

On the Cost of Finite Block Length in Quantizing Unbounded Memoryless Sources

Tamás Linder and Kenneth Zeger, *Senior Member, IEEE*

Abstract—The problem of fixed-rate block quantization of an unbounded real memoryless source is studied. It is proved that if the source has a finite sixth moment, then there exists a sequence of quantizers Q_n of increasing dimension n and fixed rate R such that the mean squared distortion $\Delta(Q_n)$ is bounded as $\Delta(Q_n) \leq D(R) + O(\sqrt{\log n/n})$, where $D(R)$ is the distortion-rate function of the source. Applications of this result include the evaluation of the distortion redundancy of fixed-rate universal quantizers, and the generalization to the non-Gaussian case of a result of Wyner on the transmission of a quantized Gaussian source over a memoryless channel.

I. INTRODUCTION

SHANNON'S source coding theorem with a fidelity criterion [5] showed that by increasing the blocklength n of a lossy source code, it is possible to have the mean squared error approach the distortion-rate lower bound arbitrarily closely. Pilc [12] showed that for finite alphabet sources the convergence of the expected distortion to the distortion-rate function occurs at a rate $O(\log n/n)$. Wyner [14] proved the same convergence rate for memoryless Gaussian sources and squared distortion. Zhang *et al* [18] in a recent paper settled the question of the convergence rate for finite alphabet memoryless sources by showing that the exact asymptotic rate is $c \log n/n$, where they also identified the constant c .

For nondiscrete sources it has recently been shown [10] that for a bounded real memoryless source and squared distortion the convergence occurs at a rate $O(\sqrt{\log n/n})$. This result was used in [10] to analyze the performance of a certain universal quantization scheme. On the other hand, the assumption of bounded support is sometimes a severe restriction in signal quantization, especially since some of the most popular source models have unbounded support, such as the Laplacian. The convergence rate results mentioned above also assume that binary information is transmitted across a lossless channel. In the present paper we eliminate the bounded support require-

ment and also consider transmission across a noisy channel. In addition, we are able to obtain a rate of convergence result for universal lossy source coding.

For a given source let $D_n(R)$ be the n th-order operational distortion-rate function, that is, the minimum distortion achievable by an n -dimensional quantizer having rate less than or equal to R . For a given quantizer Q let $\Delta(Q)$ denote its expected distortion, and let $D(R) = \lim_n D_n(R)$ be the operational distortion-rate function of the source. In universal quantization (see e.g., [2], [9], [10], [16], [17], [20]) one is often interested in the quantity $\Delta(Q_n) - D(R)$, called the *distortion redundancy*, where $\{Q_n\}$ is a sequence of quantizers of rate $R_n \rightarrow R$. In fact, the fixed rate codes which satisfy $\lim_n (\Delta(Q_n) - D(R)) = 0$ for any source in a given class of sources are termed *universal block-to-block source codes* (see, e.g., [19], [11], and [7]). The distortion redundancy is usually decomposed into two terms, the "cost of universality" $\Delta(Q_n) - D_n(R)$, and the "cost of finite blocklength" $D_n(R) - D(R)$. These two quantities can be analyzed independently, and the asymptotic behavior of $D_n(R) - D(R)$ is often of interest in its own right. The cost of finite blocklength can also be used to give bounds on the distortion redundancy when combined with bounds on the cost of universality.

Our first result (Theorem 1) shows that $D_n(R) - D(R) \leq O(\sqrt{\log n/n})$ for memoryless real sources with finite sixth moment. Furthermore, the upper bound is achieved by a sequence of quantizers whose codevectors satisfy a power constraint uniformly in n . It turns out that the proof of this result poses substantially more technical difficulties than its counterpart for bounded sources.

Two corollaries of the main result are given. The first proves that for any real memoryless source of finite sixth moment with distortion-rate function $D(R)$ and a discrete memoryless channel of capacity C it is possible to transmit the block-quantized source through the channel so that the resulting mean square distortion is $D(C) + O(\sqrt{\log n/n})$, where n is the coding delay. This corollary generalizes a result of Wyner [14] who showed the same result for a memoryless Gaussian source.

The second corollary uses a result in [9] to show that for the class of real memoryless sources of finite k th absolute moment ($k > 8$) and for any $c > 2/(k - 4)$, there exists a universal quantizer whose distortion redundancy satisfies

$$\Delta(Q_n) - D(R) = O((\log n/n)^{(1/2)-\epsilon}).$$

Manuscript received October 10, 1994; revised July 20, 1995. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Whistler, BC, Canada, September 1995. This research was supported in part by the National Science Foundation, the Joint Services Electronics Program, the Foundation for Hungarian Higher Education and Research, and the Hungarian National Foundation for Scientific Research.

T. Linder is with the Department of Mathematics and Computer Science, Technical University of Budapest, Budapest, Hungary.

K. Zeger is with the Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, IL 61801 USA.

Publisher Item Identifier S 0018-9448(96)01032-2.

Thus for classes sources with finite moments of high enough order, the distortion redundancy arbitrarily approaches $O(\sqrt{\log n/n})$.

The results are stated in the next section. Their proofs along with the statements of some lemmas and their proofs are in Section III.

II. RESULTS

An n -dimensional quantizer with N codevectors

$$\{y_1, \dots, y_N\} \subset \mathbb{R}^n$$

is a measurable mapping

$$Q: \mathbb{R}^n \rightarrow \{y_1, \dots, y_N\}.$$

The rate $r(Q)$ of Q is defined as $r(Q) = (\log N)/n$, the logarithm being base 2. The mean-squared distortion between two n -vectors $x^n = (x_1, \dots, x_n)$ and $y^n = (y_1, \dots, y_n)$ is

$$d_n(x^n, y^n) = \frac{1}{n} \|x^n - y^n\|^2 = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|^2.$$

For a real i.i.d. source X_1, X_2, \dots the quantizer's distortion is given as

$$\Delta(Q) = \mathbf{E}d_n(X^n, Q(X^n)) = \frac{1}{n} \mathbf{E}\|X^n - Q(X^n)\|^2,$$

where $X^n = (X_1, \dots, X_n)$. Denote by $D_n(R)$ the minimum distortion achievable by quantizers of rate less than or equal to R

$$D_n(R) = \inf_{Q: r(Q) \leq R} \Delta(Q).$$

In what follows $D(R)$ will denote the distortion-rate function of the memoryless source $\{X_i\}$ relative to the mean-squared distortion, i.e.

$$D(R) = \inf\{\mathbf{E}|X - Y|^2 : I(X; Y) \leq R\}$$

where the infimum is taken over all joint distributions of pairs of real random variables (X, Y) such that the first marginal has the common distribution of the X_i and the mutual information between X and Y is at most R . We will prove the following result.

