Error Exponents in Scalable Source Coding

Ertem Tuncel and Kenneth Rose
{ertem,rose}ece.ucsb.edu
Dept. of ECE, University of California, Santa Barbara, CA 93106

Abstract— **We extend the error exponent results to 2-layer scalable source coding. We consider separate error events at each layer so as to allow a trade-off analysis for the error exponents when the rate and distortion values are fixed. Given a discrete memoryless source, we derive the single-letter characterization of the region of all achievable 6-tuples $(R_1, R_2, D_1, D_2, E_1, E_2)$, i.e., the rate, distortion, and error exponent levels at each layer. We also analyze the special case of successive refinability, where the triplets (R_1, D_1, E_1) and (R_2, D_2, E_2) individually achieve the nonscalable bounds. It is surprising that for any D_1, D_2, and R_1, there exists an $\hat{R}_2$ such that successive refinability is ensured for all $R_2 \geq \hat{R}_2$.**

Keywords— **Scalable source coding, error exponents, successive refinement.**

I. Introduction

The rate-distortion function $R(D)$ indicates the minimum rate required to (asymptotically) achieve an *average* distortion D. A more demanding rate-distortion problem arises from statistical consideration of the *error event*, i.e., the event that a source vector is compressed at distortion exceeding D. While the rate $R(D)$ is sufficient to ensure that the error probability vanishes as the block length n tends to infinity, a major concern is with its asymptotic rate of decay. The asymptotic decay is typically quantified by the error exponent $E = -\frac{1}{n}\log \Pr[\text{error}]$. Thus, the rate-distortion problem may be generalized to ask one of the two questions: (i) What is the minimum rate required to achieve an error exponent at or above a given level? (ii) What is the maximum error exponent achievable at or below a given coding rate? The standard rate-distortion problem corresponds to the special case of (i) with required error exponent $E \longrightarrow 0$.

The best error exponent for nonscalable source coding was first characterized by Marton [7]. Given a discrete memoryless source (DMS) with distribution P, and given distortion and rate levels D and R, respectively, the best error exponent is

$$E_P(D,R) = \inf_{P': R < R_{P'}(D)} \mathcal{D}(P'||P) , \qquad (1)$$

where $\mathcal{D}(P'||P)$ is the information divergence and $R_P(D)$ is the rate-distortion function for source P, which is given by

$$R_P(D) = \min_{Q(y|x): E_{P,Q}\{d(X,Y)\} \leq D} I_{P,Q}(X;Y) . \qquad (2)$$

* This work was supported in part by the NSF under grant no. MIP-9707764, EIA 9986057, the UC MICRO program, Conexant Systems, Inc., Dolby Laboratories, Inc., Lucent Technologies, Inc., Media Stream, Inc., and Qualcomm, Inc.

This result is valid under the condition $R \geq R_P(D)$. Marton also discussed the existence of discontinuities in $E_P(D,R)$ as a function of R, possibly for a countable set of rates. Sufficient conditions for continuity of $E_P(D,R)$ for all R were derived in [7] and [1].

In this paper, we extend these results to 2-layer scalable coding, and derive a characterization for $E_P(D_1, D_2, R_1, R_2, E_1)$, the best error exponent achievable in the second layer given the distortion and rate constraints for both layers, and the error exponent constraint for the first layer. Kanlis and Narayan [6] previously considered an extension of the nonscalable result, however they defined as error the event that either the first layer distortion exceeds D_1, or the second layer distortion exceeds D_2, precluding a possible trade-off analysis between the exponents of the two error probabilities. Haroutunian et al. [5] analyzed the special "successive refinability" case, i.e., the conditions under which

$$E_P(D_1, D_2, R_1, R_2, E_P(D_1, R_1)) = E_P(D_2, R_2)$$

is satisfied. We further use our characterization of the function $E_P(D_1, D_2, R_1, R_2, E_1)$ to analyze the special case of successive refinability, and prove a necessary and sufficient condition which is fundamentally different from the condition in [5]. In particular, it implies that for every D_1, D_2, and R_1, there exists a $\hat{R}_2$ such that successive refinability is ensured for all $R_2 \geq \hat{R}_2$.

