1) Pick a subinterval $\mathcal{C}_1$ (e.g., $\mathcal{C}_1 = [79, 86]$)
2) Let (cl_1, cl_2, cl_3) run through the interval $\mathcal{C}_1$. For each frame $j = 1, \ldots, m$ calculate

$$p^j_{(cl_1,cl_2,cl_3)} = \sum_{v=\mathcal{I}} P((cl_1, cl_2, cl_3) \text{ in } v\text{:th pos.}) \cdot [O^j_{(cl_1,cl_2,cl_3,v-100)} = 0] + 1/2 \cdot (1 - \sum_{v=\mathcal{I}} P((cl_1, cl_2, cl_3) \text{ in } v\text{:th pos.}))$$

Calculate the log-likelihood ratio of the weighted probability over all frames

$$\Lambda_{(cl_1,cl_2,cl_3)} = \sum_{i=1}^{m} \ln \frac{p^j_{(cl_1,cl_2,cl_3)}}{1 - p^j_{(cl_1,cl_2,cl_3)}}$$

Estimate the linear combination

$$s^1_{cl_1} + s^2_{cl_2} + s^3_{cl_3} = \mathrm{HD}(\Lambda_{(cl_1,cl_2,cl_3)})$$

using a hard decision (HD) on the value of $\Lambda_{(cl_1,cl_2,cl_3)}$.
3) Decode the generated linear code

$$s^1_{cl_1} + s^2_{cl_2} + s^3_{cl_3} = \mathrm{HD}(\Lambda_{(cl_1,cl_2,cl_3)})$$

for (cl_1, cl_2, cl_3) in interval $\mathcal{C}_1$ using a ML decoding through exhaustive search. Save the T closest solutions.
4) Repeat steps 1 to 3 for each new subinterval $\mathcal{C}_2, \mathcal{C}_3, \ldots$ until totally 64 bits of the shift register sequences are recovered.
5) Combine the solutions from each subinterval and check the validness of the solutions.

Fig. 3. A summary of the proposed attack.

VII. Conclusion

We have proposed a new attack on the A5/1 stream cipher, based on an identified correlation. In contrast to previous attacks, this is not a time–memory tradeoff attack, but uses completely different properties of the cipher. It explores the weak key initialization which allows to separate the session key from the frame number in binary linear expressions.

The complexity of the attack is only linear in the length of the shift registers and depends instead on the number of irregular clockings before the keystream is produced. The implemented attack needs the 40 first bits from about 2^{16} (possible nonconsecutive) frames, which corresponds to about 5 min of GSM conversation. Our implementation of the attack shows that we have a high success rate; more than 70%. This can be improved by using larger list size and/or larger interval size. The complexity of the attack using the parameters presented here is quite low and the attack can be carried out on a modern PC in less than 5 min using very little precomputation time and memory.

The improvements compared to previous work are the following. All previous attacks have a complexity exponential in the shift-register length. The complexity of the attack presented in this correspondence is roughly linear in the shift-register lengths.

Previous attacks also need either much precomputation and/or memory or they have a high time complexity. The proposed attack is simple to implement, has been implemented, and completes its task in less than 5 min.

Finally, the presented attack also enlightens new interesting design weaknesses in A5/1 that should be considered when constructing new stream ciphers.

References

[1] E. Biham and O. Dunkelman, "Cryptanalysis of the A5/1 GSM stream cipher," in *Indocrypt 2000 (Lecture Notes in Computer Science)*, vol. 1977, 2000, pp. 43–51.

[2] A. Biryukov, A. Shamir, and D. Wagner, "Real time cryptanalysis of A5/1 on a PC," in *FSE'2000 (Lecture Notes in Computer Science)*, vol. 1978, 2001, pp. 1–18.

[3] M. Briceno, I. Goldberg, and D. Wagner. (1999, May) A pedagogical implementation of A5/1. [Online]. Available: http://scard.org

[4] J. Golic, "Cryptanalysis of alleged A5 stream cipher," in *Eurocrypt'97 (Lecture Notes in Computer Science)*, vol. 1233, 1997, pp. 239–255.

[5] T. Johansson and F. Jönsson, "Improved fast correlation attacks on stream ciphers via convolutional codes," in *Eurocrypt'99 (Lecture Notes in Computer Science)*, vol. 1592, 1999, pp. 347–362.

[6] M. Krause, "BDD-based cryptanalysis of keystream generators," presented at the EUROCRYPT 2002, [Online] Available: http://www.iacr.org.