Theorem 1: Let X_1, X_2, \dots be a real i.i.d. source with $\mathbf{E}|X_1|^2 = M_2$ and $\mathbf{E}|X_1|^6 < \infty$. Let $0 < R_1 < R_2$ and assume that $D(R_2) > 0$. Then for any $R \in [R_1, R_2]$ there exists an n -dimensional quantizer Q_n with rate $r(Q_n) \leq R$ such that

$$\Delta(Q_n) \leq D(R) + B\sqrt{\frac{\log n}{n}}$$

for all $n \geq 1$, where the constant B depends only on R_1, R_2 , and the source distribution. Furthermore, the quantizers satisfy

$$\max_{x \in \mathbb{R}^n} \frac{1}{n} \|Q_n(x)\|^2 \leq 2M_2.$$

Remarks: a) The theorem implies that $\Delta(Q_n) - D(R) = O(\sqrt{\log n/n})$ for all $R > 0$ such that $D(R) > 0$. The theorem also shows the stronger statement that the convergence is uniform in any closed interval of rates containing neither zero nor $\min\{R : D(R) = 0\}$. This fact is used in the corollaries where sequences of quantizers with rates changing with n are considered.

b) Let the single-letter distortion measure d_n be given as

$$d_n(x^n, y^n) = n^{-1} \sum_{i=1}^n \rho(|x_i - y_i|)$$

where ρ is a strictly convex function with $\rho(0) = 0$ which satisfies $\rho(|x+y|) \leq c\rho(|x|) + c\rho(|y|)$ for some constant c . As will be discussed in Section IV in more detail, it is not hard to show (by slightly altering the proof) that Theorem 1 holds with the condition $\mathbf{E}\rho(|X_1|)^3 < \infty$ replacing $\mathbf{E}|X_1|^6 < \infty$. The additional assumption we need is that d_n satisfy the following: for any n , any closed convex set $K \subset \mathbb{R}^n$, and probability distribution P on K , the generalized centroid of K with respect to d_n is in K , i.e.

$$\min_{y \in K} \int_K d_n(x, y) P(dx) = \min_{y \in \mathbb{R}^n} \int_K d_n(x, y) P(dx). \quad (1)$$

One way to show (1) is to prove that for any closed convex set K and any $y \notin K$ there exists a $y' \in K$ such that

$$d_n(x, y') \leq d_n(x, y) \quad \text{for all } x \in K. \quad (2)$$

Interestingly, when $\rho(z) = z^r$ for some $r > 2$ (d_n is called the r th power distortion in this case), then (2) holds when $n = 2$ [8], but it fails to hold for $n \geq 3$ when K is a half-space.

In [14] Wyner proved that $D_n(R) - D(R) = O(\log n/n)$ for memoryless Gaussian sources, and in [15] showed that $D_n(R) - D(R) = O(\sqrt{\log n/n})$ for any correlated Gaussian source with a sufficiently well-behaved spectral density. Recently, Zamir and Feder [17] showed that an $O(\log n/n)$ convergence rate is achievable for correlated Gaussian sources by means of a variable-rate coding scheme using subtractive dither. The main result in [14] states that it is possible to transmit a memoryless Gaussian source over a memoryless channel of capacity C causing distortion $D(C) + O(\sqrt{\log n/n})$, where n is the coding delay. It turns out that our Theorem 1 is sufficient to obtain this result without assuming the source to be Gaussian. The joint source-channel coding scheme in the next corollary is the same as in [14] (see also, e.g., [5]) except that to simplify matters we let the channel accept one input per source output.

Corollary 1: Suppose we are given a real memoryless source X_1, X_2, \dots with distortion-rate function $D(R)$, satisfying $\mathbf{E}|X|^6 < \infty$, and a discrete memoryless channel of capacity C , accepting one input per source output. Then there exists a source-channel coding scheme with delay n , such that denoting by $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n$ the channel decoder output, we have

$$\frac{1}{n} \mathbf{E} \left(\sum_{i=1}^n |X_i - \hat{X}_i|^2 \right) \leq D(C) + O \left(\sqrt{\frac{\log n}{n}} \right).$$

The second corollary combines Theorem 1 with a result in [9] on the cost of universality in universal quantization. The class of sources considered here is the family of all real memoryless sources which satisfy the moment constraint $\mathbf{E}|X_1|^k < \infty$ for some $k > 8$.

Corollary 2: For any $R > 0$, $k > 8$, and $\epsilon > 2/(k-4)$ there exists a sequence of universal quantizers $\{Q_n\}$ such that

$$r(Q_n) - R = O\left(\left(\frac{\log n}{n}\right)^{(1/2)-\epsilon}\right)$$

and for any memoryless real source with $\mathbf{E}|X_1|^k < \infty$

$$\Delta(Q_n) - D(R) = O\left(\left(\frac{\log n}{n}\right)^{(1/2)-\epsilon}\right).$$

III. LEMMAS AND PROOFS

The proof of Theorem 1 uses similar techniques as Wyner's proof for correlated Gaussian sources. The main difficulty here is that unlike in the Gaussian case, the rate-distortion function of the source is not given explicitly, nor is the joint distribution achieving the distortion-rate bound. To overcome this problem we need some lemmas. First we cite three results (Lemmas 1-3) and then we prove Lemma 4, which will be the key to the main theorem.

In what follows the joint distribution of a pair of real random variables (X, Y) is denoted by $P_{X,Y}$, while $P_X \times P_Y$ refers to the product of the marginals of $P_{X,Y}$. If $D(R)$ is the distortion-rate function of the memoryless source X_1, X_2, \dots and X has the common distribution of the X_i , we say that a pair (X, Y) achieves the point $(R, D(R))$ (on the distortion-rate curve) if

$$I(X; Y) = R \quad \text{and} \quad \mathbf{E}|X - Y|^2 = D(R).$$

The following lemma on the parametric representation of $D(R)$ is due to Csiszár [4]. Although the result holds under much more general conditions, we state it here for real sources and the squared distortion.

