *must* be compared at the same rate  $R^L$  rather than across rates at the same Lagrangian parameter  $\beta^L$  since the Lagrangian

$$\hat{J}^{\text{fr}}(\alpha^L, \beta^L, P_\theta) = \inf_{R^L} [\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) + \beta_\ell R_\ell]$$

does not have an ergodic decomposition [14].

The rate at which the expected redundancy of a universal sequence of multiresolution quantizers converges to zero serves as a comparative performance measure for competing algorithms. The redundancies  $\Delta_n^{\text{fr}}(\alpha^L, R^L, P_\theta, Q^{L,n})$  and  $\Delta_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta, Q^{L,n})$  may each be broken into a sum of two nonnegative terms, where the first term describes the difference between the performance of a particular code and the  $n$ th-order OPTA and the second term describes the difference between the  $n$ th-order OPTA and the OPTA. This work follows the lead of [12] by focusing exclusively on the first term, called the  *$n$ th-order redundancy* and given, for the fixed- and variable-rate cases, respectively, by

$$\begin{aligned}\Delta_n^{\text{fr}}(\alpha^L, R^L, P_\theta, Q^{L,n}) &= \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}) - \hat{J}_n^{\text{fr}}(\alpha^L, R^L, P_\theta) \\ \Delta_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) &= \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) - \hat{J}_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta).\end{aligned}$$

The  $n$ th-order redundancy describes the price paid for universality and is the only part of the redundancy that varies from quantizer to quantizer. The rate of convergence of the  $n$ th-order OPTA to the OPTA, which describes the penalty associated with finite dimensionality, has been studied for single-resolution quantizers in works such as [20], [21], [9], [22]; for multiresolution quantizers, this rate of convergence remains an interesting open topic for future investigation.

To simplify the following discussion on rates of convergence, the notation for fixed- and variable-rate coding are here combined into a single notation

$$\begin{aligned}\Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n}) &= \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) - \hat{J}_n(\alpha^L, \beta^L, P_\theta).\end{aligned}\quad (13)$$

For variable-rate coding

$$\hat{J}_n(\alpha^L, \beta^L, P_\theta) = \hat{J}_n^{\text{vr}}(\alpha^L, \beta^L, P_\theta)$$

and  $\Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n})$  is exactly the  $n$ th-order redundancy of code  $Q^{L,n}$ . For fixed-rate coding

$$\begin{aligned}\Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n}) &= \left[ \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell R_\ell(P_\theta, Q^{L,n}) \right] \\ &\quad - \left[ \hat{J}_n^{\text{fr}}(\alpha^L, R^L, P_\theta) + \sum_{\ell=1}^L \beta_\ell R_\ell \right] \\ &= \left[ \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}) - \hat{J}_n^{\text{fr}}(\alpha^L, R^L, P_\theta) \right] \\ &\quad + \sum_{\ell=1}^L \beta_\ell \left[ \frac{1}{n} R_\ell(P_\theta, Q^{L,n}) - R_\ell \right] \\ &= \Delta_n^{\text{fr}}(\alpha^L, R^L, P_\theta, Q^{L,n}) \\ &\quad + \sum_{\ell=1}^L \beta_\ell \left[ \frac{1}{n} R_\ell(P_\theta, Q^{L,n}) - R_\ell \right].\end{aligned}$$

Strictly speaking, this redundancy should be calculated only for quantizers in the class  $\mathcal{Q}^{\text{fr}, R^L}(L, n)$  of fixed-rate quantizers with

$$\frac{1}{n} R_\ell(P_\theta, Q^{L,n}) \leq R_\ell, \quad \text{for all } \ell \in \{1, \dots, L\}.$$

However, for any sequence of quantizers  $\{Q^{L,n}\}_{n=1}^\infty$  such that  $Q^{L,n} \in \mathcal{Q}^{\text{fr}}(L, n)$  for all  $n \geq 1$  but  $(1/n)R_\ell(P_\theta, Q^{L,n}) > R_\ell$  for one or more  $\ell \in \{1, \dots, L\}$ , if  $\Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n})$  is bounded above by some sequence  $c_n(\theta)$  that converges to zero, then there exists another sequence of fixed-rate quantizers  $\{Q'^{L,n}\}_{n=1}^\infty$  such that  $Q'^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$  and  $\Delta_n^{\text{fr}}(\alpha^L, R^L, P_\theta, Q'^{L,n})$  likewise converges to zero for any  $\beta^L$  appropriately chosen. A formal statement and proof of this result appear in Lemma 1 in the Appendix. This argument, which parallels an argument given for single-resolution codes, justifies the use of the same form of redundancy term in both fixed- and variable-rate multiresolution quantizers.

#### IV. TWO-STAGE UNIVERSAL MULTIREOLUTION QUANTIZERS

A two-stage single-resolution quantizer (e.g., [23]–[25], [12]) is a source code that describes a data sequence in two stages. The first-stage description encodes the choice of a single block-length- $n$  quantizer from some collection of available block-length- $n$  source codes. The second-stage description encodes a data vector  $x^n$  using the chosen code. The discussion that follows treats the rate of convergence of weighted universal sequences of two-stage multiresolution quantizers. The two-stage multiresolution quantizers introduced here generalize the two-stage approach to universal lossy source coding described in [12] from the single-resolution case to the multiresolution case.

An interesting question arises in the generalization of two-stage coding from single-resolution quantization to multiresolution quantization. In which resolution of a multiresolution source description should the first-stage source description be given? Clearly, one possible answer to this question is that the first-stage description can be given entirely in the first-resolution data description. One possible justification for this choice arises from the fact that of the incremental descriptions, only the first-resolution source description is used by all decoders  $g_1, g_2, \dots, g_L$ . Thus, only by incorporating the first-stage coding information into the first-resolution source description can this choice be fully known to all of the incremental decoders in  $g^L$ . This choice is also reasonable because the length of the first-stage description, when amortized over the length  $n$  of the data sequence to which it pertains, is asymptotically equal to zero for any universal code.

Yet placing the entire first-stage description in the first-resolution source description also has drawbacks. In particular, for practical coding, the coding dimension  $n$  may be quite small. In this case, the optimal per-symbol rate needed to give the first-stage description may represent a significant portion of (or even exceed) the available first-resolution coding rate  $R_1$ . Placing the entire first-stage description in the first resolution may therefore place an undue burden on the first-resolution source code.

One alternative for alleviating the burden on the first-resolution code, at least in part, is to spread the cost of the first-stage description over all  $L$  resolutions. This approach is motivated

not only from a practical coding perspective, as described above, but also from an optimal universal coding perspective. In particular, for single-resolution codes, the optimal first-stage coding rate for any dimension  $n$  may vary (by a constant factor) as a function of the target source coding rate (see, for example, [12, Sec. III, Case 1]). Since each resolution in a multiresolution source code has a different target rate, it seems reasonable to postulate that the optimal first-stage coding rate for the first resolution may differ from the optimal first-stage coding rate for later resolutions, making the idea of spreading the first-stage description over multiple resolutions even more attractive. This approach generalizes the quantization interpretation of two-stage universal source codes from single-resolution to multiresolution first-stage codes. That is, just as the first-stage encoder and decoder of a single-resolution source code may be viewed as a quantizer quantizing the space of dimension- $n$  single-resolution quantizers, the first-stage encoder and decoder of a multiresolution quantizer may be viewed as a multiresolution quantizer quantizing—first coarsely and then to higher and higher resolutions—the space of dimension- $n$  multiresolution quantizers.

This section includes descriptions of a variety of rate of convergence results for weakly minimax and weighted universal sequences of multiresolution quantizers. The arguments appearing in this section give upper bounds on the optimal rate of convergence using constructive proofs with two-stage multiresolution quantizers. The use of single-resolution first-stage source descriptions leads to arguments similar to those resulting from the quantization interpretation of two-stage coding given in [12]. As a result, the proofs of results using single-resolution first-stage descriptions are excluded for the sake of brevity. The use of multiresolution first-stage source descriptions requires additional care, and, thus, the multiresolution first-stage source coding results are included in their entirety.

Let  $\tilde{f}: A^n \rightarrow \tilde{S}^L$  be a measurable mapping from the source alphabet  $A^n$  to some  $L$ -resolution binary prefix code  $\tilde{S}^L$ , and let  $\{Q_{\tilde{s}^L}^{L,n}\}_{\tilde{s}^L \in \tilde{S}^L}$ , be a collection of block-length- $n$   $L$ -resolution quantizers. For each  $\tilde{s}^L \in \tilde{S}^L$ , the quantizer  $Q_{\tilde{s}^L}^{L,n} = g_{\tilde{s}^L}^L \circ f_{\tilde{s}^L}^L$  has measurable  $L$ -resolution encoder

$$f_{\tilde{s}^L}^L = (f_{\tilde{s}^1,1}, f_{\tilde{s}^2,2}, \dots, f_{\tilde{s}^L,L})$$

and decoder

$$g_{\tilde{s}^L}^L = (g_{\tilde{s}^1,1}, g_{\tilde{s}^2,2}, \dots, g_{\tilde{s}^L,L}).$$

Thus, for each  $\ell \in \{1, \dots, L\}$ , the  $\ell$ th incremental encoder  $f_{\tilde{s}^{\ell},\ell}^L$  and decoder  $g_{\tilde{s}^{\ell},\ell}^L$  are indexed by the first  $\ell$  increments of the description of  $\tilde{s}^L$ . A multiresolution quantizer  $Q^{L,n} = g^L \circ f^L$  is a two-stage multiresolution quantizer if its encoder  $f^L = (f_1, f_2, \dots, f_L)$  and decoder  $g^L = (g_1, g_2, \dots, g_L)$  take the following forms:

$$f^L(x^n) = \left( \tilde{f}_1(x^n) f_{\tilde{f}_1(x^n),1}^L(x^n), \tilde{f}_2(x^n) f_{\tilde{f}_2(x^n),2}^L(x^n), \dots, \tilde{f}_L(x^n) f_{\tilde{f}_L(x^n),L}^L(x^n) \right)$$

and

$$g^L(f^L(x^n)) = \left( g_{\tilde{f}_1(x^n),1}^L \left( f_{\tilde{f}_1(x^n),1}^L(x^n) \right), g_{\tilde{f}_2(x^n),2}^L \left( f_{\tilde{f}_2(x^n),2}^L(x^n) \right), \dots, g_{\tilde{f}_L(x^n),L}^L \left( f_{\tilde{f}_L(x^n),L}^L(x^n) \right) \right).$$

Here  $f_{\tilde{s}^{\ell}}(x^n) = \tilde{s}_{\ell} s_{\ell}$  indicates that  $f_{\tilde{s}^{\ell}}(x^n)$  is formed by concatenating binary strings  $\tilde{s}_{\ell}$  and  $s_{\ell}$  and

$$f_{\tilde{f}^{\ell}(x^n)}^L(x^n) = \left( f_{\tilde{f}_1(x^n),1}^L(x^n), f_{\tilde{f}_1(x^n),\tilde{f}_2(x^n),2}^L(x^n), \dots, f_{\tilde{f}_1(x^n),\tilde{f}_2(x^n),\dots,\tilde{f}_{\ell}(x^n),\ell}^L(x^n) \right).$$

The case where the entire first-stage source description is given in the first resolution (using a single-resolution first-stage code) may be treated by setting  $f_{\tilde{s}^{\ell}}(x^n)$  to the empty string for all  $\ell > 1$  and all  $x^n$ . (Rate-0 descriptions in the second-stage code are again disallowed.)