We begin with some preliminaries in the following section. In Section 3, we employ the type covering lemmas [3], [6], to construct a coding strategy and in Section 4 we prove, by extending the approach of [7], that no better coding strategy exists. Finally, in Section 5, we analyze the special case of successive refinability.

II. Preliminaries and Basic Definitions

We denote the source and the reproduction alphabets by $\mathcal{X}$ and $\hat{\mathcal{X}}$, respectively. We assume a single-letter distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, \infty)$, i.e., the distortion measure extends to n dimensions as $d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^n d(x_i, \hat{x}_i)$. In the scalable coding scenario, for generality, we allow for two layer-specific single-letter distortion measures d_1 and d_2. We denote by $|f|$ the cardinality of the range of a function f.

Definition 1: (R, D, E) is called an achievable rate-distortion-exponent triplet if for any given $\delta > 0$, there exist an encoding function

$$f : \mathcal{X}^n \longrightarrow \{1, 2, \ldots, 2^{nR}\} ,$$

and a decoding function

$$g : \{1, 2, \ldots, 2^{nR}\} \longrightarrow \hat{\mathcal{X}}^n ,$$

such that

$$-\frac{1}{n} \log \Pr[d(X^n, g(f(X^n))) > D] > E - \delta$$

for all $n \geq n_0(\delta)$.

From (1), it is clear that for a source P, and for $E > 0$, (R, D, E) is achievable if and only if

$$R \geq R_P(D) \quad \text{and} \quad \inf_{P' : R < R_{P'}(D)} \mathcal{D}(P'||P) > E , \tag{3}$$

or in other words, if and only if

$$\forall P' : \mathcal{D}(P'||P) \leq E \Longrightarrow R \geq R_{P'}(D) . \tag{4}$$

Definition 2: The rate-distortion-exponent function, $R_P(D, E)$, is defined as the minimum R such that (R, D, E) is achievable for source P.

The condition (4) implies that

$$R_P(D, E) = \sup_{P' : \mathcal{D}(P'||P) \leq E} R_{P'}(D) . \tag{5}$$

In the special case of $E \longrightarrow 0$, (R, D, E) is achievable if and only if $R \geq R_P(D)$ is satisfied. Another extreme case is when $E \longrightarrow \infty$, for which (5) yields $R^0(D)$, i.e., the "zero-error" rate-distortion function [3, Theorem 2.4.2].

Definition 3: $(R_1, R_2, D_1, D_2, E_1, E_2)$ with $D_2 \leq D_1$ and $R_2 \geq R_1$ is called an achievable 2-stage rate-distortion-exponent 6-tuple if for any given $\delta_1 > 0$ and $\delta_2 > 0$, there exist stage-encoding functions

$$f_1 : \mathcal{X}^n \longrightarrow \{1, 2, \ldots, 2^{nR_1}\}$$
$$f_2 : \mathcal{X}^n \longrightarrow \{1, 2, \ldots, 2^{n[R_2 - R_1]}\}$$

and stage-decoding functions

$$g_1 : \{1, 2, \ldots, 2^{nR_1}\} \longrightarrow \hat{\mathcal{X}}^n$$
$$g_2 : \{1, 2, \ldots, 2^{nR_1}\} \times \{1, 2, \ldots, 2^{n[R_2 - R_1]}\} \longrightarrow \hat{\mathcal{X}}^n ,$$

such that

$$-\frac{1}{n} \log \Pr[d_1(X^n, g_1(f_1(X^n))) > D_1] > E_1 - \delta_1$$

and

$$-\frac{1}{n} \log \Pr[d_2(X^n, g_2(f_1(X^n), f_2(X^n))) > D_2] > E_2 - \delta_2$$

for all $n \geq n_0(\delta_1, \delta_2)$.