Lemma 1 ([4, Lemma 1.4]): Let X be a real random variable with $\mathbf{E}|X|^2 < \infty$, and let $D(R)$ be the distortion-rate function of the memoryless source generated by X . Let $R > 0$ such that $D(R) > 0$, and assume that (X, Y) achieves $(R, D(R))$. Then the Radon-Nikodym derivative $a(x, y)$ of $P_{X,Y}$ with respect to $P_X \times P_Y$ is given as

$$a(x, y) = \alpha(x)2^{-s|x-y|^2} \quad (3)$$

for some $s \geq 0$ and $\alpha(x) \geq 1$ satisfying for all $y \in \mathbb{R}$

$$\int_{\mathbb{R}} \alpha(x)2^{-s|x-y|^2} P_X(dx) \leq 1. \quad (4)$$

Furthermore, $-1/s = D'(R)$, the derivative of $D(\cdot)$ at $R > 0$.

The next lemma is due to Wyner [15]. Wyner's lemma is a slight modification of [5, Lemma 9.3.1] by Gallager, the main step in the classic random coding proof of the source coding theorem. The only change here is that we replaced the probability density functions with R-N derivatives to deal with arbitrary source distributions, which is easily seen to be legitimate.

Lemma 2 ([15, Lemma 3]): Let X_1, X_2, \dots be an i.i.d. source and let $(X_1, Y_1), (X_2, Y_2), \dots$ be an i.i.d. sequence of pairs such that $I(X_1; Y_1) < \infty$. Then for any α, β, γ and any positive integer N there exists an n -dimensional quantizer Q with N codevectors such that

$$\mathbf{P}\left\{\frac{1}{n}\|X^n - Q(X^n)\|^2 > \alpha\right\} \leq P_n^{(1)}(\alpha) + P_n^{(2)}(\beta) + P_n^{(3)}(\gamma) + \exp(-(N-1)2^{-n\beta})$$

where

$$P_n^{(1)}(\alpha) = \mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n |X_i - Y_i|^2 > \alpha\right\}$$

$$P_n^{(2)}(\beta) = \mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n \log a(X_i, Y_i) > \beta\right\}$$

the function $a(x, y)$ being the R-N derivative $\frac{dP_{X_1, Y_1}}{dP_{X_1} \times P_{Y_1}}$, and

$$P_n^{(3)}(\gamma) = \mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n |Y_i|^2 > \gamma\right\}.$$

Furthermore, Q satisfies for all $x^n \in \mathbb{R}^n$

$$\frac{1}{n}\|Q(x^n)\|^2 \leq \gamma \quad \text{and} \quad \|x^n - Q(x^n)\| \leq \|x^n\|. \quad (5)$$

The following is the Berry-Esseen theorem [3, Corollary 9.1.4], a uniform version of the central limit theorem for i.i.d. variables with finite third moment.

Lemma 3: Let Z_1, \dots, Z_n be i.i.d. random variables with $\mathbf{E}Z_1 = 0$, $\mathbf{E}|Z_1|^2 = \sigma^2$, and $\mathbf{E}|Z_1|^3 = \gamma < \infty$. Then

$$\sup_{-\infty < x < \infty} \left| \mathbf{P}\left\{\sum_{i=1}^n Z_i < x\sigma\sqrt{n}\right\} - \Phi(x) \right| \leq \frac{1}{\sqrt{n}} \frac{\gamma}{\sigma^3}$$

where

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x e^{-t^2/2} dt.$$

Using the well-known inequality

$$1 - \Phi(x) < (x\sqrt{2\pi})^{-1} e^{-x^2/2}$$

for all $x > 0$, and choosing $x = u\sqrt{n}/\sigma$ we obtain the following: for all $u > 0$

$$\mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n Z_i > u\right\} \leq \frac{\sigma}{u\sqrt{n}\sqrt{2\pi}} e^{-nu^2/(2\sigma^2)} + \frac{1}{\sqrt{n}} \frac{\gamma}{\sigma^3}. \quad (6)$$

The last lemma will be essential in establishing our main result. It shows that there always exists an (X, Y) achieving $(R, D(R))$ such that the absolute moments of Y are bounded above by those of X . An interesting twist in the proof is that while usually the existence of near-optimal quantizers are derived using the existence an optimizing (X, Y) , this time the existence of an optimizing (X, Y) with certain properties is inferred from the source coding theorem and properties of vector quantizers.

Lemma 4: Let X be a real random variable satisfying $\mathbf{E}|X|^r < \infty$ for some $r \geq 2$. Let $D(R)$ be the distortion-rate function of X relative to the squared error criterion, and let r_1, \dots, r_l be positive reals satisfying $2 \leq r_j \leq r$, $j = 1, \dots, l$. Then for any $R > 0$ there exists a Y such that (X, Y) achieves the point $(R, D(R))$, and the absolute r_j th moments of Y satisfy

$$\mathbf{E}|Y|^{r_j} \leq \mathbf{E}|X|^{r_j}, \quad j = 1, \dots, l.$$

Proof: The proof has two steps. First we will show that for all $\epsilon > 0$

$$D(R) = \inf\{\mathbf{E}|X - Y|^2 : I(X; Y) \leq R, \mathbf{E}|Y|^{r_j} \leq M_{r_j} + \epsilon, j = 1, \dots, l\} \quad (7)$$

where $M_{r_j} = \mathbf{E}|X|^{r_j}$. The next step is to show that (7) implies the claim of the lemma.