[7] W. Meier and O. Staffelbach, "Fast correlation attacks on certain stream ciphers," *J. Cryptol.*, vol. 1, pp. 159–176, 1989.

[8] A. Menezes, P. van Oorschot, and S. Vanstone, *Handbook of Applied Cryptography*. Boca Raton, FL: CRC, 1997.

Error Exponents in Scalable Source Coding

Ertem Tuncel, *Student Member, IEEE,* and Kenneth Rose, *Fellow, IEEE*

***Abstract*—The characterization of the set of achievable rate and distortion values for scalable source coding is extended to additionally account for error exponents, namely, the negative normalized asymptotic log likelihood of error events at different layers. The "error" at each layer is defined as the event that the source block is not reproduced within the prespecified fidelity at the corresponding decoder. We consider separate error events at each layer so as to allow a tradeoff analysis for the error exponents when the rate and distortion values are fixed. For two-step coding of discrete memoryless sources, we derive a single-letter characterization of the region of all achievable 6-tuples $(R_1, R_2, E_1, E_2, D_1, D_2)$, i.e., the rate, error exponent, and distortion levels at each layer. We also analyze the special case of successive refinability, where (R_1, E_1, D_1) and (R_2, E_2, D_2) individually achieve the nonscalable bounds. A surprising outcome of the analysis is that for any D_1, D_2, and E_1, there exists a finite threshold $\hat{E}_2 \geq E_1$ such that successive refinability is ensured for all $E_2 \geq \hat{E}_2$.**

***Index Terms*—Error exponents, large deviations, reliability, scalable source coding, successive refinement.**

Manuscript received March 20, 2001; revised May 30, 2002. This work was supported in part by the National Science Foundation under Grants EIA-9986057 and EIA-0080134, the University of California MICRO Program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, Inc., and Qualcomm, Inc. The material in this correspondence was presented in part at the Canadian Workshop on Information Theory, Vancouver, BC, Canada, June 2001.

The authors are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560 USA (e-mail: ertem@ece.ucsb.edu; rose@ece.ucsb.edu).

Communicated by P. Narayan, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2002.806142

I. Introduction

Successive refinement of information, or scalable source coding, has received increasing attention in recent years, mainly due to the growing prevalence of heterogeneous communication networks such as the Internet. This setting involves the embedding of descriptions, ranging from coarse to fine, into a single bit stream so that the signal can be reproduced at higher quality as a larger portion of the bit stream is being accessed.

The early results on rate-distortion theoretic analysis of scalable coding is due to Koshelev [7], [8], and later to Equitz and Cover [4]. Most early work was mainly concerned with successive refinability of the source, i.e., the conditions under which it is possible to perform scalable coding without compromising the rate-distortion performance. The set of all achievable rates and distortions in scalable coding has been derived independently by Koshelev [7] and Rimoldi [10]. Work by Kanlis and Narayan [6] and Haroutunian and Harutyunyan [5] offered extensions to account for rate, distortion, and *error exponent* (or *reliability* as referred to in [5]). However, an exact characterization of the entire region of achievable rates, distortions, and error exponents has not been fully derived so far and is the subject of this correspondence.

The rate-distortion function $R_P(D)$ for the memoryless source P 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 $R_P(D)$ 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 maximum error exponent for nonscalable source coding was first characterized by Marton [9]. Given a discrete memoryless source (DMS) with distribution P, and given distortion and rate levels D and R, respectively, the best error exponent, denoted by $E_P(D, R)$, is characterized in terms of the information divergence $\mathcal{D}(Q\|P)$ and the rate-distortion functions $R_Q(D)$, for all sources Q. Considering the best error exponent as function of rate, Marton also discussed the existence of (possibly infinite but countable number of) discontinuities. Sufficient conditions for continuity of the maximum error exponent for all rates were derived in [9] and [1].