The special case $E_1, E_2 \longrightarrow 0$ corresponds to Rimoldi's successive refinement characterization [8], as the 6-tuple $(R_1, R_2, D_1, D_2, E_1, E_2)$ becomes achievable if and only if $R_1 \geq R_P(D_1)$ and $R_2 \geq R_P(D_1, D_2, R_1)$, where $R_P(D_1, D_2, R_1)$ is given by

$$R_P(D_1, D_2, R_1) = \min_{\substack{Q(y,z|x) \,:\, I_{P,Q}(X;Y) \leq R_1 \\ E_{P,Q}\{d_1(X,Y)\} \leq D_1 \\ E_{P,Q}\{d_2(X,Z)\} \leq D_2}} I_{P,Q}(X; Y, Z) \tag{6}$$

if $R_1 \leq R_P(D_2)$, and

$$R_P(D_1, D_2, R_1) = R_1 \tag{7}$$

otherwise[1].

Definition 4: Given source P, distortion levels $D_1 \geq D_2$, and rate $R_1 \geq R_P(D_1, E_1)$, the *scalable rate-distortion-exponent function* $R_P(D_1, D_2, R_1, E_1, E_2)$ is defined as the minimum R_2 such that $(R_1, R_2, D_1, D_2, E_1, E_2)$ is achievable.

Definition 5: Similarly, the *scalable error exponent function* $E_P(D_1, D_2, R_1, R_2, E_1)$ for source P, defined under the conditions $D_1 \geq D_2$, $R_1 \geq R_P(D_1, E_1)$, and $R_2 \geq R_P(D_1, D_2, R_1)$, is the minimum E_2 such that $(R_1, R_2, D_1, D_2, E_1, E_2)$ is achievable.

III. Sufficient Conditions for Achievability

We derive sufficient conditions for achievability by constructing an actual scalable coding strategy. To this end, we employ the type covering lemma [3], and its scalable extension proved by Kanlis and Narayan [6]. The strategy exploits a fundamental property of types: the number of distinct types for sequences of length n grows at most polynomially with n. Hence, we may tailor encoding functions to each type separately, without compromising the overall coding rate asymptotically [3].

We denote by T_P^n the set of all source vectors x^n having type P. We separately analyze the two cases $E_1 < E_2$ and $E_1 \geq E_2$. Note that in order for $(R_1, R_2, D_1, D_2, E_1, E_2)$ to be achievable, it is necessary to satisfy condition (4) for the first layer:

$$\forall P' : \mathcal{D}(P'||P) \leq E_1 \Longrightarrow R_1 \geq R_{P'}(D_1) . \tag{8}$$

Case I: $E_1 < E_2$.
We adopt the following strategy: For type P':

- If $\mathcal{D}(P'||P) \leq E_1$, then we can generate 2^{nR_1} balls of radius D_1 in the first layer, and for each D_1-ball, generate $2^{n[R_{P'}(D_1, D_2, R_1) - R_1]}$ balls of radius D_2 in the second layer, such that for every source vector $x^n \in T_{P'}^n$, there exists a pair of D_1- and D_2-balls covering x^n. Since $R_1 \geq R_{P'}(D_1)$ from (8), this 2-layer covering is indeed possible for large n, as proved in [6].
- If $E_1 < \mathcal{D}(P'||P) \leq E_2$, then we cover $T_{P'}^n$ with $2^{nR_{P'}(D_2)}$ balls, only for the purpose of using the ball centers as second layer codevectors. In the first stage, we send the first R_1 bits, and do not reproduce anything at the decoder, and in the second stage send the rest of the bits (if any), and reproduce the ball centers at the decoder.
- If $E_2 < \mathcal{D}(P'||P)$, then we do not perform any covering.

[1] Observe that if $R_1 > R_P(D_2)$, the minimum in (6) is $R_P(D_2)$, which makes R_1 greater than the achieved minimum. On the other hand, if $R_1 \leq R_P(D_2)$, then the minimum in (6) is always greater than or equal to $R_P(D_2)$.

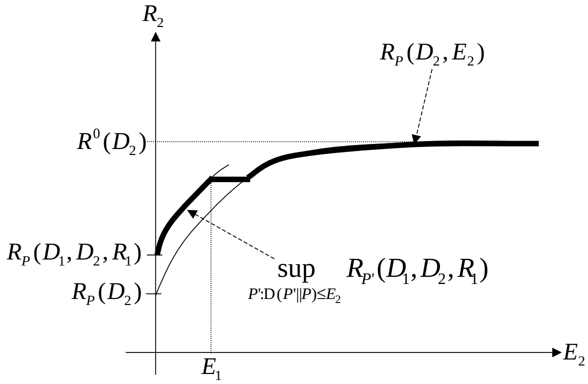

Fig. 1. A typical curve of achievable R_2 versus E_2, given fixed D_1, D_2, R_1, and E_1, where $R_1 \geq R_P(D_1, E_1)$.