It suffices to show that $D(R)$ is greater than or equal to the right-hand side of (7) for any $\epsilon > 0$ since the opposite inequality is obvious. We prove this using the following construction. For any r_j and $\epsilon > 0$ let $B_{n,\epsilon}^{r_j}$ be defined as

$$B_{n,\epsilon}^{r_j} = \{x^n = (x_1, \dots, x_n) \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^{r_j} \leq n(M_{r_j} + \epsilon)\}$$

and let

$$B_{n,\epsilon} = \bigcap_{j=1}^l B_{n,\epsilon}^{r_j}. \quad (8)$$

If $X^n = (X_1, \dots, X_n)$ is a vector of i.i.d. copies of X , the weak law of large numbers and the union bound imply that $\mathbf{P}\{X^n \notin B_{n,\epsilon}\} \rightarrow 0$ as $n \rightarrow \infty$. Then we have by symmetry

$$\mathbf{E}\left[\frac{1}{n} \sum_{i=1}^n |X_i|^2 I_{\{X^n \notin B_{n,\epsilon}\}}\right] = \mathbf{E}[|X_1|^2 I_{\{X^n \notin B_{n,\epsilon}\}}] \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (9)$$

Consider now a sequence of n -dimensional quantizers $\{Q_n\}$ such that Q_n has $N_n \leq 2^{nR}$ codevectors for each n and $(1/n)\mathbf{E}\|X^n - Q_n(X^n)\|^2 \rightarrow D(R)$ as $n \rightarrow \infty$. Such a sequence of quantizers exists by the source coding theorem (see, e.g., [5, Theorem 9.6.2]). Denote the codevectors of Q_n by $\{y_1, \dots, y_{N_n}\}$, where we can assume that $y_1 = 0$ since adding one extra codeword to the codebook asymptotically does not affect the quantizer's rate. We construct a new codebook $\{\hat{y}_1, \dots, \hat{y}_{N_n}\}$ by projecting the y_i lying outside $B_{n,\epsilon}$ onto $B_{n,\epsilon}$, that is, we redefine the y_i for $i = 1, \dots, N_n$ as

$$\hat{y}_i = \arg \min\{\|x^n - y_i\| : x^n \in B_{n,\epsilon}\}.$$

Since the r_j are greater than 1, the $B_{n,\epsilon}^{r_j}$ are convex, and it follows that $B_{n,\epsilon}$ is convex. Thus the \hat{y}_i exist and are unique. By a well-known property of convex sets and the Euclidean distance (which follows from the Pythagorean identity) we have for all $x^n \in B_{n,\epsilon}$, and for all $i = 1, \dots, N_n$

$$\|x^n - \hat{y}_i\| \leq \|x^n - y_i\|. \quad (10)$$

Now define a new quantizer \hat{Q}_n as

$$\hat{Q}_n(x^n) = \begin{cases} \hat{y}_i & \text{if } \|x^n - \hat{y}_i\| = \min_{1 \leq j \leq N_n} \|x^n - \hat{y}_j\| \\ & \text{and } x^n \in B_{n,\epsilon} \\ y_1 = 0, & \text{if } x^n \notin B_{n,\epsilon} \end{cases} \quad (11)$$

where ties are broken arbitrarily. By (10) we have

$$\|x^n - \hat{Q}_n(x^n)\| \leq \|x^n - Q_n(x^n)\| \quad \text{if } x^n \in B_{n,\epsilon}.$$

From this and (9) we obtain

$$\frac{1}{n} \mathbf{E}\|X^n - \hat{Q}_n(X^n)\|^2 = D(R) + \delta_n \quad (12)$$

where $\delta_n \rightarrow 0$ as $n \rightarrow \infty$, while the rate of \hat{Q}_n is at most R , and all the codevectors of \hat{Q}_n are in $B_{n,\epsilon}$. In particular, (8) implies that

$$\frac{1}{n} \mathbf{E}(\|\hat{Q}_n(X^n)\|_{r_j}^{r_j}) \leq M_{r_j} + \epsilon \quad (13)$$

for all r_j , where

$$\|x^n\|_{r_j}^{r_j} = \sum_{i=1}^n |x_i|^{r_j}.$$

Let $Y^n = (Y_1, \dots, Y_n) = \hat{Q}_n(X^n)$. Then by the properties of mutual information (see [5, (9.6.2), p. 470])

$$R \geq \frac{1}{n} I(X^n; Y^n) \geq \frac{1}{n} \sum_{i=1}^n I(X_i, Y_i). \quad (14)$$

Denoting the distribution of the pair (X_i, Y_i) by $P_{X_i Y_i}$, we define the pair of random variables (X, \bar{Y}) to have the distribution

$$P_{X\bar{Y}} = \frac{1}{n} \sum_{i=1}^n P_{X_i Y_i}.$$

Then the convexity of $I(X, Y)$ in $P_{Y|X}$ for fixed P_X implies that

$$\frac{1}{n} \sum_{i=1}^n I(X_i, Y_i) \geq I(X, \bar{Y}). \quad (15)$$

By the definition of the distribution of \bar{Y} we have

$$\mathbf{E}|X - \bar{Y}|^2 = \frac{1}{n} \mathbf{E}\|X^n - Y^n\|^2 = D(R) + \delta_n$$

by (12), and

$$\mathbf{E}|\bar{Y}|^{r_j} = \mathbf{E}\left(\frac{1}{n} \sum_{i=1}^n |Y_i|^{r_j}\right) \leq M_{r_j} + \epsilon$$

by (13). Combining this with (14) and (15) we obtain

$$D(R) \geq \inf\{\mathbf{E}|X - Y|^2 : I(X; Y) \leq R, \mathbf{E}|Y|^{r_j} \leq M_{r_j} + \epsilon, j = 1, \dots, l\}$$

which completes the proof of (7).

Next we show that (7) indeed implies the existence of the said random variable Y . Clearly, the union over all $0 < \epsilon \leq 1$ of the families of probability distributions on \mathbb{R}^2 considered on the right-hand side of (7) is tight. It follows from Prohorov's theorem [1] that there exists a sequence of random variables Y_1, Y_2, \dots such that (X, Y_n) converges in distribution to a pair (X, Y) , and $I(X, Y_n) \leq R$, $\lim_n \mathbf{E}|X - Y_n|^2 = D(R)$, and $\limsup_n \mathbf{E}|Y_n|^{r_j} \leq M_{r_j}$, $j = 1, \dots, l$. Then, by [1, Theorem 5.3] we obtain

$$\mathbf{E}|X - Y|^2 \leq D(R), \quad \mathbf{E}|Y|^{r_j} \leq M_{r_j}, j = 1, \dots, l.$$

Finally, the lower semicontinuity property of the mutual information (see [4, Lemma 2.2]) implies $R \geq \liminf_n I(X; Y_n) \geq I(X, Y)$, so that the pair (X, Y) satisfies the claim of the lemma. The proof of the lemma is thus complete. \square