In this correspondence, we derive a single-letter characterization of $E_P(D_1, D_2, R_1, R_2, E_1)$, 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. We also provide an equivalent characterization in terms of the minimum second layer rate as a function of the other parameters, $R_P(D_1, D_2, E_1, E_2, R_1)$, which is an extension of the rate-reliability-distortion function $R_P(D_1, E_1)$ of [5]. Kanlis and Narayan [6] previously considered an extension of the nonscalable error exponent result of Marton, however, they mainly analyzed the case where the error exponent at the first layer coincides with $E_P(D_1, R_1)$, precluding a possible tradeoff analysis between the error exponents at separate layers. Haroutunian and Harutyunyan [5] analyzed the special "successive refinability" case, i.e., the conditions under which

$$R_P(D_1, D_2, E_1, E_2, R_P(D_1, E_1)) = R_P(D_2, E_2)$$

is satisfied. We further use our characterization of the function $R_P(D_1, D_2, E_1, E_2, R_1)$ to analyze the above equality, and prove a necessary and sufficient condition for successive refinability, which is fundamentally different from the condition provided in [5]. In particular, it implies that for every D_1, D_2, and E_1, there exists an $\hat{E}_2 \geq E_1$ such that successive refinability is ensured for all $E_2 \geq \hat{E}_2$.

We begin with some preliminaries in the following section. In Section III, we employ the type covering lemmas [3], [6], to construct a coding strategy and in Section IV we prove, by extending the approach of [9], that no better coding strategy exists. Finally, in Section V, we analyze the special case of successive refinability.

II. Preliminaries

Let $\{X_t\}_{t=1}^{\infty}$ be a sequence of independent and identically distributed (i.i.d.) random variables taking values from the finite source alphabet $\mathcal{X}$, with probability mass function (pmf) P. We assume, without loss of generality, that $P(x) > 0$ for all $x \in \mathcal{X}$. Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ denote the first- and the second-layer finite reproduction alphabets, respectively. We assume, for both layers $i = 1, 2$, single-letter distortion measures $d_i\colon \mathcal{X} \times \mathcal{Y}_i \longrightarrow [0, \infty)$, i.e., d_i extends to n dimensions as

$$d_i(x^n, y^n) = \frac{1}{n}\sum_{t=1}^{n} d_i(x_t, y_t).$$

In nonscalable source coding, a pair (R_1, D_1) is *achievable* if for every $\epsilon > 0, \delta > 0$, there exist a sequence of block-encoding functions $f_1^{(n)}\colon \mathcal{X}^n \longrightarrow \mathcal{M}_1^{(n)}$ and a sequence of block-decoding functions $g_1^{(n)}\colon \mathcal{M}_1^{(n)} \longrightarrow \mathcal{Y}_1^n$, such that

$$\frac{1}{n}\log\left|\mathcal{M}_1^{(n)}\right| \leq R_1 + \delta$$

and

$$\Pr\left[d_1\left(X^n, g_1^{(n)}\left(f_1^{(n)}(X^n)\right)\right) \leq D_1\right] \geq 1 - \epsilon$$

for large enough n. The well-known rate-distortion function $R_P(D_1)$, given by

$$R_P(D_1) = \min_{W \in \mathcal{W}_P(D_1)} \mathbf{I}_{PW}(X; Y_1), \tag{1}$$

where

$$\mathcal{W}_P(D_1) = \{W(y_1|x)\colon \mathbf{E}_{PW}\{d_1(X, Y_1)\} \leq D_1\}$$

characterizes the minimum achievable rate for distortion D_1. Here, $\mathbf{E}$ and $\mathbf{I}$, respectively, denote expectation and mutual information, and $PW(x, y_1) = P(x)W(y_1|x)$ is the joint pmf for random variables X and Y_1.

In scalable source coding, achievability of a quadruple (R_1, R_2, D_1, D_2) is considered. Adopting a slightly modified version of Rimoldi's definition [10], we say that quadruple (R_1, R_2, D_1, D_2) with $R_2 \geq R_1$ is *successively achievable* if for every $\epsilon > 0, \delta > 0$, there exist a sequence of block-encoding functions $f_i^{(n)}\colon \mathcal{X}^n \longrightarrow \mathcal{M}_i^{(n)}$ for $i = 1, 2$, and a sequence of block-decoding functions $g_1^{(n)}\colon \mathcal{M}_1^{(n)} \longrightarrow \mathcal{Y}_1^n$ and $g_2^{(n)}\colon \mathcal{M}_1^{(n)} \times \mathcal{M}_2^{(n)} \longrightarrow \mathcal{Y}_2^n$, such that

$$\frac{1}{n}\log\left|\mathcal{M}_1^{(n)}\right| \leq R_1 + \delta$$
$$\frac{1}{n}\log\left|\mathcal{M}_2^{(n)}\right| \leq R_2 - R_1 + \delta$$