It is clear that the first and second-layer error exponents are at least E_1, and E_2, respectively. The achieved rate at the second layer is

$$R_2 = \max \left\{ \sup_{P': E_1 < \mathcal{D}(P'||P) \leq E_2} R_{P'}(D_2), \sup_{P': \mathcal{D}(P'||P) \leq E_1} R_{P'}(D_1, D_2, R_1) \right\} . \quad (9)$$

or in a more compact form

$$R_2 = \max \left\{ R_P(D_2, E_2), \sup_{P': \mathcal{D}(P'||P) \leq E_1} R_{P'}(D_1, D_2, R_1) \right\} \quad (10)$$

since $R_{P'}(D_1, D_2, R_1) \geq R_{P'}(D_2)$ for all P'.

Case II: $E_1 \geq E_2$.

We adopt the following strategy: For type P':

- If $\mathcal{D}(P'||P) \leq E_2$, then similarly to the first case, we perform a two-layer covering of type P' using 2^{nR_1} balls of radius D_1 in the first layer, and $2^{nR_{P'}(D_1,D_2,R_1)}$ balls of radius D_2 in the second layer. Since $\mathcal{D}(P'||P) \leq E_2 \leq E_1$, it follows from (8) that $R_1 \geq R_{P'}(D_1)$, and hence this 2-layer covering is again possible.
- If $E_2 < \mathcal{D}(P'||P) \leq E_1$, then we cover $T^n_{P'}$ with $2^{nR_{P'}(D_1)}$ balls, only for the purpose of using the ball centers as first layer codevectors.
- If $E_1 < \mathcal{D}(P'||P)$, then we do not perform any covering.

It is easily verified that the first and second-layer error exponents are at least E_1, and E_2, respectively. The achieved rate at the second layer is given by

$$R_2 = \sup_{P': \mathcal{D}(P'||P) \leq E_2} R_{P'}(D_1, D_2, R_1) . \quad (11)$$

Combining (9) and (11), we observe that if $R_2 \geq R^*(D_1, D_2, R_1, E_1, E_2)$, where $R^*(D_1, D_2, R_1, E_1, E_2)$ is given in (12) at the bottom, then $(R_1, R_2, D_1, D_2, E_1, E_2)$ is achievable. However, for the purpose of the proof that $R^*(D_1, D_2, R_1, E_1, E_2)$ actually specifies the entire achievable region, it will be more convenient to work with infimum of $\mathcal{D}(P'||P)$ over certain sets. The corresponding sufficient condition for achievability is given by

$$E_2 \leq E^*(D_1, D_2, R_1, R_2, E_1) ,$$

where $E^*(D_1, D_2, R_1, R_2, E_1)$ is defined as in (13) with the standard convention that infimum over an empty set yields infinity. A fairly general curve of the achieved R_2 with this coding strategy is shown in Figure 1.

IV. Necessary Conditions for Achievability

For any coding strategy given by encoding and decoding functions f_1, f_2, g_1, and g_2, as described in Definition 3, we introduce the notation

$$\mathcal{U}_1(f_1, g_1) = \{x^n : d_1(x^n, g_1(f_1(x^n))) > D_1\} ,$$

for the set of points in $\mathcal{X}^n$ that are not reproduced within distortion D_1 at the first stage. Similarly, the set of points that are not reproduced within distortion D_2 at the second stage is denoted by

$$\mathcal{U}_2(f_1, f_2, g_2) = \{x^n : d_2(x^n, g_2(f_1(x^n), f_2(x^n))) > D_2\} .$$