Proof of Theorem 1: Let $R \in [R_1, R_2]$ be fixed and let X have the common distribution of the X_i . Then by Lemma 4 there exists a Y such that the pair (X, Y) achieves $(R, D(R))$, and

$$\begin{aligned} \mathbf{E}|Y|^2 &\leq \mathbf{E}|X|^2 = M_2 \\ \mathbf{E}|Y|^4 &\leq \mathbf{E}|X|^4 \end{aligned}$$

and

$$\mathbf{E}|Y|^6 \leq \mathbf{E}|X|^6. \quad (16)$$

Let $(X_1, Y_1), (X_2, Y_2), \dots$ be i.i.d. copies of (X, Y) . For a given $u > 0$ let $\alpha = D(R) + u$, $\beta = R + u/2$, $\gamma = 2M_2$, and let Q be the n -dimensional quantizer with $N = \lfloor 2^{n(R+u)} \rfloor$ codevectors whose existence is assured by Lemma 2 with the given α , β , and γ . Then we have that

$$\frac{1}{n} \|Q(x^n)\|^2 \leq 2M_2 \quad \text{for all } x \in \mathbb{R}^n.$$

The distortion of Q can be written as

$$\begin{aligned} \Delta(Q) &= \frac{1}{n} \mathbf{E} \left\{ \|X^n - Q(X^n)\|^2 I_{\{\|X^n\|^2 \leq 2nM_2\}} \right\} \\ &\quad + \frac{1}{n} \mathbf{E} \left\{ \|X^n - Q(X^n)\|^2 I_{\{\|X^n\|^2 > 2nM_2\}} \right\}. \quad (17) \end{aligned}$$

Using (5) and the Cauchy-Schwarz and Chebyshev inequalities we can upper-bound the last term in (17) as

$$\begin{aligned} &\frac{1}{n} \mathbf{E} \left\{ \|Q(X^n) - X^n\|^2 I_{\{\|X^n\|^2 > 2nM_2\}} \right\} \\ &\leq \frac{1}{n} \sqrt{\mathbf{E} \|X^n\|^4} \sqrt{\mathbf{P}\{\|X^n\|^2 > 2nM_2\}} \\ &\leq \sqrt{\mathbf{E} |X_1|^4} \sqrt{\frac{\mathbf{Var}(X_1^2)}{nM_2^2}} \\ &= \frac{c}{\sqrt{n}} \quad (18) \end{aligned}$$

where c depends only on the second and fourth moments of X .

To upper-bound the first term on the right side of (17) we again use the fact that $\|x^n - Q(x^n)\| \leq \|x\|$. Letting

$$W = n^{-1} \|Q(X^n) - X^n\|^2 I_{\{\|X^n\|^2 \leq 2nM_2\}}$$

we can write (using the fact that $W \leq 2M_2$ a.s.)

$$\begin{aligned} \mathbf{E}W &\leq D(R) + u + 2M_2 \mathbf{P}\{W > D(R) + u\} \\ &\leq D(R) + u + 2M_2 \mathbf{P}\{n^{-1} \|X^n - Q(X^n)\|^2 > D(R) + u\}. \quad (19) \end{aligned}$$

By Lemma 2 we have

$$\begin{aligned} \mathbf{P}\{n^{-1} \|X^n - Q(X^n)\|^2 > D(R) + u\} \\ &\leq P_n^{(1)}(D(R) + u) + P_n^{(2)}(R + u/2) \\ &\quad + P_n^{(3)}(2M_2) + \exp(-(N-1)2^{-n\beta}). \quad (20) \end{aligned}$$

The third probability

$$P_n^{(3)}(2M_2) = \mathbf{P}\left\{ \sum_{i=1}^n |Y_i|^2 > 2nM_2 \right\}$$

is easily bounded using the Chebyshev inequality and the fact that $\mathbf{E}|Y|^4 \leq \mathbf{E}|X|^4$ as

$$\begin{aligned} \mathbf{P}\left\{ \sum_{i=1}^n |Y_i|^2 > 2nM_2 \right\} &\leq \frac{\mathbf{Var}(|Y|^2)}{nM_2^2} \\ &\leq \frac{c_2}{n} \quad (21) \end{aligned}$$

where c_2 depends only on the moments of X . The last term in (20) is no greater than $\exp(-2^{nu/2} - 2)$ by the choice of N . In what follows we provide uniform upper bounds for the two remaining probabilities in (20).

Uniform Bound on $P_n^{(1)}(D(R) + u)$

Using (6) (the corollary of the Berry-Esseen theorem) with $Z_i = |X_i - Y_i|^2 - D(R)$ we can upper-bound $P_n^{(1)}(D(R) + u)$ as

$$\begin{aligned} P_n^{(1)}(D(R) + u) &\leq \frac{\sqrt{A(R)}}{u\sqrt{n}2\pi} e^{-nu^2/(2A(R))} \\ &\quad + \frac{1}{\sqrt{n}} \frac{B(R)}{A(R)^{3/2}} \quad (22) \end{aligned}$$

where $A(R) = \mathbf{Var}(|X - Y|^2)$ and $B(R) = \mathbf{E} | |X - Y|^2 - D(R) |^3$. The moment conditions (16) and the inequality

$$(a + b)^p \leq 2^{p-1}(a^p + b^p), \quad a, b > 0, p \geq 1$$

imply that both $A(R)$ and $B(R)$ are uniformly upper-bounded for all R in $[R_1, R_2]$. Thus we only have to show that $A(R)$ is bounded away from zero in $[R_1, R_2]$.