Now suppose that for each  $\tilde{s}^L \in \tilde{S}^L$ ,  $Q_{\tilde{s}^L}^{L,n}$  is a product code created by using some block-length- $m$  multiresolution quantizer  $Q_{\tilde{s}^L}^{L,m}$   $n/m$  times. (For simplicity, assume that  $m$  divides  $n$  evenly.) Then for any  $x^n = (x_{(1)}^m, x_{(2)}^m, \dots, x_{(n/m)}^m)$  and any  $\ell \in \{1, \dots, L\}$

$$f_{\tilde{s}^{\ell}}(x^n) = \tilde{f}_{\tilde{s}^{\ell}}(x^n) f_{\tilde{f}^{\ell}(x^n),\ell} \left( x_{(1)}^m \right) f_{\tilde{f}^{\ell}(x^n),\ell} \left( x_{(2)}^m \right) \dots f_{\tilde{f}^{\ell}(x^n),\ell} \left( x_{(n/m)}^m \right)$$

and

$$g_{\tilde{s}^{\ell}}(f^{\ell}(x^n)) = g_{\tilde{f}^{\ell}(x^n),\ell} \left( f_{\tilde{f}^{\ell}(x^n),\ell}^{\ell} \left( x_{(1)}^m \right) \right) g_{\tilde{f}^{\ell}(x^n),\ell} \left( f_{\tilde{f}^{\ell}(x^n),\ell}^{\ell} \left( x_{(2)}^m \right) \right) \dots g_{\tilde{f}^{\ell}(x^n),\ell} \left( f_{\tilde{f}^{\ell}(x^n),\ell}^{\ell} \left( x_{(n/m)}^m \right) \right).$$

The first stage of our multiresolution quantizer may itself be thought of as a multiresolution quantizer  $\tilde{Q}^{L,n} = \tilde{g}^L \circ \tilde{f}^L$  with encoder  $\tilde{f}^L: A^n \rightarrow \tilde{S}^L$  as described above and decoder  $\tilde{g}^L: \tilde{S}^L \rightarrow \mathcal{Q}(L, m) = \{Q_{\tilde{s}^L}^{L,m}: \tilde{s} \in \tilde{S}\}$ . Thus, the first-stage quantizer maps the input space of possible data vectors to the output space of multiresolution quantizers by way of a binary multiresolution prefix code  $\tilde{S}^L$ .

Given a two-stage multiresolution quantizer  $Q^{L,n}$  with first-stage multiresolution quantizer  $\tilde{Q}^{L,n}$ , the (instantaneous) rate and distortion of the two-stage code are

$$\begin{aligned} \mathbf{r}(x^n, Q^{L,n}) &= \left( \left| \tilde{f}_1(x^n) \right| + \sum_{i=1}^{n/m} \left| f_{\tilde{f}_1(x^n),1}^L \left( x_{(i)}^m \right) \right|, \dots, \right. \\ &\quad \left. \left| \tilde{f}_L(x^n) \right| + \sum_{i=1}^{n/m} \left| f_{\tilde{f}_L(x^n),L}^L \left( x_{(i)}^m \right) \right| \right) \\ &= \tilde{\mathbf{r}} \left( x^n, \tilde{Q}^{L,n} \right) + \sum_{i=1}^{n/m} \mathbf{r} \left( x_{(i)}^m, Q_{\tilde{f}_L(x^n)}^{L,m} \right) \\ \mathbf{d}(x^n, Q^{L,n}) &= \left( \sum_{i=1}^{n/m} \rho_{(1)} \left( x_{(i)}^m, g_{\tilde{f}_1(x^n),1}^L \left( f_{\tilde{f}_1(x^n),1}^L \left( x_{(i)}^m \right) \right) \right), \dots, \right. \\ &\quad \left. \sum_{i=1}^{n/m} \rho_{(L)} \left( x_{(i)}^m, g_{\tilde{f}_L(x^n),L}^L \left( f_{\tilde{f}_L(x^n),L}^L \left( x_{(i)}^m \right) \right) \right) \right) \\ &= \sum_{i=1}^{n/m} \mathbf{d} \left( x_{(i)}^m, Q_{\tilde{f}_L(x^n)}^{L,m} \right) \end{aligned}$$

where

$$\tilde{r}(x^n, \tilde{Q}^{L,n}) = \left( \left| \tilde{f}_1(x^n) \right|, \dots, \left| \tilde{f}_L(x^n) \right| \right)$$

denotes the (instantaneous) rate of the first-stage encoder  $\tilde{f}^L$ . Combining these two results, the redundancy  $\Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n})$  of the two-stage code  $Q^{L,n}$  equals

$$\begin{aligned} & \Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \\ &= \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) - \hat{J}_n(\alpha^L, \beta^L, P_\theta) \\ &= \frac{1}{n} \int \left[ \sum_{\ell=1}^L \alpha_\ell \sum_{i=1}^{n/m} d_\ell(x_{(i)}^m, Q_{\tilde{f}^L(x^n)}^{L,m}(x_{(i)}^m)) \right. \\ & \quad \left. + \sum_{\ell=1}^L \beta_\ell \left( \tilde{r}_\ell(x^n, \tilde{Q}^{L,n}) \right. \right. \\ & \quad \left. \left. + \sum_{i=1}^{n/m} r_\ell(x_{(i)}^m, Q_{\tilde{f}^L(x^n)}^{L,m}) \right) \right] dP_\theta(x^n) \\ & \quad - \hat{J}_n(\alpha^L, \beta^L, P_\theta) \\ &= \frac{1}{n} \int \left[ \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell(x^n, \tilde{Q}^{L,n}) \right. \\ & \quad \left. + \sum_{i=1}^{n/m} j^{\text{op}}(\alpha^L, \beta^L, x_{(i)}^m, Q_{\tilde{f}^L(x^n)}^{L,m}) \right] dP_\theta(x^n) \\ & \quad - \hat{J}_n(\alpha^L, \beta^L, P_\theta). \end{aligned} \quad (14)$$

The first-stage encoder that minimizes (14) minimizes the integrand pointwise, giving the following result.

*Nearest Neighbor First-Stage Multiresolution Encoders:* For any two-stage multiresolution source code  $Q^{L,n}$  with first-stage quantizer  $\tilde{Q}^{L,n} = \tilde{g}^L \circ \tilde{f}^L$ , there exists a two-stage multiresolution source code  $Q_*^{L,n}$  with first-stage quantizer  $\tilde{Q}_*^{L,n} = \tilde{g}^L \circ \tilde{f}_*^L$  such that

$$\Delta_n(\alpha^L, \beta^L, P_\theta, Q_*^{L,n}) \leq \Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \quad (15)$$

for all  $\theta \in \Lambda$ . The first-stage encoder  $\tilde{f}_*^L$  is given by

$$\begin{aligned} & \tilde{f}_*^L(x^n) \\ &= \arg \min_{\tilde{s}^L \in \tilde{\mathcal{S}}^L} \left[ \sum_{\ell=1}^L \beta_\ell |\tilde{s}_\ell| + \sum_{i=1}^{n/m} j^{\text{op}}(\alpha^L, \beta^L, x_{(i)}^m, Q_{\tilde{s}^L}^{L,m}) \right] \end{aligned} \quad (16)$$

for each  $x^n \in A^n$  and is called an ‘‘optimal’’ or ‘‘nearest neighbor’’ first-stage multiresolution encoder with respect to first-stage multiresolution decoder  $\tilde{g}^L$ . When  $\tilde{\mathcal{S}}^L$  is countably infinite, a proof is required to demonstrate that the minimum in (16) is always achieved. This proof appears in Lemma 2 in the Appendix.

The nearest neighbor first-stage multiresolution encoder defined above gives a two-stage coding analogy to the nearest neighbor multiresolution quantizer defined in [1], [2] and used

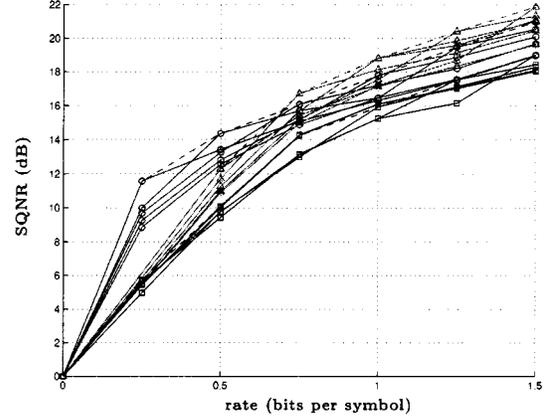


Fig. 3. Experimental results for single- and multiresolution nonuniversal and universal codes. Three sets of curves are included. The squares show the performance of a collection of VQs (connected by a dashed line to note that the performances are achieved by *different* codes) and a family of MRVQs. The circles and triangles show corresponding results for WUVQ (dashed line) and WUMRVQ (solid lines) with two different first-stage coding rates. The weighted universal codes give up to 6 dB of performance improvement over their nonuniversal counterparts on the collection of medical brain scans used in these experiments.

there for practical multiresolution vector quantizer (MRVQ) design. Just as single-resolution vector quantization (VQ) generalizes to weighted universal vector quantization (WUVQ) using an iterative descent design technique [23], [12], MRVQ likewise generalizes to weighted universal multiresolution vector quantization (WUMRVQ) using a similar approach. Results demonstrating the experimental performance benefits of WUMRVQ over MRVQ appear in Fig. 3.

While the optimal encoder is easily implemented in practical weighted universal multiresolution vector quantizers, for the purpose of theoretical analysis, the optimal encoder is more difficult to analyze than is the omniscient multiresolution first-stage encoder defined next and used through the remainder of this work.