and

$$\Pr\Big[d_1\left(X^n, g_1^{(n)}\left(f_1^{(n)}(X^n)\right)\right) \leq D_1,$$
$$d_2\left(X^n, g_2^{(n)}\left(f_1^{(n)}(X^n), f_2^{(n)}(X^n)\right)\right) \leq D_2\Big] \geq 1 - \epsilon$$

for large enough n.[1] A single-letter characterization of the set of achievable quadruples has been provided independently by Koshelev [7] and Rimoldi [10]. Kanlis and Narayan [6] stated as a corollary to Rimoldi's theorem that (R_1, R_2, D_1, D_2) is 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[2]

$$R_P(D_1, D_2, R_1) = \max\left\{R_1, \min_{W \in \mathcal{W}_P(D_1, D_2, R_1)} \mathbf{I}_{PW}(X; Y_1, Y_2)\right\} \quad (2)$$

with

$$\begin{aligned}\mathcal{W}_P(D_1, D_2, R_1) = \{&W(y_1, y_2|x): \\ &\mathbf{E}_{PW}\{d_1(X, Y_1)\} \leq D_1, \\ &\mathbf{E}_{PW}\{d_2(X, Y_2)\} \leq D_2, \\ &\mathbf{I}_{PW}(X; Y_1) \leq R_1\}.\end{aligned}$$

It obviously follows from (2) that $R_P(D_1, D_2, R_1)$ is nonincreasing in the distortion arguments. Moreover, since $\mathbf{I}_{PW}(X; Y_1)$, $\mathbf{I}_{PW}(X; Y_1, Y_2)$, $\mathbf{E}_{PW}\{d_1(X, Y_1)\}$, and $\mathbf{E}_{PW}\{d_2(X, Y_2)\}$, are all convex functions of $W(y_1, y_2|x)$, it also follows that $R_P(D_1, D_2, R_1)$ is a convex (and hence continuous) function in all its arguments. (The standard proof in, e.g., [2, Lemma 13.4.1], can easily be extended for $R_P(D_1, D_2, R_1)$.) We will need this important fact in the proof of Lemma 3 in Section IV.

More demanding rate-distortion problems arise from the consideration of the asymptotic behavior of the probability of the error event. The asymptotic decay of the probability of the error event is typically quantified by the *error exponent* $E = -\frac{1}{n}\log \Pr[\text{error}]$. This quantity is also referred to as the *reliability*, and we shall use these terms interchangeably.

Definition 1: (R_1, E_1, D_1) with $E_1 > 0$ is an *achievable* rate–reliability–distortion triple if for any given $\epsilon > 0$ and $\delta > 0$ there exist a sequence of block-encoding functions $f_1^{(n)}: \mathcal{X}^n \longrightarrow \mathcal{M}_1^{(n)}$ and a sequence of block-decoding functions $g_1^{(n)}: \mathcal{M}_1^{(n)} \longrightarrow \mathcal{Y}_1^n$, such that

$$\frac{1}{n}\log\left|\mathcal{M}_1^{(n)}\right| \leq R_1 + \delta$$

and

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

for large enough n.

The maximum error exponent for given rate and distortion values were characterized by Marton [9] (cf. also Csiszár and Körner [3]).

Theorem 1—Marton [9]: (R_1, E_1, D_1) is achievable if and only if $0 < E_1 \leq E_P(D_1, R_1)$, where

$$E_P(D_1, R_1) = \inf_{Q:\, R_Q(D_1) > R_1} \mathcal{D}(Q\|P) \quad (3)$$

and where $\mathcal{D}(Q\|P)$ is the Kullback–Leibler (information) divergence between Q and P.

Remarks:

1) An obvious necessary condition for the achievability of (R_1, E_1, D_1) is $R_1 \geq R_P(D_1)$, because otherwise $E_P(D_1, R_1) = 0$ as is immediately seen by setting $Q = P$ in (3). More generally, the achievability of (R_1, E_1, D_1) implies that

$$\mathcal{D}(Q\|P) \leq E_1 \Longrightarrow R_1 \geq R_Q(D_1), \qquad \forall Q. \quad (4)$$

An illustration of (4) is provided in Fig. 1.