This is seen by the following argument. Suppose

$$\inf_{R \in [R_1, R_2]} A(R) = 0.$$

Then there exists a sequence $\{R_k\}$ in $[R_1, R_2]$ converging to an $\hat{R} \in [R_1, R_2]$, such that the associated pairs (X, Y_k) achieve the points $(R_k, D(R_k))$ and satisfy

$$\mathbf{Var}(|X - Y_k|^2) = \mathbf{E} \left\{ | |X - Y_k|^2 - D(R_k) |^2 \right\} \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Also, we can assume by Lemma 4 that $\mathbf{E}|Y_k|^2 \leq \mathbf{E}|X|^2$, and this moment constraint implies that the family of probability measures induced by the (X, Y_k) on \mathbb{R}^2 is tight. Thus there exists a subsequence R_{k_j} and a random variable \hat{Y} such that (X, Y_{k_j}) converges in distribution to (X, \hat{Y}) , and for simplicity we will write R_k instead of R_{k_j} . It follows that

$$|X - Y_k|^2 - D(R_k) \rightarrow |X - \hat{Y}|^2 - D(\hat{R})$$

in distribution (see [1, Theorem 5.5]). Then a simple truncation argument shows that

$$\mathbf{E} \left| |X - \hat{Y}|^2 - D(\hat{R}) \right|^2 \leq \liminf_k \mathbf{E} \left| |X - Y_k|^2 - D(R_k) \right|^2 = 0$$

(see, e.g., [1, Theorem 5.3]). This implies that $|X - \hat{Y}| = c$ almost surely. Now since

$$\hat{a}(x, y) = \frac{dP_{X, \hat{Y}}}{dP_X \times P_{\hat{Y}}}(x, y) = \hat{\alpha}(x) 2^{-\hat{s}|x-y|^2} > 0$$

we have that $P_X \times P_{\hat{Y}}$ is absolutely continuous with respect to $P_{X, \hat{Y}}$. It follows that $P_X \times P_{\hat{Y}}$ and $P_{X, \hat{Y}}$ have the same support. By assumption the support of $P_{X, \hat{Y}}$ is contained in the set $S = \{(x, y) \in \mathbb{R}^2 : |x - y| = c\}$, and thus so is the support of $P_X \times P_{\hat{Y}}$. But the support of $P_X \times P_{\hat{Y}}$ is a product set $A \times B$, and $A \times B \subset S$ can only happen if $A = \{x\}$ for some $x \in \mathbb{R}$. This would mean that P_X is concentrated on x , which contradicts our assumption that $D(R_2) > 0$. We conclude that there exist constants C_1, C_2 , and C_3 such that

$$P_n^{(1)}(D(R) + u) \leq \frac{C_1}{u\sqrt{n}} e^{-nu^2/(2C_2)} + \frac{C_3}{\sqrt{n}} \quad (23)$$

for all $R \in [R_1, R_2]$.

Uniform Bound on $P_n^{(2)}(R + u/2)$

Using the Berry–Esseen inequality again we obtain

$$P_n^{(2)}(R + u/2) \leq \frac{\sqrt{\hat{A}(R)}}{u\sqrt{n}2\pi} e^{-nu^2/(2\hat{A}(R))} + \frac{1}{\sqrt{n}} \frac{\hat{B}(R)}{\hat{A}(R)^{3/2}}$$

where $\hat{A}(R) = \mathbf{Var}(\log a(X, Y))$, $\hat{B}(R) = \mathbf{E}|\log a(X, Y) - R|^3$, and

$$a(x, y) = \frac{dP_{X_1, Y_1}}{dP_{X_1} \times P_{Y_1}}(x, y).$$

To show that $\hat{B}(R)$ is uniformly upper-bounded in $[R_1, R_2]$ we will make use of Lemma 1. Since (X, Y) achieves $(R, D(R))$, by (3) we have

$$\begin{aligned} \mathbf{E}|\log a(X, Y)|^3 &= \mathbf{E}|\log \alpha(X) - s|X - Y|^2|^3 \\ &\leq 2^2 (\mathbf{E}|\log \alpha(X)|^3 + s^3 \mathbf{E}|X - Y|^6). \end{aligned} \quad (24)$$

Since

$$\alpha(x) = \left(\int 2^{-s|x-y|^2} P_Y(dy) \right)^{-1} \geq 1$$

for P_X a.e., by Jensen's inequality applied to $-\log z$ and $|z|^3$ we have

$$\begin{aligned} \mathbf{E}|\log \alpha(X)|^3 &= \mathbf{E}(\log \alpha(X))^3 \\ &= \mathbf{E} \left(-\log \left(\int 2^{-s|x-y|^2} P_Y(dy) \right) \right)^3 \\ &\leq \mathbf{E} \left(\int s|x-y|^2 P_Y(dy) \right)^3 \\ &\leq s^3 \int \int |x-y|^6 P_Y(dy) P_X(dx) \\ &\leq s^3 2^5 (\mathbf{E}|X|^6 + \mathbf{E}|Y|^6) \quad (25) \\ &\leq s^3 2^6 \mathbf{E}|X|^6. \quad (26) \end{aligned}$$

Thus $\mathbf{E}|\log a(X, Y)|^3$ is upper-bounded by $2^6 s^3 \mathbf{E}|X|^6$. Since $-1/s$ is the derivative of the decreasing convex function $D(R)$, it is readily seen that $R_1 > 0$, $D(R_2) > 0$, and the fact that $\lim_{R \rightarrow \infty} D(R) = 0$ imply that the values of s associated with $R \in [R_1, R_2]$ are bounded and bounded away from zero.

Next we show (the intuitively obvious fact) that $\hat{A}(R)$ is bounded away from zero in $[R_1, R_2]$. Assume the contrary. Then there exists a sequence R_k in $[R_1, R_2]$ such that $R_k \rightarrow \hat{R} \in [R_1, R_2]$, and for which $\mathbf{Var}(\log a_k(X, Y_k)) \rightarrow 0$ as $k \rightarrow \infty$ (the R-N derivatives a_k are also indexed since they depend on (X, Y_k)). Using the same argument about the tightness of the family of distributions induced by the (X, Y_k) as above, we can again assume that (X, Y_k) converges in distribution to a pair (X, \hat{Y}) . As before we obtain that (X, \hat{Y}) achieves the point $(\hat{R}, D(\hat{R}))$ on the distortion-rate curve. By the parametric representation of $a_k(x, y)$ we have

$$a_k(x, y) = \alpha_k(x) 2^{-s_k|x-y|^2}$$

where

$$\alpha_k(x) = \left(\int 2^{-s_k|x-y|^2} P_{Y_k}(dy) \right)^{-1}.$$

Since $D(R)$ is continuously differentiable in $[R_1, R_2]$ (a fact which follows from the existence of the derivative for any $R \in [R_1, R_2]$), we have $\lim_k s_k = \hat{s} > 0$, where \hat{s} is the parameter in the parametric representation of