*Omniscient First-Stage Multiresolution Encoders:* Consider a two-stage multiresolution source code  $Q_o^{L,n}$  with first-stage quantizer  $\tilde{Q}_o^{L,n} = \tilde{g}^L \circ \tilde{f}_o^L$ . The first-stage encoder  $\tilde{f}_o^L$  is called an omniscient first-stage encoder if it works directly on  $\theta \in \Lambda$  rather than on  $x^n \in A^n$ . Thus, the omniscient first-stage encoder  $\tilde{f}_o^L: \Lambda \rightarrow \tilde{\mathcal{S}}^L$  chooses a second-stage code based on the true underlying source in operation (rather than the vector  $x^n$  produced by that source). The omniscient first-stage encoder, like the original first-stage encoder, is assumed to be measurable. For any omniscient first-stage encoder  $\tilde{f}_o^L$

$$\Delta_n(\alpha^L, \beta^L, P_\theta, Q_*^{L,n}) \leq \Delta_n(\alpha^L, \beta^L, P_\theta, Q_o^{L,n})$$

for all  $\theta \in \Lambda$  by (15). Thus, the omniscient multiresolution first-stage encoder, while not achievable in practice, does no better than the optimal (nearest neighbor) multiresolution first-stage encoder, which can be achieved in practice. As a result, the redundancy of the best two-stage multiresolution quantizer with an omniscient first-stage encoder gives an upper bound on the redundancy for the best two-stage multiresolution quantizer. The remainder of this section focuses on the derivation of such a bound.

Given a two-stage multiresolution quantizer  $Q_o^{L,n}$  with omniscient first-stage quantizer  $\tilde{Q}_o^{L,n} = \tilde{g}^L \circ \tilde{f}_o^L$ , the (instantaneous) rate and distortion of the two-stage code are

$$\begin{aligned} \mathbf{r}(x^n, Q_o^{L,n}) &= \left( |\tilde{f}_{o,1}(\theta)| + \sum_{i=1}^{n/m} \left| f_{\tilde{f}_o^L(\theta),1} \left( x_{(i)}^m \right) \right|, \dots, \right. \\ &\quad \left. |\tilde{f}_{o,L}(\theta)| + \sum_{i=1}^{n/m} \left| f_{\tilde{f}_o^L(\theta),L} \left( x_{(i)}^m \right) \right| \right) \\ &= \tilde{\mathbf{r}} \left( \theta, \tilde{Q}_o^{L,n} \right) + \sum_{i=1}^{n/m} \mathbf{r} \left( x_{(i)}^m, Q_{\tilde{f}_o^L(\theta)}^{L,m} \right), \\ \mathbf{d}(x^n, Q_o^{L,n}) &= \left( \sum_{i=1}^{n/m} \rho_{(1)} \left( x_{(i)}^m, g_{\tilde{f}_o^L(\theta),1} \left( f_{\tilde{f}_o^L(\theta)}^1 \left( x_{(i)}^m \right) \right) \right) \right), \dots, \\ &\quad \left( \sum_{i=1}^{n/m} \rho_{(L)} \left( x_{(i)}^m, g_{\tilde{f}_o^L(\theta),L} \left( f_{\tilde{f}_o^L(\theta)}^L \left( x_{(i)}^m \right) \right) \right) \right) \\ &= \sum_{i=1}^{n/m} \mathbf{d} \left( x_{(i)}^m, Q_{\tilde{f}_o^L(\theta)}^{L,m} \right) \end{aligned}$$

where  $\tilde{\mathbf{r}}(\theta, \tilde{Q}_o^{L,n})$  denotes the (instantaneous) rate of the omniscient first-stage encoder  $\tilde{f}_o^L$ . Thus, the redundancy of the two-stage code  $Q_o^{L,n}$  equals

$$\begin{aligned} \Delta_n(\alpha^L, \beta^L, P_\theta, Q_o^{L,n}) &= \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q_o^{L,n}) - \hat{J}_n(\alpha^L, \beta^L, P_\theta) \\ &= \frac{1}{n} \left[ \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell \left( \theta, \tilde{Q}_o^{L,n} \right) \right. \\ &\quad \left. + E_\theta \sum_{i=1}^{n/m} j^{\text{op}} \left( \alpha^L, \beta^L, X_{(i)}^m, Q_{\tilde{f}_o^L(\theta)}^{L,m} \right) \right] \\ &\quad - \hat{J}_n(\alpha^L, \beta^L, P_\theta) \\ &= \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell \left( \theta, \tilde{Q}_o^{L,n} \right) \\ &\quad + \frac{1}{m} J^{\text{op}} \left( \alpha^L, \beta^L, P_\theta, Q_{\tilde{f}_o^L(\theta)}^{L,m} \right) \\ &\quad - \hat{J}_n(\alpha^L, \beta^L, P_\theta). \end{aligned}$$

The resulting redundancy can be broken into two terms, the first of which compares the given code's performance to the  $m$ th-order OPTA and the second of which compares the  $m$ th-order OPTA to the  $n$ th-order OPTA. The focus of this work is on the first of those terms, called the “ $(n, m)$ th-order redundancy” and defined as

$$\begin{aligned} \Delta_{n,m}(\alpha^L, \beta^L, P_\theta, Q_o^{L,n}) &= \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q_o^{L,n}) - \hat{J}_m(\alpha^L, \beta^L, P_\theta). \end{aligned}$$

Thus,

$$\begin{aligned} \Delta_{n,m}(\alpha^L, \beta^L, P_\theta, Q_o^{L,n}) &= \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell \left( \theta, \tilde{Q}_o^{L,n} \right) + \Delta_m \left( \alpha^L, \beta^L, P_\theta, Q_{\tilde{f}_o^L(\theta)}^{L,m} \right). \end{aligned}$$

By convention, the  $(n, m)$ th-order redundancy will only be calculated for codes with second-stage coding dimension  $m$ .

The  $(n, m)$ th-order redundancy is of particular interest for practical multiresolution source coding. For practical codes, the code dimension cannot be allowed to grow without bound due the complexity implications of such unrestricted growth. When the coding dimension is fixed at some finite value  $m$ , the asymptotic performance of the  $(n, m)$ th-order redundancy as  $n$  grows without bound describes the behavior of the given code on arbitrarily large data sequences. If  $\Delta_{n,m} \rightarrow 0$  as  $n \rightarrow \infty$ , then the given code's performance using a code of dimension  $m$  approaches the  $m$ th-order OPTA for the source in operation. Thus, the code's performance is asymptotically optimal subject to the fixed constraint on the code's dimension.

Bounding the performance of two-stage multiresolution quantizers with omniscient first-stage encoders leads to bounds on the redundancy achievable in two-stage multiresolution coding. Theorems 2 and 3 use single-resolution first-stage quantizers. The coding strategy involves first describing the source parameter  $\theta$  and then describing the data using a multiresolution code matched to the described source. The parameter description is treated as part of the first-resolution description. Proof of universality involves demonstrating, for a particular set  $\Lambda$  of possible sources, that the rate cost of describing  $\theta$  and the Lagrangian performance penalty for any inaccuracies in that description are asymptotically negligible. Theorems 2 and 3 give two such scenarios. The proofs follow similar arguments to the corresponding proofs for single-resolution codes [12, Theorems 3, 4] and are, therefore, omitted.

*Theorem 2:* Consider a space  $\Lambda$  of possible distribution parameters such that  $\Lambda \subseteq \mathbb{R}^k$ . The set  $\Lambda$  is assumed to be bounded for fixed-rate coding but may be unbounded for variable-rate coding. For each  $\theta \in \Lambda$ , suppose that there exists a multiresolution quantizer  $Q_\theta^{L,n}$  achieving the  $n$ th-order OPTA on source  $P_\theta$ . Further, assume that the cost, in  $n$ th-order redundancy, associated with coding source  $P_\theta$  with quantizer  $Q_\theta^{L,n}$  is locally quadratic in  $\theta$ . That is, for each  $\theta \in \Lambda$  and  $n \geq 1$ , there exists a neighborhood  $S_\theta^n$  of  $\theta$  and constant  $c_\theta^n$  such that

$$\Delta_n(\alpha^L, \beta^L, P_\theta, Q_\theta^{L,n}) \leq c_\theta^n \|\theta - \hat{\theta}\|^2$$

for all  $\hat{\theta} \in S_\theta^n$ . Then, for any  $n$  and any  $m$  that divides  $n$  evenly, there exists a two-stage multiresolution quantizer  $Q^{L,n}$  such that

$$\Delta_{n,m}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \leq \beta_1 \frac{k \log n}{2} + \frac{a_\theta^m}{n}$$

where  $a_\theta^m \in \mathbb{R}$  is a constant dependent on  $\theta$  and  $m$ . Further, if  $c_\theta^n$  and  $S_\theta^n$  are independent of  $n$  for each  $\theta$ , then  $a_\theta^n$  is independent of  $n$  for each  $\theta$  and, setting  $m = n$ , there exists a weakly min-max sequence of multiresolution quantizers  $\{Q^{L,n}\}_{n=1}^\infty$  such that

$$\Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \leq \beta_1 \frac{k \log n}{2} + \frac{a_\theta}{n}.$$

Theorem 2 is interesting only if there exist collections of sources  $\{P_\theta, \theta \in \Lambda\}$  satisfying the locally quadratic property assumed there. In [12], the corresponding property for single-resolution quantizers is shown to hold for any collection of random processes  $\{P_\theta: \theta \in \Lambda\}$  parameterized by the vector  $\theta = (\mu, \sigma)$  describing the sources with unknown mean  $\mu$  and standard deviation  $\sigma$ . The same result holds for multiresolution quantizers, as shown in Lemma 3 in the Appendix.

*Theorem 3:* Consider a discrete space  $\Lambda$  of possible distribution parameters. For fixed-rate coding, assume that  $\Lambda$  is finite. For variable-rate coding,  $\Lambda$  may be countably infinite. For each  $\theta \in \Lambda$ , suppose that there exists a multiresolution quantizer  $Q_\theta^{L,n}$  achieving the  $n$ th-order OPTA on source  $P_\theta$ . Then, for any  $n$  and any  $m$  that divides  $n$  evenly, there exists a two-stage multiresolution quantizer  $Q^{L,n}$  such that

$$\Delta_{n,m}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \leq \beta_1 \frac{a_\theta}{n}$$

where  $a_\theta \in \mathbb{R}$  is a constant dependent on  $\theta$ . Thus, setting  $m = n$ , there exists a weakly minimax sequence of multiresolution quantizers  $\{Q^{L,n}\}_{n=1}^\infty$  with

$$\Delta_n(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \leq \beta_1 \frac{a_\theta}{n}.$$

For  $\Lambda$  finite, the codes are strongly minimax universal as well.