Theorem 1: Given a discrete source with probability distribution P, let $R_1 \geq R_P(D_1, E_1)$. A coding strategy given by f_1, f_2, g_1, and g_2 satisfies

$$\begin{aligned} \frac{1}{n} \log |f_1| &< R_1 \\ \frac{1}{n} \log |f_1 \times f_2| &< R_2 \\ -\frac{1}{n} \log P^n(\mathcal{U}_1(f_1, g_1)) &> E_1 - \delta_1 \end{aligned}$$

for any given $\delta_1 > 0$, and for all $n \geq n_0(\delta_1)$ only if

$$\limsup_{n\to\infty} \left\{ -\frac{1}{n} \log P^n(\mathcal{U}_2(f_1, f_2, g_2)) \right\} \leq E^*(D_1, D_2, R_1, R_2, E_1 - \delta_1) . \quad (14)$$

$$R^*(D_1, D_2, R_1, E_1, E_2) \triangleq \max \left\{ R_P(D_2, E_2), \sup_{P': \mathcal{D}(P'||P) \leq \min(E_1, E_2)} R_{P'}(D_1, D_2, R_1) \right\} \quad (12)$$

$$E^*(D_1, D_2, R_1, R_2, E_1) \triangleq \min \left\{ E_P(D_2, R_2), \inf_{\substack{P' : \mathcal{D}(P'||P) \leq E_1 \text{ and} \\ R_2 < R_{P'}(D_1, D_2, R_1)}} \mathcal{D}(P'||P) \right\} \quad (13)$$

Remark: It follows from the theorem that the achievable region constructed in the previous section is the largest possible achievable region. In other words,

$$E_P(D_1,D_2,R_1,R_2,E_1) = E^*(D_1,D_2,R_1,R_2,E_1)$$
$$R_P(D_1,D_2,R_1,E_1,E_2) = R^*(D_1,D_2,R_1,E_1,E_2)\ .$$

Sketch of Proof: First, we define three sets

$$\begin{aligned}\mathcal{Q}_1 &= \{Q : \mathcal{D}(Q||P) < E_1 - \delta_1\} \\ \mathcal{Q}_2 &= \{Q : R_2 < R_Q(D_1,D_2,R_1)\} \\ \mathcal{Q}_3 &= \{Q : R_2 < R_Q(D_2)\}\ .\end{aligned}$$

We observe that for any $Q \in \mathcal{Q}_1$, $Q^n(\mathcal{U}_1(f_1,g_1)) \longrightarrow 0$ as $n \longrightarrow \infty$. Using this observation, we prove that for all $Q \in \mathcal{Q}_3 \cup (\mathcal{Q}_1 \cap \mathcal{Q}_2)$, there exists a constant $\alpha(Q,D_1,D_2,R_1,R_2,E_1) > 0$ such that

$$Q^n(\mathcal{U}_2(f_1,f_2,g_2)) \geq \alpha(Q,D_1,D_2,R_1,R_2,E_1)\ . \tag{15}$$

Next, we define

$$G^n \triangleq \left\{x^n : \left|\frac{1}{n}\log\frac{Q^n(x^n)}{P^n(x^n)} - \mathcal{D}(Q||P)\right| < \eta\right\}\ .$$

The weak law of large numbers ensures that

$$Q^n(G^n) > 1 - \frac{1}{2}\alpha(Q,D_1,D_2,R_1,R_2,E_1)\ , \tag{16}$$

for all $Q \in \mathcal{Q}_3 \cup (\mathcal{Q}_1 \cap \mathcal{Q}_2)$, for sufficiently large n. We next consider the error probability at the second layer:

$$\begin{aligned}&P^n(\mathcal{U}_2(f_1,f_2,g_2)) \\ &\geq P^n(\mathcal{U}_2(f_1,f_2,g_2) \cap G^n) \\ &= \sum_{x^n \in \mathcal{U}_2(f_1,f_2,g_2)\cap G^n} P^n(x^n) \\ &= \sum_{x^n \in \mathcal{U}_2(f_1,f_2,g_2)\cap G^n} Q^n(x^n)\exp\left\{-\log\frac{Q^n(x^n)}{P^n(x^n)}\right\} \\ &\geq Q^n(\mathcal{U}_2(f_1,f_2,g_2)\cap G^n)\exp\{-n[\mathcal{D}(Q||P)+\eta]\} \\ &\geq \frac{1}{2}\alpha(Q,D_1,D_2,R_1,R_2,E_1)\exp\{-n[\mathcal{D}(Q||P)+\eta]\}\end{aligned}$$