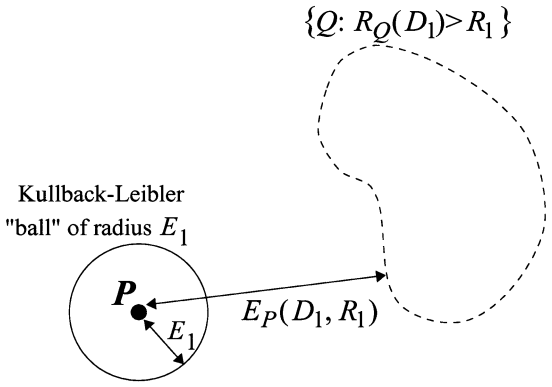

Fig. 1. If $(\boldsymbol{R_1, E_1, D_1})$ is achievable, i.e., $\boldsymbol{E_1 \leq E_P(D_1, R_1)}$, then the Kullback–Leibler ball $\boldsymbol{\{Q: \mathcal{D}(Q\|P) \leq E_1\}}$ and the set $\boldsymbol{\{Q: R_Q(D_1) > R_1\}}$ are disjoint.

2) An equivalent characterization of achievable (R_1, E_1, D_1) is via the rate–reliability–distortion function $R_P(D_1, E_1)$, which is defined in [5] as the minimum R_1 such that (R_1, E_1, D_1) is achievable

$$R_P(D_1, E_1) = \sup_{Q:\, \mathcal{D}(Q\|P) \leq E_1} R_Q(D_1). \quad (5)$$

3) As the error exponent becomes arbitrarily small, i.e., as $E_1 \to 0$, the minimum achievable R_1 tends to $R_P(D_1)$ from above. Another extreme case is when $E_1 \to \infty$. By inspection of (3), or equivalently (5), it follows that $R_1 \geq R^0(D_1)$ must be satisfied to ensure that (R_1, ∞, D_1) is achievable, where

$$R^0(D_1) = \max_Q R_Q(D_1) \quad (6)$$

is the "zero-error"' rate-distortion function [3, Theorem 2.4.2].

4) It should be noted that the coding strategy that achieves $(R_1, E_P(D_1, R_1), D_1)$ is *universal* in that the same construction achieves $E_P(D_1, R_1)$, with the given rate budget R_1 and the distortion constraint D_1, for all sources P. Thus, the coding method does not make use of prior knowledge about P. (See [9] and [3, Theorem 2.4.5].)

The generalization of the rate–reliability–distortion analysis to scalable coding was first addressed by Kanlis and Narayan [6]. They mainly analyzed the case where the error exponent at the first layer coincides with $E_P(D_1, R_1)$, precluding a possible tradeoff analysis between the error exponents at separate layers. Haroutunian *et al.* [5], on the other hand, only considered the conditions for successive refinability, i.e., whether or not a scalable coder can achieve Marton's error exponent function $E_P(D_i, R_i)$ at both layers $i = 1, 2$. In this work, we characterize the entire set of successively achievable 6-tuples $(R_1, R_2, E_1, E_2, D_1, D_2)$.

Definition 2: $(R_1, R_2, E_1, E_2, D_1, D_2)$ with $E_1, E_2 > 0$ and $R_2 \geq R_1$ is a *successively achievable* rate–reliability–distortion 6-tuple if for any given $\epsilon > 0$ and $\delta > 0$, there exist a sequence of block-encoding functions $f_i^{(n)}: \mathcal{X}^n \longrightarrow \mathcal{M}_i^{(n)}$ for $i = 1, 2$, and

[1] Rimoldi's original definition does not enforce $\boldsymbol{R_2 \geq R_1}$, and requires $\frac{1}{n}\log|\boldsymbol{\mathcal{M}_1^{(n)}}|\,|\boldsymbol{\mathcal{M}_2^{(n)}}| \leq \boldsymbol{R_2 + \delta}$, instead of $\frac{1}{n}\log|\boldsymbol{\mathcal{M}_2^{(n)}}| \leq \boldsymbol{R_2 - R_1 + \delta}$.