The use of multiresolution first-stage quantizers complicates the story considerably. Given a two-stage multiresolution code with a multiresolution, omniscient first-stage quantizer, the first-stage quantizer's description of  $\theta$  is a multiresolution description. Thus, only the first-resolution description of  $\theta$  is used in choosing the code for the first-resolution description of  $\{X_{(i)}^m\}_{i=1}^{n/m}$ ; only the first- and second-resolution descriptions of  $\theta$  are used in choosing the code for the second-resolution description of  $\{X_{(i)}^m\}_{i=1}^{n/m}$ ; and so on.

Multiresolution first-stage quantizers are motivated by universal coding problems where the accuracy required in the description of  $\theta$  varies as a function of the second-stage coding rate. In such examples, the only hope for achieving optimal rates of convergence lies in using a multiresolution first-stage quantizer. Only by using a multiresolution first-stage quantizer can the first-stage description rate (and thus the resulting redundancy) vary as a function of the resolution.

A variety of results on multiresolution first-stage quantization follow. These results treat weighted universal codes. In weighted universal multiresolution source coding, the goal is to minimize the expected  $(n, m)$ th-order redundancy

$$\begin{aligned} E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q_\Theta^{L,n}) \\ = \int \left[ \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell(\Theta, \tilde{Q}_\Theta^{L,n}) \right. \\ \left. + \Delta_m(\alpha^L, \beta^L, P_\Theta, Q_{\tilde{f}_\Theta^L}^{L,m}) \right] dW(\Theta). \end{aligned}$$

The first component of this expression treats the expected rate spent in providing a (multiresolution) description of random variable  $\Theta$ . The other describes the expected “distortion”—actually a redundancy—achieved by this description. Thus, minimization of the expected redundancy requires optimization of

some sort of “rate–distortion” tradeoff. The following definitions help make these ideas concrete.

Define the rates of the multiresolution first-stage quantizer as

$$\tilde{r}_\ell(\tilde{Q}_\Theta^{L,n}) = \int \tilde{r}_\ell(\Theta, \tilde{Q}_\Theta^{L,n}) dW(\Theta)$$

for each  $\ell \in \{1, \dots, L\}$ , and define the corresponding (weighted) “distortion” as

$$\begin{aligned} \Delta_m(\alpha^L, \beta^L, \tilde{Q}_\Theta^{L,n}) \\ = \int \Delta_m(\alpha^L, \beta^L, P_\Theta, \tilde{g}^L(\tilde{f}^L(\Theta))) dW(\Theta) \\ = \int \Delta_m(\alpha^L, \beta^L, P_\Theta, Q_{\tilde{f}_\Theta^L}^{L,m}) dW(\Theta). \end{aligned}$$

The operational “distortion–rate” function  $\tilde{D}_m(\tilde{R}^L)$  captures the tradeoff between rate and distortion as

$$\begin{aligned} \tilde{D}_m(\tilde{R}^L) = \inf_{\tilde{Q}_\Theta^{L,n}} \left\{ \Delta_m(\alpha^L, \beta^L, \tilde{Q}_\Theta^{L,n}) : \right. \\ \left. \tilde{r}_\ell(\tilde{Q}_\Theta^{L,n}) \leq \tilde{R}_\ell \forall \ell \in \{1, \dots, L\} \right\}. \end{aligned}$$

By definition of  $\tilde{D}_m(\tilde{R}^L)$ , for any  $\tilde{R}_\ell \geq 0$  for all  $\ell \in \{1, \dots, L\}$ , any  $\epsilon > 0$ , and any  $n$ , there exists an omniscient first-stage quantizer  $\tilde{Q}_\Theta^{L,n}$  such that

$$\tilde{r}_\ell(\tilde{Q}_\Theta^{L,n}) \leq \tilde{R}_\ell$$

for all  $\ell \in \{1, \dots, L\}$  and

$$\Delta_m(\alpha^L, \beta^L, \tilde{Q}_\Theta^{L,n}) \leq \tilde{D}_m(\tilde{R}^L) + \epsilon/n^2.$$

The expected redundancy of the two-stage multiresolution code  $Q_\Theta^{L,n}$  with the multiresolution first-stage quantizer  $\tilde{Q}_\Theta^{L,n}$  is

$$\begin{aligned} E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q_\Theta^{L,n}) \\ = \Delta_m(\alpha^L, \beta^L, \tilde{Q}_\Theta^{L,n}) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell(\tilde{Q}_\Theta^{L,n}) \\ \leq \tilde{D}_m(\tilde{R}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{R}_\ell + \frac{\epsilon}{n^2}. \end{aligned}$$

If, for each  $m$ ,  $\tilde{D}_m(\tilde{R}^L) \rightarrow 0$  as  $\tilde{R}_\ell \rightarrow \infty$  for all  $\ell \in \{1, \dots, L\}$ , then the expected  $n$ th-order redundancy  $E\Delta_n(\alpha^L, \beta^L, P_\Theta, Q_\Theta^{L,n})$  can be made arbitrarily small by appropriate choice of  $m$ ,  $\tilde{R}^L$ ,  $n$ , and  $\epsilon$ . In particular, choosing  $m$  sufficiently large makes the difference between  $E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q_\Theta^{L,n})$  and  $E\Delta_n(\alpha^L, \beta^L, P_\Theta, Q_\Theta^{L,n})$  small; then choosing  $\tilde{R}^L$  sufficiently large makes  $\tilde{D}_m(\tilde{R}^L)$  small; next choosing  $n$  sufficiently large makes  $(1/n) \sum_{\ell=1}^L \beta_\ell \tilde{R}_\ell$  small; and, finally, choosing  $\epsilon$  small gives the desired result.

The above discussion demonstrates the existence of weighted universal codes for sources where, for each  $m$ ,  $\tilde{D}_m(\tilde{R}^L)$  decays to zero as  $\tilde{R}^L$  grows without bound. More careful examination yields insight into the rate of convergence of  $E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q_\Theta^{L,n})$  for these universal codes. Finding the optimal rate of convergence for  $E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q_\Theta^{L,n})$  involves finding the value of

$\tilde{R}^L$  at which  $\tilde{D}_m(\tilde{R}^L) + \sum_{\ell=1}^L (\beta_\ell/n)\tilde{R}_\ell$  is minimized. Note that  $\tilde{D}_m(\tilde{R}^L) + \sum_{\ell=1}^L (\beta_\ell/n)\tilde{R}_\ell$  may be interpreted as a Lagrangian for the minimization of  $\tilde{D}_m(\tilde{R}^L)$  subject to a collection of constraints on  $\tilde{R}_\ell$ ,  $\ell \in \{1, \dots, L\}$ . Note further that as  $n$  grows, the Lagrangian constants  $\beta_1/n, \dots, \beta_L/n$  decay to zero, thereby loosening the constraints on  $\tilde{R}^L$ .

If  $\tilde{D}_m(\tilde{R}^L)$  is differentiable with respect to  $\tilde{R}_\ell$  for each  $\ell \in \{1, \dots, L\}$ , then the optimal  $\tilde{R}^L$  for any  $n$  can be found by taking partial derivatives of  $\tilde{D}_m(\tilde{R}^L) + \sum_{\ell=1}^L (\beta_\ell/n)\tilde{R}_\ell$  with respect to  $\tilde{R}_\ell$  for each  $\ell \in \{1, \dots, L\}$  and setting those partial derivatives equal to zero. In the case where  $\tilde{D}_m(\tilde{R}^L)$  is not differentiable, a convex, differentiable, decreasing upper bound on  $\tilde{D}_m(\tilde{R}^L)$  can be used as a replacement for  $\tilde{D}_m(\tilde{R}^L)$  in the minimization. The optimal performance with respect to this upper bound provides an upper bound to the optimal performance with respect to  $\tilde{D}_m(\tilde{R}^L)$ .

*Theorem 4:* Let  $\tilde{D}_m(\tilde{R}^L)$  be a convex, differentiable upper bound on the ‘‘distortion-rate’’ function  $\tilde{D}_m(\tilde{R}^L)$  for first-stage omniscient multiresolution quantizers  $\tilde{Q}_o^{L,n}: \Lambda \rightarrow \mathcal{Q}(L, m)$ . Further, assume that, for each  $m$ ,  $\tilde{D}_m(\tilde{R}^L) \rightarrow 0$  as  $\tilde{R}_\ell \rightarrow \infty$  for each  $\ell \in \{1, \dots, L\}$ . Then, for each  $n$  and each  $m$  dividing  $n$ , for any  $\epsilon > 0$  there exists a two-stage multiresolution source code  $Q^{L,n}$  with

$$\begin{aligned} E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \\ \leq \min_{\tilde{R}^L} \left[ \tilde{D}_m(\tilde{R}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{R}_\ell + \frac{\epsilon}{n^2} \right] \\ = \tilde{D}_m(\tilde{R}_{*,n}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{R}_{*,n,\ell} + \frac{\epsilon}{n^2} \end{aligned}$$

where  $\tilde{R}_{*,n}^L = (\tilde{R}_{*,n,1}, \dots, \tilde{R}_{*,n,L})$  satisfies

$$\left. \frac{\partial \tilde{D}_m(\tilde{R}^L)}{\partial \tilde{R}_\ell} \right|_{\tilde{R}_{*,n}^L} + \frac{\beta_\ell}{n} = 0 \quad (17)$$

for each  $\ell \in \{1, \dots, L\}$ . Further, if  $\tilde{D}_m(\tilde{R}^L)$  is independent of  $m$ , so that  $\tilde{D}_m(\tilde{R}^L) = \check{D}(\tilde{R}^L)$  for all  $m$ , then (setting  $m = n$ ) there exists a weighted universal sequence of multiresolution source codes  $Q^{L,n}$  such that

$$\begin{aligned} E\Delta_n(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \\ \leq \check{D}(\tilde{R}_{*,n}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{R}_{*,n,\ell} + \frac{\epsilon}{n^2}. \end{aligned}$$

*Proof:* By definition of  $\tilde{D}_m(\tilde{R}^L)$  and the given properties of  $\tilde{D}_m(\tilde{R}^L)$ , for any  $\tilde{R}^L \geq \mathbf{0}$  and any  $\epsilon > 0$ , there exists an omniscient first-stage quantizer  $\tilde{Q}_o^{L,n}$  such that  $\tilde{r}_\ell(\tilde{Q}_o^{L,n}) \leq \tilde{R}_\ell$  for all  $\ell \in \{1, \dots, L\}$  and

$$\Delta_m(\alpha^L, \beta^L, \tilde{Q}_o^{L,n}) \leq \tilde{D}_m(\tilde{R}^L) + \epsilon/n^2 \leq \check{D}_m(\tilde{R}^L) + \epsilon/n^2.$$