$$\hat{a}(x, y) = \frac{dP_{X, \hat{Y}}}{dP_X \times P_{\hat{Y}}}(x, y).$$

Using a standard argument involving truncation of integrals and the fact that $Y_k \rightarrow \hat{Y}$ in distribution one can see that

$$\lim_{k \rightarrow \infty} \alpha_k(x_k) = \hat{\alpha}(x) = \left(\int 2^{-\hat{s}|x-y|^2} P_{\hat{Y}}(dy) \right)^{-1}$$

for any $x \in \mathbb{R}$ and any sequence $x_k \rightarrow x$. Thus we also have for all $x, y \in \mathbb{R}$ and all sequences $x_k \rightarrow x$, that

$$\begin{aligned} \log a_k(x_k, y) &= \log \alpha_k(x_k) - s_k|x_k - y|^2 \\ &\rightarrow \log \hat{\alpha}(x) - \hat{s}|x - y|^2 \\ &= \log \hat{a}(x, y) \end{aligned}$$

as $k \rightarrow \infty$. This implies that $\log a_k(X, Y_k) \rightarrow \log \hat{a}(X, \hat{Y})$ in distribution by [1, Theorem 5.5], and thus we also have

$$\mathbf{Var}(\log \hat{a}(X, \hat{Y})) = \lim_{k \rightarrow \infty} \mathbf{Var}(\log a_k(X, Y_k)) = 0.$$

Thus $\log \hat{a}(x, y) = \hat{R}$ for $P_{X, \hat{Y}}$ a.e. But by a previous argument $P_X \times P_{\hat{Y}}$ is absolutely continuous with respect to $P_{X, \hat{Y}}$, implying

$$\hat{a}(x, y) = 2^{\hat{R}} \quad \text{for } P_X \times P_{\hat{Y}} \text{ a.e. } (x, y).$$

This gives $2^{\hat{R}} = 1$, a contradiction, since $\hat{R} \geq R_1 > 0$. Thus we have shown that $\hat{A}(R)$ is bounded away from zero in $[R_1, R_2]$.

We conclude that there exist constants \hat{C}_1 , \hat{C}_2 , and \hat{C}_3 such that

$$P_n^{(2)}(R + u/2) \leq \frac{\hat{C}_1}{u\sqrt{n}} e^{-nu^2/(2\hat{C}_2)} + \frac{\hat{C}_3}{\sqrt{n}} \quad (27)$$

for all $R \in [R_1, R_2]$.

Combining (17) and (19) with the derived bounds on the the probabilities in (20) we can conclude that

$$\begin{aligned} \frac{1}{n} \mathbf{E} \|X^n - Q(X^n)\|^2 &\leq D(R) + u + \frac{A}{\sqrt{n}} + e^{-2nu+2} \\ &\quad + \frac{C_1}{u\sqrt{n}} e^{-nu^2/(2C_2)} \\ &\quad + \frac{\hat{C}_1}{u\sqrt{n}} e^{-nu^2/(2\hat{C}_2)} \end{aligned} \quad (28)$$

holds for all $u > 0$, all positive integers n , and all $R \in [R_1, R_2]$, where A is a constant. The asymptotically optimal choice of u is, of course, $u = \sqrt{C_4(\log n)/n}$, where $C_4 = (\log e) \max\{C_2, \hat{C}_2\}$. With this u we conclude that for any $R \in [R_1, R_2]$ there exists a quantizer Q whose rate is bounded as

$$r(Q) \leq R + C_4 \sqrt{\frac{\log n}{n}}$$

and whose distortion satisfies: There exists a constant C_5 such that for all $n \geq 1$

$$\frac{1}{n} \mathbf{E} \|X^n - Q(X^n)\|^2 \leq D(R) + C_5 \sqrt{\frac{\log n}{n}}$$

Finally, we replace R by $R_n = R - \sqrt{C_4(\log n)/n}$ in the above bound and assume that n is large enough to ensure that $\sqrt{C_4(\log n)/n} \leq \delta$, for a given $\delta > 0$. We then use a first-order Taylor expansion of $D(R)$ around R and the fact that the derivative of $D(R)$ is bounded in $[R_1, R_2]$. This proves for any $R \in [R_1 + \delta, R_2]$, there exists a quantizer Q of rate $r(Q) \leq R$ such that

$$\Delta(Q) \leq D(R) + B \sqrt{\frac{\log n}{n}}$$

for some constant B . Now renaming $R_1 + \delta$ as R_1 (and repeating the same argument) we obtain the claim of the theorem. \square

Proof of Corollary 1: The corollary is straightforwardly obtained by combining Wyner's proof of the source-channel transmission theorem for a Gaussian source in [14] with our Theorem 1. The point is that the result holds for any source for which there exists a sequence of quantizers Q_n ($n = 1, 2, \dots$) such that $r(Q_n) \rightarrow R$, the codewords satisfy the power constraint

$$\max_{x^n \in \mathbb{R}^n} \|Q_n(x^n)\|^2 \leq nd$$

for some $d < \infty$, and for which we have

$$\Delta(Q_n) \leq D(r(Q_n)) + O(\sqrt{\log n/n}).$$

The sketch of the argument separating the source and channel coding is as follows.

Let $R_n \leq C$ and consider an n -dimensional quantizer Q_n with rate $R(Q_n) \leq R_n$ and

$$\max_{x^n \in \mathbb{R}^n} \|Q_n(x^n)\|^2 \leq nd.$$

Since the source produces one sample per channel symbol, the channel code will have $N_n = 2^{nR_n}$ codewords of length n . Denoting the codevectors of Q_n by $\{y_1, \dots, y_{N_n}\}$, the channel decoder maps n -blocks of the channel output symbols into $\{y_1, \dots, y_{N_n}\}$. Letting \hat{X}^n be the output of the channel decoder we can write

$$\begin{aligned} \frac{1}{n} \mathbf{E} \|X^n - \hat{X}^n\|^2 &\leq \left(\sqrt{\frac{1}{n} \mathbf{E} \|X^n - Q_n(X^n)\|^2} \right. \\ &\quad \left. + \sqrt{\frac{1}{n} \mathbf{E} \|Q_n(X^n) - \hat{X}^n\|^2} \right)^2. \end{aligned} \quad (29)$$

Now let

$$P_{\max}^{(n)} = \max_{1 \leq i \leq N_n} \mathbf{P}\{\hat{X}^n \neq y_i \mid Q(X^n) = y_i\}.$$