[2] The reason for adding the external maximization to the original version of [6] is that $\boldsymbol{R_2 \geq R_1}$ must be satisfied. Observe that if $\boldsymbol{R_1 > R_P(D_2)}$, the minimum in (2) is $\boldsymbol{R_P(D_2)}$, which makes $\boldsymbol{R_1}$ greater than the achieved minimum. On the other hand, if $\boldsymbol{R_1 \leq R_P(D_2)}$, then the minimum in (2) is always greater than or equal to $\boldsymbol{R_P(D_2)}$.

a sequence of block-decoding functions $g_1^{(n)}\colon\ \mathcal{M}_1^{(n)} \longrightarrow \mathcal{Y}_1^n$ and $g_2^{(n)}\colon \mathcal{M}_1^{(n)} \times \mathcal{M}_2^{(n)} \longrightarrow \mathcal{Y}_2^n$, such that

$$\frac{1}{n}\log\left|\mathcal{M}_1^{(n)}\right| \le R_1 + \delta$$
$$\frac{1}{n}\log\left|\mathcal{M}_2^{(n)}\right| \le R_2 - R_1 + \delta$$

and

$$-\frac{1}{n}\log \Pr\left[d_1\left(X^n,\, g_1^{(n)}\left(f_1^{(n)}(X^n)\right)\right) > D_1\right] \ge E_1 - \epsilon$$
$$-\frac{1}{n}\log \Pr\left[d_2\left(X^n,\, g_2^{(n)}\left(f_1^{(n)}(X^n),\, f_2^{(n)}(X^n)\right)\right) > D_2\right] \ge E_2 - \epsilon$$

for large enough n.

Remark: The special case $E_1, E_2 \to 0$ corresponds to Rimoldi's successive refinement characterization [10], i.e., $(R_1, R_2, E_1, E_2, D_1, D_2)$ with $E_1, E_2 \to 0$ is successively achievable if and only if $R_1 \ge R_P(D_1)$ and $R_2 \ge R_P(D_1, D_2, R_1)$.

As in the case of nonscalable coding, functions that characterize the set of all successively achievable $(R_1,\, R_2,\, E_1,\, E_2,\, D_1,\, D_2)$ can be defined in two equivalent ways:

Definition 3: Given $E_1, E_2 > 0$, and rate $R_1 \ge R_P(D_1, E_1)$, the *scalable rate–reliability–distortion function* $R_P(D_1, D_2, E_1, E_2, R_1)$ is defined as the minimum R_2 such that the 6-tuple $(R_1, R_2, E_1, E_2, D_1, D_2)$ is successively achievable.

Definition 4: Similarly, the *scalable error exponent function* $E_P(D_1,\, D_2,\, R_1,\, R_2,\, E_1)$, defined under the condition $R_1 \ge R_P(D_1, E_1)$, is the maximum E_2 such that $(R_1,\, R_2,\, E_1,\, E_2,\, D_1,\, D_2)$ is successively achievable.

We make heavy use of the method of types in the next section where we construct codes that achieve $E_P(D_1,\, D_2,\, R_1,\, R_2,\, E_1)$. The few properties of types that we provide in the remainder of this section are sufficient to follow the sequel. The reader is referred to [3] for an extensive discussion.

The type of a vector $x^n \in \mathcal{X}^n$ is the empirical distribution given by

$$P(a) = \frac{1}{n}\, N(a|x^n)$$

where $N(a|x^n)$ denotes the number of occurrences of a in x^n. We denote by T_Q^n the type class Q, i.e., the set of all source vectors x^n having type Q. A most fundamental property of types is the type counting lemma which states that the number of distinct types for sequences of length n grows at most polynomially with n [3, Lemma 1.2.2]. Two other properties are crucial for error exponent analysis in source coding:

Type Class Probabilities [3, Lemma 1.2.6]: The probability that $X^n \in T_Q^n$ when X^n is generated i.i.d. with pmf P, decays exponentially as n increases, with the exponent $\mathcal{D}(Q\|P)$. More precisely

$$(n+1)^{-|\mathcal{X}|}\exp\{-n\mathcal{D}(Q\|P)\} \le P^n(T_Q^n) \le \exp\{-n\mathcal{D}(Q\|P)\}. \tag{7}$$

Type Covering Lemma [3, Lemma 2.4.1]: For any distortion measure d_1 on $\mathcal{X}\times\mathcal{Y}_1$, type Q on $\mathcal{X}$, and numbers $D_1 \ge 0, \delta_1 > 0$, there exists a set $\mathcal{B}_1 \in \mathcal{Y}_1^n$ satisfying

$$\frac{1}{n}\log|\mathcal{B}_1| \le R_Q(D_1) + \delta_1,$$

and

$$\forall\, x^n \in T_Q^n, \quad \exists\, y_1^n \in \mathcal{B}_1 \quad \text{s.t. } d_1(x^n,\, y_1^n) \le D_1.$$

Finally, we will also need the scalable extension of the type covering lemma[3] proved in [6].