Thus, there exists a two-stage multiresolution code  $Q_o^{L,n}$  with an omniscient first-stage quantizer  $\tilde{Q}_o^{L,n}$  such that

$$\begin{aligned} E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q_o^{L,n}) \\ \leq \check{D}_m(\tilde{R}_{*,n}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{R}_{*,n,\ell} + \frac{\epsilon}{n^2}. \end{aligned}$$

If  $\check{D}_m(\tilde{R}^L) = \check{D}(\tilde{R}^L)$  for all  $m$ , then, setting  $m = n$

$$\begin{aligned} E\Delta_n(\alpha^L, \beta^L, P_\Theta, Q_o^{L,n}) \\ \leq \check{D}(\tilde{R}_{*,n}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{R}_{*,n,\ell} + \frac{\epsilon}{n^2}. \end{aligned}$$

Since  $\check{D}(\tilde{R}_{*,n}^L)$  approaches 0 as  $\tilde{R}_\ell \rightarrow \infty$  for each  $\ell \in \{1, \dots, L\}$ ,  $\check{D}(\tilde{R}^L) + \sum_{\ell=1}^L (\beta_\ell/n)\tilde{R}_\ell$  can be made arbitrarily small by first choosing the components of  $\tilde{R}^L$  to be large enough such that  $\check{D}(\tilde{R}^L)$  is small, and then choosing  $n$  large enough so that  $\sum_{\ell=1}^L (\beta_\ell/n)\tilde{R}_\ell$  is small. Thus, since  $\tilde{R}_{*,n}^L$  is the  $\tilde{R}^L$ -vector that minimizes  $\check{D}(\tilde{R}^L) + \sum_{\ell=1}^L (\beta_\ell/n)\tilde{R}_\ell$  for each  $n$ ,  $\check{D}(\tilde{R}_{*,n}^L) + \sum_{\ell=1}^L (\beta_\ell/n)\tilde{R}_{*,n,\ell}$  must approach zero as  $n$  grows without bound. Hence there exists a sequence of two-stage multiresolution source codes  $\{Q^{L,n}\}_{n=1}^\infty$  with  $E\Delta_n(\alpha^L, \beta^L, P_\Theta, Q_o^{L,n})$  approaching zero at the given rate.  $\square$

The following corollary demonstrates the implications of Theorem 4 under a variety of conditions on  $\tilde{D}_m(\tilde{R}^L)$ .

*Corollary 1:* If

$$\check{D}_m(\tilde{R}^L) = \sum_{\ell=1}^L A_\ell 2^{-b_\ell \tilde{R}_\ell}$$

and  $\check{D}_m(\tilde{R}^L) \rightarrow 0$  as  $\tilde{R}_\ell \rightarrow \infty$ , then for any  $\epsilon > 0$ , there exists a sequence  $\{Q^{L,n}\}_{n=1}^\infty$  of multiresolution source codes such that

$$\begin{aligned} E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \\ \leq \sum_{\ell=1}^L \left[ \left( \frac{A_\ell}{c_\ell} \right) \frac{1}{n} + \beta_\ell \left( \frac{1}{b_\ell} \right) \frac{\log(c_\ell n)}{n} \right] + \frac{\epsilon}{n^2} \end{aligned}$$

where  $c_\ell = A_\ell b_\ell / (\beta_\ell \log e)$ . In particular, we have the following conditions.

- When  $\Lambda$  has finite dimension  $k$

$$\check{D}_m(\tilde{R}^L) = A 2^{-2\tilde{R}_1/k}$$

(see Lemma 4 in the Appendix). Since  $\check{D}_m(\tilde{R}^L)$  is independent of  $m$ , setting  $m = n$  and allowing  $n$  to grow without bound gives a weighted universal sequence of multiresolution source codes with performance that converges to the  $n$ th-order OPTA as

$$\begin{aligned} E\Delta_n(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \\ \leq \frac{A}{c} \frac{1}{n} + \beta_1 \frac{k}{2} \frac{\log(c_1 n)}{n} + \frac{\epsilon}{n^2} \\ = \beta_1 \frac{k}{2} \frac{\log n}{n} + O\left(\frac{k_\ell \log k_\ell}{n}\right). \end{aligned}$$

- For fixed-rate- $R^L$  coding

$$\check{D}_m(\tilde{R}^L) = \sum_{\ell=1}^L A_\ell 2^{-2\tilde{R}_\ell/k_\ell}$$

where  $k_\ell = m 2^m \sum_{j=1}^\ell R_j$  (see Lemmas 5 and 6 in the Appendix). Thus, for any  $m$ , there exists a sequence of

multiresolution source codes with performance that converges to the  $m$ th-order OPTA as

$$\begin{aligned} E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) &\leq \sum_{\ell=1}^L \left(\frac{A_\ell}{c_\ell}\right) \frac{1}{n} + \sum_{\ell=1}^L \beta_\ell \frac{k_\ell}{2} \frac{\log(c_\ell n)}{n} + \frac{\epsilon}{n^2} \\ &= \sum_{\ell=1}^L \beta_\ell \frac{k_\ell}{2} \frac{\log n}{n} + O\left(\frac{k_\ell \log k_\ell}{n}\right). \end{aligned}$$

Allowing  $m$  to grow with  $n$  at the optimal rate gives a weighted universal code with  $O(\sqrt{\log \log n / \log n})$  convergence.

*Proof:* Given  $\check{D}_m(\tilde{R}^L) = \sum_{\ell=1}^L A_\ell 2^{-b_\ell \tilde{R}_\ell}$

$$\frac{\partial \check{D}_m(\tilde{R}^L)}{\partial \tilde{R}_\ell} = -\left(\frac{A_\ell b_\ell}{\log e}\right) 2^{-b_\ell \tilde{R}_\ell} = -\beta_\ell c_\ell 2^{-b_\ell \tilde{R}_\ell}.$$

Hence  $\tilde{R}_{*,n}^L$  may be found as the solution to the  $L$  equations described in (17), giving

$$\tilde{R}_{*,n,\ell} = \frac{1}{b_\ell} \log(c_\ell n)$$

and

$$\check{D}_m(\tilde{R}_{*,n}^L) = \sum_{\ell=1}^L \left(\frac{A_\ell}{c_\ell}\right) \frac{1}{n}.$$

Combining these results as

$$\check{D}(\tilde{R}_{*,n}^L) + \sum_{\ell=1}^L (\beta_\ell/n) \tilde{R}_{*,n,\ell} + \epsilon/n^2$$

gives the desired result.  $\square$

*Corollary 2:* If

$$\check{D}_m(\tilde{R}^L) = 0$$

when  $\tilde{R}_\ell \geq \tilde{r}_\ell$  for each  $\ell \in \{1, \dots, L\}$ , then for any  $\epsilon > 0$ , there exists a sequence  $\{Q^{L,n}\}_{n=1}^\infty$  of multiresolution source codes such that

$$E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \leq \sum_{\ell=1}^L \beta_\ell \frac{\tilde{r}_\ell}{n} + \frac{\epsilon}{n^2}.$$

In particular, we have the following condition.

- If  $\Lambda$  is countable and  $H(\Theta)$  is finite, then

$$\check{D}_m(\tilde{R}^L) = 0$$

for all  $\tilde{R}^L \geq (H(\Theta), 0, \dots, 0)$  since  $H(\Theta)$  bits in the first resolution of the first-stage source description suffices to describe  $\Theta$  perfectly. In this case, there exists a sequence  $\{Q^{L,n}\}_{n=1}^\infty$  of two-stage multiresolution source codes such that

$$E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \leq \beta_1 \frac{H(\Theta)}{n} + \frac{\epsilon}{n^2}.$$

*Proof:* Given  $\check{D}_m(\tilde{R}^L) = 0$  for any  $\tilde{R}_\ell \geq \tilde{r}_\ell$ , by Theorem 4, there exists a sequence  $\{Q^{L,n}\}_{n=1}^\infty$  of multiresolution source codes such that

$$\begin{aligned} E\Delta_{n,m}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) &\leq \min_{\tilde{R}^L} \left[ \check{D}_m(\tilde{R}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{R}_\ell \right] + \frac{\epsilon}{n^2} \\ &\leq \left( \check{D}_m(\tilde{r}^L) + \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell \right) + \frac{\epsilon}{n^2} \\ &= \frac{1}{n} \sum_{\ell=1}^L \beta_\ell \tilde{r}_\ell + \frac{\epsilon}{n^2} \end{aligned}$$

which gives the desired result.  $\square$

## V. SUMMARY

This work introduces the concept of universal multiresolution source coding and considers the existence, rate of convergence, and design of universal multiresolution codes. Central to the discussion is a generalization of the quantization interpretation of two-stage universal source coding from single- to multiresolution source coding. Two ways of achieving the generalization are considered. When the rate required for the code description in the first-stage description is independent of the desired coding rate, then a single-resolution first-stage quantizer is used with a multiresolution second-stage code. The single-resolution first-stage quantizer's role in quantizing the space of possible sources (or, equivalently, the space of possible multiresolution codes) is a straightforward extension of the role played by the first-stage quantizer in traditional two-stage codes. When the rate required for the code description in the first-stage description is a function of the desired coding rate, then the single-resolution first-stage quantizer is replaced with a multiresolution first-stage quantizer. In this case, in resolution  $\ell$  the first-stage quantizer may describe the source or code to the accuracy required for the resolution- $\ell$  data description. The first-stage description may then be refined in future resolutions if greater accuracy in the source description is required. Proof of existence and redundancy results are given under both scenarios.

## APPENDIX

*Theorem 1:* Assume that  $\{X_i\}$  is stationary and ergodic for each  $\theta \in \Lambda$ . Weighted universal multiresolution codes exist for  $(\{X_i\}, \Theta)$  if and only if the OPTA has an ergodic decomposition.