for sufficiently large n, where the last inequality follows from (15) and (16). This implies that

$$\limsup_{n\to\infty}\left\{-\frac{1}{n}\log P^n(\mathcal{U}_2(f_1,f_2,g_2))\right\} \leq \mathcal{D}(Q||P) + \eta$$

for all $Q \in \mathcal{Q}$ and all η. The result follows after taking the infimum of both sides over the set $\mathcal{Q}$, and letting $\eta \longrightarrow 0$.

V. Successive Refinability

Let $\hat{R}_2$ be defined as

$$\hat{R}_2 = \sup_{P':\mathcal{D}(P'||P)\leq E_P(D_1,R_1)} R_{P'}(D_1,D_2,R_1)\ . \tag{17}$$

Pictorially, $\hat{R}_2$ corresponds to the straight line in Figure 1 when E_1 is set to $E_P(D_1,R_1)$. It follows from (13) that successive refinement is achievable, i.e.,

$$E_P(D_1,D_2,R_1,R_2,E_P(D_1,R_1)) = E_P(D_2,R_2)$$

at all second layer rates $R_2 \geq \hat{R}_2$. If $R_2 < \hat{R}_2$, then by (13), successive refinability requires

$$\inf_{\substack{P' : \mathcal{D}(P'||P) \leq E_P(D_1,R_1) \text{ and} \\ R_2 < R_{P'}(D_1,D_2,R_1)}} \mathcal{D}(P'||P) = E_P(D_2,R_2)\ . \tag{18}$$

The successive refinability condition for the case $R_2 < \hat{R}_2$ may be restated equivalently as

$$E_P(D_1,R_1) \geq E_P(D_2,R_2) \tag{19}$$
$$R_{P^*}(D_1,D_2,R_1) = R_{P^*}(D_2)\ , \tag{20}$$

where distribution P^* achieves the infimum in (18). Note that for the special case $R_1 = R_P(D_1)$ and $R_2 = R_P(D_2)$, it follows that $E_P(D_1,R_1) = E_P(D_2,R_2) = 0$ and the above conditions for successive refinability reduce to

$$R_P(D_2) = R_P(D_1,D_2,R_P(D_1))\ ,$$

for which the necessary and sufficient condition is the well-known Markovian condition given in [4].

VI. Conclusion

We characterized the region of all achievable 6-tuples $(R_1,R_2,D_1,D_2,E_1,E_2)$ for the scalable source coding scenario. Given source P, the characterization is in terms of the information divergence $\mathcal{D}(P'||P)$ and the rate distortion functions $R_{P'}(D_2)$ and $R_{P'}(D_1,D_2,R_1)$, for all sources P'. We specialized the necessary and sufficient achievability conditions to the successive refinability case, and obtained the surprising result that it is possible to achieve the bounds $E_1 = E_P(D_1,R_1)$, and $E_2 = E_P(D_2,R_2)$, for all second layer rates R_2 above a specified threshold.

References

[1] R. Ahlswede. Extremal properties of rate-distortion functions. *IEEE Trans. on Information Theory* 36(1):166–171, January 1990.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[3] I. Csiszár. *Information Theory. Coding Theorems for Discrete Memoryless Sources.* Akadémiai Kiadó, Budapest, Hungary, 1981.

[4] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Trans. on Information Theory* 37(2):269–275, March 1991.

[5] E. A. Haroutunian and A. N. Harutyunyan. Successive refinement of information with reliability criterion *IEEE International Symposium on Information Theory*, page 205, Sorrento, Italy, June 2000.

[6] A. Kanlis and P. Narayan. Error exponents for successive refinement by partitioning. *IEEE Trans. on Information Theory*, 42(1):275–282, January 1996.

[7] K. Marton. Error exponent for source coding with a fidelity criterion. *IEEE Trans. on Information Theory*, 20(2):197–199, March 1974.

[8] B. Rimoldi. Successive refinement of information: characterization of the achievable rates. *IEEE Trans. on Information Theory*, 40(1):253–259, January 1994.