Lemma 1—Kanlis and Narayan [6]: For distortion measures d_1 on $\mathcal{X}\times\mathcal{Y}_1$ and d_2 on $\mathcal{X}\times\mathcal{Y}_2$, type Q on $\mathcal{X}$, and numbers $D_1,\, D_2 \ge 0$, $R_1 \ge R_Q(D_1)$, $\delta_1,\, \delta_2 > 0$, there exist sets $\mathcal{B}_1 \in \mathcal{Y}_1^n$ and $\mathcal{B}_2(y_1^n) \in \mathcal{Y}_2^n$ for $y_1^n \in \mathcal{B}_1$ satisfying

$$\frac{1}{n}\log|\mathcal{B}_1| \le R_1 + \delta_1$$
$$\frac{1}{n}\log|\mathcal{B}_2(y_1^n)| \le R_Q(D_1,\, D_2,\, R_1) - R_1 + \delta_2, \qquad \forall\, y_1^n \in \mathcal{B}_1$$

and

$$\forall\, x^n \in T_Q^n, \qquad \exists\, \{y_1^n \in \mathcal{B}_1,\, y_2^n \in \mathcal{B}_2(y_1^n)\} \quad \text{s.t.}$$
$$d_1(x^n,\, y_1^n) \le D_1$$
$$d_2(x^n,\, y_2^n) \le D_2.$$

III. Sufficient Conditions for Successive Achievability

In this section, we derive sufficient conditions for successive achievability by constructing an actual scalable source coder that satisfies the given constraints $(D_1,\, D_2,\, E_1,\, E_2,\, R_1)$. The coder's second layer rate R_2 is obviously an upper bound for the scalable rate–reliability–distortion function $R_P(D_1,\, D_2,\, E_1,\, E_2,\, R_1)$.

The coding strategy consists of separate construction of encoding and decoding functions for each type. Thus, initially, the type of the source vector is losslessly transmitted to ensure that the decoders utilize the correct lookup tables for reconstruction. Since, according to the type counting lemma, there are at most polynomially many distinct types, lossless transmission of the source type has an asymptotically negligible impact on the overall coding rate.

To prove the existence of encoding and decoding functions that operate at given $(D_1,\, D_2,\, R_1)$, we employ the nonscalable and scalable type covering lemmas. According to (7), for each type class T_Q^n, we can afford the possibility of having an error event at the first and the second layers if $\mathcal{D}(Q\|P) > E_1$ and $\mathcal{D}(Q\|P) > E_2$, respectively. We utilize this fact when we decide on which type covering lemma to employ at each type.

We separately analyze the two possible cases $E_1 < E_2$ and $E_1 \ge E_2$. Recall that (4) is by definition a necessary condition for achievability of $(R_1,\, R_2,\, E_1,\, E_2,\, D_1,\, D_2)$.

Case I: $E_1 < E_2$. We adopt the following source-coding strategy for each type class T_Q^n.

- If $\mathcal{D}(Q\|P) \le E_1$: Since $\mathcal{D}(Q\|P) \le E_2$ is also implied, we employ the scalable type covering lemma to prevent the error event at both layers. Note that from (4), $R_1 \ge R_Q(D_1)$ follows. Thus, for any $\delta_1,\, \delta_2 > 0$ and large enough n, we generate $2^{n[R_1+\delta_1]}$ balls of radius D_1, and for each D_1-ball, generate $2^{n[R_Q(D_1,\, D_2,\, R_1) - R_1 + \delta_2]}$ balls of radius D_2, such that for every $x^n \in T_Q^n$, there exists a pair of D_1- and D_2-balls containing x^n. The encoder sends the corresponding D_1-ball index in the first layer, and the D_2-ball index in the second layer, so that the ball centers y_1^n and y_2^n can be recovered at the corresponding decoder layers without error.
- If $E_1 < \mathcal{D}(Q\|P) \le E_2$: We need to prevent only the second-layer error event. Although we are in pursuit of constructing a scalable source coder, we utilize the nonscalable type covering lemma for generating $2^{n[R_Q(D_2)+\delta_2]}$ balls of radius D_2, such that for every $x^n \in T_Q^n$, there exists a D_2-ball

[3]This is, in fact, a weaker statement than (and implied by) the result of Kanlis and Narayan.

containing x^n. The encoder sends in the first layer the first R_1 bits out of the total $R_Q(D_2)$ necessary for transmitting the corresponding D_2-ball index, and sends the remaining bits (if any) in the second layer. The first-layer decoder reproduces an arbitrary y_1^n, and the second-layer decoder reproduces the center of the D_2-ball.