*Proof:* By the definition of weighted universal coding for fixed-rate multiresolution quantizers, a sequence  $\{Q^{L,n}\}_{n=1}^\infty$  of quantizers with  $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$  for all  $n \geq 1$  is a weighted universal sequence of fixed-rate multiresolution quantizers if and only if

$$E\Delta^{\text{fr}}(\alpha^L, R^L, P_\Theta, Q^{L,n}) \rightarrow 0, \quad \text{as } n \rightarrow \infty$$

where

$$\begin{aligned} E\Delta^{\text{fr}}(\alpha^L, R^L, P_\Theta, Q^{L,n}) \\ &= \int \left[ \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}) \right. \\ &\quad \left. - \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) \right] dW(\theta) \\ &= \frac{1}{n} J^{\text{op}}(\alpha^L, P, Q^{L,n}) - \int \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) dW(\theta). \end{aligned}$$

Thus, if the ergodic decomposition

$$\hat{J}^{\text{fr}}(\alpha^L, R^L, P) = \int \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) dW(\theta) \quad (18)$$

holds, then, because there exists  $\{Q^{L,n}\}_{n=1}^\infty$  such that  $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$  for all  $n \geq 1$  and

$$(1/n)J^{\text{op}}(\alpha^L, P, Q^{L,n}) \rightarrow \hat{J}^{\text{fr}}(\alpha^L, R^L, P)$$

(18) implies that  $E\Delta^{\text{fr}}(\alpha^L, R^L, P_\Theta, Q^{L,n}) \rightarrow 0$  for the same sequence  $\{Q^{L,n}\}_{n=1}^\infty$  of fixed-rate multiresolution quantizers. On the other hand, notice that for any  $n \geq 1$  and any  $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}(L, n)$

$$\frac{1}{n} J^{\text{op}}(\alpha^L, P, Q^{L,n}) \geq \hat{J}^{\text{fr}}(\alpha^L, R^L, P)$$

and

$$\begin{aligned} \hat{J}^{\text{fr}}(\alpha^L, R^L, P) \\ &= \lim_{n \rightarrow \infty} \inf_{Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}} \int \frac{1}{n} J^{\text{op}}(\alpha^L, P_x, Q^{L,n}) dP(x) \\ &\geq \lim_{n \rightarrow \infty} \int \inf_{Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}} \frac{1}{n} J^{\text{op}}(\alpha^L, P_x, Q^{L,n}) dP(x) \\ &\geq \int \lim_{n \rightarrow \infty} \inf_{Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L}} \frac{1}{n} J^{\text{op}}(\alpha^L, P_x, Q^{L,n}) dP(x) \\ &= \int \hat{J}^{\text{fr}}(\alpha^L, R^L, P_x) dP(x) \\ &= \int J(\alpha^L, R^L, P_x) dP(x) \\ &= \iint J(\alpha^L, R^L, P_x) dP_\theta(x) dW(\theta) \\ &= \int \bar{J}(\alpha^L, R^L, P_\theta) dW(\theta) \\ &= \int \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) dW(\theta) \end{aligned}$$

so  $E\Delta^{\text{fr}}(\alpha^L, R^L, P_\Theta, Q^{L,n})$  can only approach zero if (18) holds.

The argument for variable-rate coding is similar. In particular, a sequence  $\{Q^{L,n}\}_{n=1}^\infty$  of quantizers with

$$Q^{L,n} \in \mathcal{Q}^{\text{vr}}(L, n), \quad \text{for all } n \geq 1$$

is a weighted universal sequence of variable-rate multiresolution quantizers if and only if  $E\Delta^{\text{vr}}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \rightarrow 0$  as  $n \rightarrow \infty$ , where

$$\begin{aligned} E\Delta^{\text{vr}}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \\ &= \int \left[ \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q^{L,n}) \right. \\ &\quad \left. - \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) \right] dW(\theta) \\ &= \frac{1}{n} J^{\text{op}}(\alpha^L, \beta^L, P, Q^{L,n}) \\ &\quad - \int \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) dW(\theta). \end{aligned}$$

Thus, by an argument similar to the one given above,  $E\Delta^{\text{vr}}(\alpha^L, \beta^L, P_\Theta, Q^{L,n}) \rightarrow 0$  if and only if the following ergodic decomposition holds:

$$\hat{J}^{\text{vr}}(\alpha^L, \beta^L, P) = \int \hat{J}^{\text{vr}}(\alpha^L, \beta^L, P_\theta) dW(\theta). \quad (19)$$

□

*Lemma 1:* Suppose that  $\{r^L(n)\}_{n=1}^\infty$  is a sequence of  $L$ -dimensional real vectors such that  $r_\ell(n) > 0$  for all  $n \geq 1$  and  $r_\ell(n) \rightarrow 0$  as  $n \rightarrow \infty$  for each  $\ell \in \{1, \dots, L\}$ . Suppose further that for each  $\theta \in \Lambda$ ,  $\{c_n(\theta)\}_{n=1}^\infty$  is a sequence of positive numbers that converges to zero as  $n$  grows without bound. If for each rate vector  $R^L$  there exists a sequence of multiresolution quantizers  $\{Q^{L,n}\}_{n=1}^\infty$  such that for each  $n \geq 1$   $Q^{L,n} \in \mathcal{Q}^{\text{fr}, R^L + r_\ell(n)}(L, n)$  and

$$\frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, Q^{L,n}) \leq \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) + c_n(\theta)$$

for each  $\theta \in \Lambda$ , then for each positive rate vector  $\hat{R}^L$  there exists a sequence of quantizers  $\{\hat{Q}^{L,n}\}_{n=1}^\infty$  such that for each  $n \geq 1$   $\hat{Q}^{L,n} \in \mathcal{Q}^{\text{fr}, \hat{R}^L}(L, n)$  and

$$\begin{aligned} \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, \hat{Q}^{L,n}) \\ \leq \hat{J}^{\text{fr}}(\alpha^L, \hat{R}^L, P_\theta) + c_n(\theta) + \sum_{\ell=1}^L \beta_\ell r_\ell(n) \end{aligned}$$

for  $n$  sufficiently large, provided that the planar direction described by  $\beta^L$  satisfies the constraints

$$\beta_\ell > -\partial(\hat{J}^{\text{fr}}(\alpha^L, \hat{R}^L, P_\theta)) / (\partial R_\ell)$$

for all  $\ell \in \{1, \dots, L\}$ .

*Proof:* Since  $\hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta)$  is convex in  $R^L$  (by a simple time-sharing argument) [14] and given the condition on  $\beta^L$ , there exists a rate vector  $\hat{R}^L$  with  $\hat{R}_\ell < \hat{R}_\ell$  for all  $\ell \in \{1, \dots, L\}$  such that for any  $R^L$  with

$$\begin{aligned} R_\ell \in [\hat{R}_\ell, \hat{R}_\ell], \quad \ell \in \{1, \dots, L\} \\ \hat{J}^{\text{fr}}(\alpha^L, R^L, P_\theta) \leq \hat{J}^{\text{fr}}(\alpha^L, \hat{R}^L, P_\theta) + \sum_{\ell=1}^L \beta_\ell (\hat{R}_\ell - R_\ell). \end{aligned}$$

Consider  $n$  sufficiently large so that

$$R_\ell = \hat{R}_\ell - r_\ell(n) \in [\check{R}_\ell, \hat{R}_\ell], \quad \text{for all } \ell \in \{1, \dots, L\}.$$

Then there exists a

$$\hat{Q}^{L,n} \in \mathcal{Q}^{\text{fir}, R^L + r^L(n)}(L, n) = \mathcal{Q}^{\text{fir}, \hat{R}^L}(L, n)$$

such that

$$\begin{aligned} & \frac{1}{n} J^{\text{op}}(\alpha^L, P_\theta, \hat{Q}^{L,n}) \\ & \leq \hat{J}^{\text{fir}}(\alpha^L, R^L, P_\theta) + c_n(\theta) \\ & \leq \hat{J}^{\text{fir}}(\alpha^L, \hat{R}^L, P_\theta) + \sum_{\ell=1}^L \beta_\ell r_\ell(n) + c_n(\theta), \end{aligned}$$

giving the desired result.  $\square$

*Lemma 2:* For each  $x^n \in A^n$  and any  $\beta^L$  such that  $\beta_\ell \geq 0$  for all  $\ell \in \{1, \dots, L\}$  and  $0 < \sum_{\ell=1}^L \beta_\ell < \infty$ , the infimum

$$\inf_{\tilde{s}^L \in \tilde{\mathcal{S}}^L} \left[ \sum_{\ell=1}^L \beta_\ell |\tilde{s}_\ell| + \sum_{i=1}^{n/m} j^{\text{op}}(\alpha^L, \beta^L, x_{(i)}^m, Q_{\tilde{s}^L}^{L,m}) \right]$$

is achieved.

*Proof:* Pick any  $\tilde{s}^L \in \tilde{\mathcal{S}}^L$ , and let

$$c = \sum_{\ell=1}^L \beta_\ell |\tilde{s}_\ell| + \sum_{i=1}^{n/m} j^{\text{op}}(\alpha^L, \beta^L, x_{(i)}^m, Q_{\tilde{s}^L}^{L,m}) < \infty.$$

Since  $\tilde{\mathcal{S}}^L$  is a multiresolution prefix code, the number of strings  $\tilde{s}'^L \in \tilde{\mathcal{S}}^L$  that satisfy  $\sum_{\ell=1}^L \beta_\ell |\tilde{s}'_\ell| \leq c$  must be finite. Thus, since  $j^{\text{op}}(\alpha^L, \beta^L, x^m, Q_{\tilde{s}^L}^{L,m}) > 0$  for all  $x^m$  and all  $Q_{\tilde{s}^L}^{L,m}$ , the infimum is achieved.  $\square$

*Lemma 3:* Let  $\{X_i\}$  be an arbitrary real-valued random process with measure  $P_\theta$ , such that  $\theta = (\mu, \sigma)$  describes the source's mean  $\mu$  and standard deviation  $0 < \sigma < \infty$ . For any  $\theta \in \Lambda$ , let  $Q_\theta^{L,m}$  be the optimal dimension- $m$  multiresolution quantizer for source  $P_\theta$ . If the  $\ell$ -th-resolution distortion measure  $\rho_{(\ell)}(x, y) = (x - y)^2$  for all  $\ell \in \{1, \dots, L\}$ , then

$$\Delta_n(\alpha^L, \beta^L, P_\theta, Q_\theta^{L,m}) \leq \left( \sum_{\ell=1}^L \alpha_\ell \right) \|\theta - \hat{\theta}\|^2$$

for all  $n \geq 1$  and all  $\theta \in \Lambda$ .