Then we have

$$\mathbf{E} \|Q_n(X^n) - \hat{X}^n\|^2 \leq P_{\max}^{(n)} 4dn$$

since

$$\max_{i,j} \|y_i - y_j\|^2 \leq 4dn.$$

By a result of Gallager [5, Problem 5.23] (see also [13]) for all $R_n \leq C$ there exists an n -length channel code of rate R_n such that

$$P_{\max}^{(n)} \leq 4e^{-n(C-R_n)^2/\alpha}$$

where $\alpha = 8e^{-2} + 2(\ln J)^2$, J being the size of the channel output alphabet. Choosing $R_n = C - \sqrt{\alpha(\ln 2)(\log n/n)}$ we obtain

$$\frac{1}{n} \mathbf{E} \|X^n - \hat{X}^n\|^2 \leq \frac{16d}{n}. \quad (30)$$

Since by Theorem 1 there exists a sequence Q_n with rate $r(Q_n) \leq R_n$ satisfying

$$\Delta(Q_n) - D(C) = O(\sqrt{\log n/n})$$

and

$$\|Q_n(x^n)\|^2 \leq n2\mathbf{E}|X_1|^2 \text{ for all } x^n \in \mathbb{R}^n$$

the corollary follows from (29) and (30). \square

Proof of Corollary 2: To prove the corollary we combine a result on the cost of universality in universal quantization with Theorem 1. In [9] it was proved that for any $k > 8$, $\epsilon > 2/(k-4)$, and $R > 0$ there exists a sequence of quantizers with rate $r(Q_n) = R + O((\log n/n)^{(1/2)-\epsilon})$ such that for any memoryless real source X_1, X_2, \dots with $\mathbf{E}|X_1|^k < \infty$

$$\Delta(Q_n) - D_n(R) = O\left(\left(\frac{\log n}{n}\right)^{(1/2)-\epsilon}\right).$$

Since $D_n(R) - D(R) = O(\sqrt{\log n/n})$ (by Theorem 1) for sources considered here, the claim of the corollary follows. \square

IV. CONCLUSION

We have investigated upper bounds on the cost of finite blocklength in block quantization of real sources of unbounded support. We obtained an $O(\sqrt{\log n/n})$ bound on the distortion redundancy $D_n(R) - D(R)$ for squared distortion and showed that the same bound is achievable when the source is transmitted through a memoryless channel of capacity $C \geq R$. We note here that it is suggestive that for at least a smaller class of real memoryless sources an $O(\log n/n)$ upper (and possibly lower) bound might be obtainable for the distortion redundancy. This is supported by Pile's $O(\log n/n)$ upper bound [12] for finite alphabet memoryless sources, by Wyner's $O(\log n/n)$ bound for Gaussian sources, and by a heuristic high-rate quantization argument concerning the so-called "coefficient of quantization" [6].

As we noted in Section II, Theorem 1 can be extended to single-letter distortion measures of the form

$$d_n(x^n, y^n) = \frac{1}{n} \sum_{i=1}^n \rho(|x_i - y_i|)$$

where ρ is a strictly convex nonnegative function with $\rho(0) = 0$. For this extension we require that ρ satisfy (1) and the following inequality (meant to replace the triangle inequality):

$$\rho(x + y) \leq c\rho(x) + c\rho(y) \quad \text{for all } x, y \geq 0 \quad (31)$$

for some constant c .

To see why this extension is possible we note that Lemma 1 (Csiszár) holds for even more general distortion measures, and that Lemma 2 (Wyner) straightforwardly extends to this case (for these two lemmas to hold neither (1) nor (31) are

needed). To prove Lemma 4 (with $\rho(|\cdot|)$ replacing $|\cdot|^2$) we need only use (1) to show the existence of the \hat{y}_i satisfying (10). The proof of Theorem 1 is virtually unchanged, except that (31) is used to replace the triangle inequality where needed. Note that Corollary 1 can be extended in the same manner. Unfortunately, however, we could not prove that the r th-power distortion satisfies the centroid condition (1), although it clearly satisfies (31) with $c = 2^{r-1}$.

REFERENCES

- [1] P. Billingsley, *Convergence of Probability Measures*. New York: Wiley, 1968.
- [2] P. A. Chou, M. Effros, and R. M. Gray, "A vector quantization approach to universal noiseless coding and quantization," submitted to *IEEE Trans. Inform. Theory*, 1994.
- [3] Y. S. Chow and H. Teicher, *Probability Theory, Independence, Interchangeability, Martingales*. New York: Springer-Verlag, 1988.
- [4] I. Csiszár, "On the extremum problem of information theory," *Studia Scient. Math. Hungarica*, pp. 57-70, 1974.
- [5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [6] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373-380, July 1979.
- [7] J. C. Kieffer, "A unified approach to weak universal source coding," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 674-682, Nov. 1978.
- [8] G. László, personal communication, 1994.
- [9] T. Linder, G. Lugosi, and K. Zeger, "Fixed rate universal lossy source coding and rate of convergence for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 41, pp. 665-676, May 1995.
- [10] ———, "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.
- [11] 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.
- [12] R. Pile, "The transmission distortion of a source as a function of the encoding block length," *Bell Syst. Tech. J.*, vol. 47, pp. 827-885, 1968.
- [13] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inform. Contr.*, vol. 1, pp. 6-25, 1957.
- [14] A. D. Wyner, "Communication of analog data from a Gaussian source over a noisy channel," *Bell Syst. Tech. J.*, pp. 801-812, May-June 1968.
- [15] ———, "On the transmission of correlated Gaussian data over a noisy channel with finite encoding block length," *Inform. Contr.*, vol. 20, pp. 193-215, 1972.
- [16] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Trans. Inform. Theory*, vol. 38, pp. 428-436, Mar. 1992.
- [17] ———, "Information rates of pre/post filtered dithered quantizers," submitted to *IEEE Trans. Inform. Theory*, 1993.
- [18] Z. Zhang, E.-H. Yang, and V. K. Wei, "The redundancy of source coding with a fidelity criterion," to appear in *IEEE Trans. Inform. Theory*, 1996.
- [19] 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.
- [20] ———, "On universal quantization," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 344-347, May 1985.