*Proof:* For any  $(\mu, \sigma) \in \Lambda$ , the quantizer  $Q_{(\mu, \sigma)}^{L,m}$  can be obtained from the quantizer  $Q_{(0,1)}^{L,m}$  by first scaling each component of  $Q_{(0,1)}^{L,m}$  by  $\sigma$  and then translating by  $\mu$ . Thus,

$$\begin{aligned} & \Delta_m(\alpha^L, \beta^L, P_\theta, Q_\theta^{L,m}) \\ & = \frac{1}{m} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q_\theta^{L,m}) \\ & \quad - \frac{1}{m} J^{\text{op}}(\alpha^L, \beta^L, P_\theta, Q_\theta^{L,m}) \end{aligned}$$

$$\begin{aligned} & = \frac{1}{m} E_\theta \sum_{\ell=1}^L \left[ \alpha_\ell \|X^m - g_{\theta, \ell}(f_\theta^\ell(X^m))\|^2 + \beta_\ell |f_{\theta, \ell}(X^m)| \right] \\ & \quad - \frac{1}{m} E_\theta \sum_{\ell=1}^L \left[ \alpha_\ell \|X^m - g_{\theta, \ell}(f_\theta^\ell(X^m))\|^2 + \beta_\ell |f_{\theta, \ell}(X^m)| \right] \\ & \leq \frac{1}{m} E_\theta \sum_{\ell=1}^L \left[ \alpha_\ell \|X^m - g_{\theta, \ell}(f_\theta^\ell(X^m))\|^2 + \beta_\ell |f_{\theta, \ell}(X^m)| \right] \\ & \quad - \frac{1}{m} E_\theta \sum_{\ell=1}^L \left[ \alpha_\ell \|X^m - g_{\theta, \ell}(f_\theta^\ell(X^m))\|^2 + \beta_\ell |f_{\theta, \ell}(X^m)| \right] \\ & = \frac{1}{m} E_\theta \sum_{\ell=1}^L \alpha_\ell \|g_{\theta, \ell}(f_\theta^\ell(X^m)) - g_{\hat{\theta}, \ell}(f_\theta^\ell(X^m))\|^2 \\ & = \frac{1}{m} E_\theta \sum_{\ell=1}^L \alpha_\ell \sum_{j=1}^m \left[ g_{\theta, \ell, j}(f_\theta^\ell(X^m)) \right. \\ & \quad \left. - \left( \frac{\hat{\sigma}}{\sigma} (g_{\theta, \ell, j}(f_\theta^\ell(X^m)) - \mu) + \hat{\mu} \right) \right]^2 \\ & = \frac{1}{m} E_\theta \sum_{\ell=1}^L \alpha_\ell \sum_{j=1}^m \left[ \left( 1 - \frac{\hat{\sigma}}{\sigma} \right) g_{\theta, \ell, j}(f_\theta^\ell(X^m)) \right. \\ & \quad \left. + \left( \frac{\mu \hat{\sigma}}{\sigma} - \hat{\mu} \right) \right]^2 \\ & = \frac{1}{m} E_\theta \sum_{\ell=1}^L \alpha_\ell \sum_{j=1}^m \\ & \quad \cdot \left[ \left( 1 - \frac{\hat{\sigma}}{\sigma} \right)^2 g_{\theta, \ell, j}^2(f_\theta^\ell(X^m)) \right. \\ & \quad \left. + 2 \left( 1 - \frac{\hat{\sigma}}{\sigma} \right) \left( \frac{\mu \hat{\sigma}}{\sigma} - \hat{\mu} \right) g_{\theta, \ell, j}(f_\theta^\ell(X^m)) \right. \\ & \quad \left. + \left( \frac{\mu \hat{\sigma}}{\sigma} - \hat{\mu} \right)^2 \right] \\ & \leq \left( \sum_{\ell=1}^L \alpha_\ell \right) \left[ \left( 1 - \frac{\hat{\sigma}}{\sigma} \right)^2 E_\theta X_1^2 \right. \\ & \quad \left. + 2 \left( 1 - \frac{\hat{\sigma}}{\sigma} \right) \left( \frac{\mu \hat{\sigma}}{\sigma} - \hat{\mu} \right) E_\theta X_1 + \left( \frac{\mu \hat{\sigma}}{\sigma} - \hat{\mu} \right)^2 \right] \\ & = \left( \sum_{\ell=1}^L \alpha_\ell \right) \left[ \left( 1 - \frac{\hat{\sigma}}{\sigma} \right)^2 (\sigma^2 + \mu^2) \right. \\ & \quad \left. + 2 \left( 1 - \frac{\hat{\sigma}}{\sigma} \right) \left( \frac{\mu \hat{\sigma}}{\sigma} - \hat{\mu} \right) \mu + \left( \frac{\mu \hat{\sigma}}{\sigma} - \hat{\mu} \right)^2 \right] \\ & = \left( \sum_{\ell=1}^L \alpha_\ell \right) [(\sigma - \hat{\sigma})^2 + (\mu - \hat{\mu})^2]. \quad \square \end{aligned}$$

*Lemma 4:* Consider a space  $\Lambda$  of possible source parameters such that  $\Lambda \subseteq \mathbb{R}^k$  for some fixed integer  $k$ . In the case of

fixed-rate coding, assume that  $\Lambda$  is bounded. In the case of variable-rate coding,  $\Lambda$  may be unbounded if the differential entropy

$$h(\Theta) = \int w(\theta) \log w(\theta) d\theta < \infty.$$

Suppose that for each  $\theta \in \Lambda$ , there exists a code  $Q_{\theta}^{L,m}$  that achieves the  $m$ th-order OPTA for source  $P_{\theta}$ . Suppose further that for all  $\theta, \hat{\theta} \in \Lambda$  and each  $m \geq 1$ , there exists a constant  $c_m$  such that

$$\Delta_m(\alpha^L, \beta^L, P_{\theta}, Q_{\theta}^{L,m}) \leq c_m \|\theta - \hat{\theta}\|^2.$$

Then there is a constant  $A_m$  dependent on  $m$  such that

$$\tilde{D}_m(\tilde{R}^L) \leq \check{D}_m(\tilde{R}^L) = A_m 2^{-2\tilde{R}_1/k}.$$

Further, if there exists a constant  $c$  such that  $c_m \leq c$  for all  $m$ , then  $A_m = A$  does not depend on  $m$ .

*Proof:* This result uses a single-resolution first-stage quantizer. Designing that first-stage quantizer involves breaking  $\mathbb{R}^k$  into nested hypercubes—each with a representative value—and using the description of the hypercube in which  $\theta$  lies as an approximate description of  $\theta$ . Since the first-stage code is a single-resolution code, the argument is similar to the corresponding argument for weighted universal single-resolution codes [12, Lemma 7].  $\square$

*Lemma 5:* Fix integer  $m$  and rate vector  $R^L$  such that  $2^m \sum_{j=1}^{\ell} R_j$  is an integer for each  $\ell \in \{1, \dots, L\}$ , and suppose that  $\|X^m\| \leq B$  almost surely- $P^m$ . Then, for fixed-rate- $R^L$  coding under the squared-error distortion criterion

$$\tilde{D}_m(\tilde{R}^L) \leq A \sum_{\ell=1}^L \alpha_{\ell} 2^{-2\tilde{R}_{\ell}/k_{\ell}}$$

for some  $A < \infty$ , where  $k_{\ell} = m 2^m \sum_{j=1}^{\ell} R_j$  equals the number of parameters in the  $\ell$ th resolution of a fixed-rate- $R^L$  code.

*Proof:* Let  $Q_{\theta}^{L,m} = g_{\theta}^L \circ f_{\theta}^L$  be the optimal fixed-rate- $R^L$  multiresolution quantizer for  $P_{\theta}$ . Since the optimal encoder  $f_{\theta}^L$  must be a nearest neighbor encoder, describing for each  $\ell \in \{1, \dots, L\}$  the  $2^m \sum_{j=1}^{\ell} R_j$   $m$ -dimensional, resolution- $\ell$  codewords suffices to completely describe  $Q_{\theta}^{L,m}$ . Further, since the codewords of  $Q_{\theta}^{L,m}$  are the only information about  $\theta$  that is required by the decoder, describing  $Q_{\theta}^{L,m}$  to the decoder is equivalent to describing  $\theta$  to the decoder. Thus, there is no loss of generality associated with treating  $\theta$  as the vector of dimension

$$\sum_{\ell=1}^L m 2^m \sum_{j=1}^{\ell} R_j = \sum_{\ell=1}^L k_{\ell}$$

that describes the full multiresolution codebook. Since  $\|X^m\| \leq B$  almost surely, under this treatment

$$\begin{aligned} \theta &\in [-B, B]^{k_1} \times [-B, B]^{k_2} \times \dots \times [-B, B]^{k_L} \\ &= [-B, B]^{\sum_{\ell=1}^L k_{\ell}} \end{aligned}$$

almost surely.

For each  $\ell \in \{1, \dots, L\}$ , partition  $[-B, B]^{k_{\ell}}$  uniformly into  $2^{\tilde{R}_{\ell}}$  bins by partitioning each dimension into  $2^{\tilde{R}_{\ell}/k_{\ell}}$  bins and then taking the cross product. Taking the cross product of these partitions gives a  $(2^{\sum_{\ell=1}^L \tilde{R}_{\ell}})$ -bin partition of  $[-B, B]^{\sum_{\ell=1}^L k_{\ell}}$ . Now consider an omniscient two-stage multiresolution quantizer  $Q_o^{L,n}$  that describes  $\theta$  by describing the bin into which  $\theta$  falls. (If  $\theta \notin [-B, B]^{\sum_{\ell=1}^L k_{\ell}}$ , then the first-stage quantizer encoder describes the bin whose center is closest to  $\theta$ .) Let the second-stage multiresolution quantizer be  $Q_{\hat{\theta}}^{L,m} = g_{\hat{\theta}}^L \circ f_{\hat{\theta}}^L$ , where  $\hat{\theta}$  is a representative value at the center of the chosen bin,  $g_{\hat{\theta}}^L$  is the multiresolution quantizer decoder described by  $\hat{\theta}$ , and  $f_{\hat{\theta}}^L$  is the corresponding nearest neighbor encoder. Notice that decoding the  $\ell$ th-resolution description of data vector  $X^m$  requires only the  $\ell$ th-resolution codewords. Thus, an  $L$ -resolution description  $\tilde{s}^L$  of  $\hat{\theta}$  suffices, where for each  $\ell \in \{1, \dots, L\}$ ,  $\tilde{s}_{\ell}$  is a rate- $\tilde{R}_{\ell}$  description of the resolution- $\ell$  codewords of  $Q_{\hat{\theta}}^{L,m}$ . Notice that

$$\|g_{\hat{\theta}, \ell}(f_{\hat{\theta}}^{\ell}(X^m)) - g_{\theta, \ell}(f_{\theta}^{\ell}(X^m))\|^2 \leq m \left[ (2B) 2^{-\tilde{R}_{\ell}/k_{\ell}} \right]^2$$

for almost all  $\theta$ . Thus,

$$\begin{aligned} \Delta_{n,m}(\alpha^L, \beta^L, \check{Q}_o^{L,n}) &= \frac{1}{m} EE_{\Theta} \left[ \sum_{\ell=1}^L \alpha_{\ell} \|X^m - g_{\tilde{s}_{\ell}, \ell}(f_{\tilde{s}_{\ell}}^{\ell}(X^m))\|^2 \right] \\ &\quad - \frac{1}{m} EE_{\Theta} \left[ \sum_{\ell=1}^L \alpha_{\ell} \|X^m - g_{\Theta, \ell}(f_{\Theta}^{\ell}(X^m))\|^2 \right] \\ &\leq \frac{1}{m} EE_{\Theta} \sum_{\ell=1}^L \alpha_{\ell} \left[ \|X^m - g_{\tilde{s}_{\ell}, \ell}(f_{\tilde{s}_{\ell}}^{\ell}(X^m))\|^2 \right. \\ &\quad \left. - \|X^m - g_{\Theta, \ell}(f_{\Theta}^{\ell}(X^m))\|^2 \right] \\ &= \frac{1}{m} EE_{\Theta} \sum_{\ell=1}^L \alpha_{\ell} \left[ \|X^m - g_{\hat{\theta}, \ell}(f_{\hat{\theta}}^{\ell}(X^m))\|^2 \right. \\ &\quad \left. - \|X^m - g_{\Theta, \ell}(f_{\Theta}^{\ell}(X^m))\|^2 \right] \\ &= \frac{1}{m} EE_{\Theta} \sum_{\ell=1}^L \alpha_{\ell} \left[ \|g_{\hat{\theta}, \ell}(f_{\hat{\theta}}^{\ell}(X^m)) - g_{\Theta, \ell}(f_{\Theta}^{\ell}(X^m))\|^2 \right] \\ &\leq \sum_{\ell=1}^L \alpha_{\ell} 4B^2 2^{-2\tilde{R}_{\ell}/k_{\ell}} \end{aligned}$$

giving the desired result.  $\square$

*Lemma 6:* Choose an integer  $m$  and rate vector  $R^L$  such that  $2^m \sum_{j=1}^{\ell} R_j$  is an integer for each  $\ell \in \{1, \dots, L\}$ . Suppose that  $E\|X^m\|^{2+\epsilon} < \infty$  for some  $\epsilon > 0$ . Then, for fixed-rate- $R^L$  coding under the squared-error distortion criterion

$$\tilde{D}_m(\tilde{R}^L) \leq A \sum_{\ell=1}^L \alpha_{\ell} 2^{-2\tilde{R}_{\ell}/k_{\ell}}$$

for some  $A < \infty$ , where  $k_{\ell} = m 2^m \sum_{j=1}^{\ell} R_j$ .

*Proof:* Given that  $E\|X^m\|^{2+\epsilon} < \infty$ , by [12, Lemma 9]

$$E \left[ \|g_{\Theta, \ell}(s^{\ell})\|^{2+\epsilon} \mid f_{\Theta}^{\ell}(X^m) = s^{\ell} \right] < \infty$$

for each  $\ell \in \{1, \dots, L\}$  and each of the  $2^m \sum_{j=1}^{\ell} R_j$  fixed-length- $(m \sum_{j=1}^{\ell} R_j)$  binary strings  $s^\ell$ . Thus, by [26, Theorem 2], for each such  $s^\ell$  there exists a constant  $A_{s^\ell} < \infty$  and a collection  $\Gamma_{s^\ell}$  containing  $\lfloor 2^{m\tilde{R}_\ell/k_\ell} \rfloor$   $m$ -vectors such that

$$\frac{1}{m} E \left[ \min_{\mathbf{g} \in \Gamma_{s^\ell}} \|\mathbf{g} - g_{\Theta}(s^\ell)\|^2 \mid f_{\Theta}^\ell(X^m) = s^\ell \right] \leq A_{s^\ell} 2^{-2\tilde{R}_\ell/k_\ell},$$

Let  $\Gamma_\ell$  be the collection of

$$\left\lfloor 2^{m\tilde{R}_\ell/k_\ell} \right\rfloor 2^m \sum_{j=1}^{\ell} R_j \leq 2(m 2^m \sum_{j=1}^{\ell} R_j) \tilde{R}_\ell/k_\ell = 2^{\tilde{R}_\ell}$$

fixed-rate- $(\sum_{j=1}^{\ell} R_j)$  codes achieved by taking the cross product of the  $\Gamma_{s^\ell}$  sets for a fixed value of  $\ell$ . Finally, let  $\Gamma$  be the collection of

$$\prod_{\ell=1}^L \left\lfloor 2^{m\tilde{R}_\ell/k_\ell} \right\rfloor 2^m \sum_{j=1}^{\ell} R_j \leq \prod_{\ell=1}^L 2^{\tilde{R}_\ell} = 2^{\sum_{\ell=1}^L \tilde{R}_\ell}$$

$L$ -resolution source codes achieved by using a codebook from  $\Gamma_\ell$  in resolution  $\ell$  for each  $\ell \in \{1, \dots, L\}$ .

By the above code construction, for any  $\theta \in \Lambda$ , there exists an  $L$ -resolution source code  $Q_{\tilde{s}^L}^{L,m} = g_{\tilde{s}^L}^L \circ f_{\tilde{s}^L}^L$  such that for each  $\ell \in \{1, \dots, L\}$  and each fixed-rate- $(\sum_{j=1}^{\ell} R_j)$  binary description  $s^\ell$

$$g_{\tilde{s}^\ell}(s^\ell) = \arg \min_{\mathbf{g} \in \Gamma_{s^\ell}} \|\mathbf{g} - g_{\Theta, \ell}(s^\ell)\|^2.$$

Let  $\tilde{Q}_o^{L,n} = \tilde{g}^L \circ \tilde{f}_o^L$  be the first-stage multiresolution quantizer such that  $\tilde{f}_o^\ell(\theta) = \tilde{s}^\ell$  and  $\tilde{g}^L(\tilde{s}^L) = Q_{\tilde{s}^L}^{L,m}$ . Then

$$\begin{aligned} \Delta_{n,m}(\alpha^L, \beta^L, \tilde{Q}_o^{L,n}) &\leq \frac{1}{m} E \sum_{\ell=1}^L \alpha_\ell \left[ \|g_{\tilde{s}^\ell, \ell}(f_{\Theta}^\ell(X^m)) - g_{\Theta, \ell}(f_{\Theta}^\ell(X^m))\|^2 \right] \\ &\leq \sum_{\ell=1}^L \alpha_\ell \sum_{s^\ell} P[f_{\Theta}^\ell(X^m) = s^\ell] \\ &\cdot \frac{1}{m} E \left[ \|g_{\tilde{s}^\ell, \ell}(s^\ell) - g_{\Theta, \ell}(s^\ell)\|^2 \mid f_{\Theta}^\ell(X^m) = s^\ell \right] \\ &\leq \sum_{\ell=1}^L \alpha_\ell \left( \sum_{s^\ell} P[f_{\Theta}^\ell(X^m) = s^\ell] A_{s^\ell} \right) 2^{-2\tilde{R}_\ell/k_\ell} \end{aligned}$$

giving the desired result.  $\square$

## REFERENCES

[1] M. Effros, "Practical multiresolution source coding: TSVQ revisited," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1998, pp. 53–62.

[2] M. Effros and D. Dugatkin, "Multiresolution vector quantization," *IEEE Trans. Inform. Theory*, submitted for publication.

[3] M. Effros, "Universal multiresolution source codes," in *Proc. Information Theory Workshop*, Santa Fe, NM, Feb. 1999, Invited Paper, p. 71.

[4] J. Ziv, "Coding of sources with unknown statistics—Part II: Distortion relative to a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 389–394, May 1972.

[5] D. L. Neuhoff, R. M. Gray, and L. D. Davisson, "Fixed rate universal block source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 511–523, Sept. 1975.

[6] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[7] J. C. Kieffer, "On the minimum rate for strong universal block coding of a class of ergodic sources," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 693–702, Nov. 1980.

[8] J. Ziv, "On universal quantization," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 344–347, May 1985.

[9] T. Linder, G. Lugosi, and K. Zeger, "Rates of convergence in the source coding theorem, in empirical quantizer design, and in universal lossy source coding," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1728–1740, Nov. 1994.

[10] —, "Fixed rate universal lossy source coding and rates of convergence for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 41, pp. 665–676, May 1995.

[11] Z. Zhang and V. K. Wei, "An on-line universal lossy data compression algorithm via continuous codebook refinement," *IEEE Trans. Inform. Theory*, submitted for publication.

[12] P. A. Chou, M. Effros, and R. M. Gray, "A vector quantization approach to universal noiseless coding and quantization," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1109–1138, July 1996.

[13] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.

[14] M. Effros, "Distortion-rate bounds for fixed- and variable-rate multiresolution source codes," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1887–1910, Sept. 1999.

[15] W. H. R. Equitz, "Successive refinement of information," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1989.

[16] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[17] B. Rimoldi, "Successive refinement of information: Characterization of achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.

[18] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[19] M. Effros, P. A. Chou, and R. M. Gray, "Variable-rate source coding theorems for stationary nonergodic sources," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1920–1925, Nov. 1994.

[20] R. J. Pile, "Coding theorems for discrete source-channel pairs," Ph.D. dissertation, Mass. Inst. Technol. (MIT), Cambridge, 1967.

[21] —, "The transmission distortion of a source as a function of the encoding block length," *Bell Syst. Tech. J.*, vol. 47, pp. 827–885, 1968.

[22] Z. Zhang, E.-h. Yang, and V. K. Wei, "The redundancy of source coding with a fidelity criterion. 1. Known statistics," *IEEE Trans. Inform. Theory*, vol. 43, pp. 71–91, Jan. 1997.

[23] P. A. Chou, "Code clustering for weighted universal VQ and other applications," in *Proc. IEEE Int. Symp. Information Theory*, Budapest, Hungary, June 1991, p. 253.

[24] K. Zeger and A. Bist, "Universal adaptive vector quantization using codebook quantization," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, San Francisco, CA, Mar. 1992, pp. III.381–384.

[25] K. Zeger, A. Bist, and T. Linder, "Universal source coding with codebook transmission," *IEEE Trans. Commun.*, vol. 42, pp. 336–346, Feb. 1994.

[26] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with  $r$ th power distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 239–347, Mar. 1982.