- If $E_2 < \mathcal{D}(Q\|P)$: The encoder does not send any bits. The decoders reproduce arbitrary y_1^n and y_2^n.

When the above strategy is applied, the achieved asymptotic rate at the second layer is

$$R_2 = \max\left\{\sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_1,\,D_2,\,R_1),\ \sup_{Q:\,E_1<\mathcal{D}(Q\|P)\le E_2} R_Q(D_2)\right\}.$$

To simplify the preceding expression, we first observe that

$$\sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_1,\,D_2,\,R_1) \ge \sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_2) \quad (8)$$

which follows from the fact that $R_Q(D_1,\,D_2,\,R_1) \ge R_Q(D_2)$ for all Q (see (2)). Therefore,

$$\begin{aligned} R_2 &= \max\left\{\sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_1,\,D_2,\,R_1),\ \sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_2),\ \sup_{Q:\,E_1<\mathcal{D}(Q\|P)\le E_2} R_Q(D_2)\right\} \\ &= \max\left\{\sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_1,\,D_2,\,R_1),\ \sup_{Q:\,\mathcal{D}(Q\|P)\le E_2} R_Q(D_2)\right\} \\ &= \max\left\{\sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_1,\,D_2,\,R_1),\ R_P(D_2,\,E_2)\right\} \quad (9) \end{aligned}$$

where the last equality follows from (5). Moreover, the exponent of the error event $\mathcal{E}_i$ which is achieved by this strategy at layer i is at least E_i, as can be shown by exploiting the type counting lemma and (7):

$$\begin{aligned} \Pr[\mathcal{E}_i] &\le \sum_{Q:\,\mathcal{D}(Q\|P)>E_i} \exp\{-n\mathcal{D}(Q\|P)\} \\ &\le \sum_{Q:\,\mathcal{D}(Q\|P)>E_i} \exp\{-nE_i\} \\ &\le \exp\{-n[E_i-\epsilon_i]\} \end{aligned}$$

where $\epsilon_i \longrightarrow 0$ as $n \longrightarrow \infty$.

Case II: $E_1 \ge E_2$. We adopt the following source-coding strategy for each type class T_Q^n.

- If $\mathcal{D}(Q\|P) \le E_2$: We perform the same two-layer type covering as in Case I. (Note from (4) that $R_1 \ge R_Q(D_1)$.) The encoder sends the corresponding D_1-ball index in the first layer, and the D_2-ball index in the second layer, so that the ball centers can be recovered at the corresponding decoder layers without error.
- If $E_2 < \mathcal{D}(Q\|P) \le E_1$: We need to prevent only the first-layer error event. We employ the nonscalable type covering lemma for generating $2^{n[R_1+\delta_1]}$ balls of radius D_1, such that for every $x^n \in T_Q^n$, there exists a D_1-ball containing x^n. (Once again, note from (4) that $R_1 \ge R_Q(D_1)$.) The encoder sends in the first layer the corresponding D_1-ball index and does not send anything in the second layer. The first-layer decoder reproduces the center of the D_1-ball without error, and the second-layer decoder reproduces an arbitrary y_2^n.
- If $E_1 < \mathcal{D}(Q\|P)$: The encoder does not send any bits. The decoders reproduce arbitrary y_1^n and y_2^n.

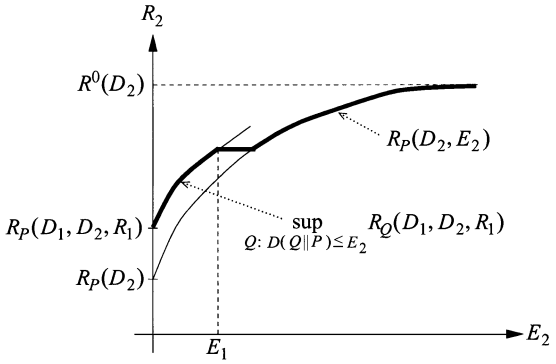

Fig. 2. Indicated in bold is a typical curve of $\boldsymbol{R_P^*(D_1, D_2, E_1, E_2, R_1)}$ as a function of $\boldsymbol{E_2}$, given fixed $\boldsymbol{D_1}$, $\boldsymbol{D_2}$, $\boldsymbol{R_1}$, and $\boldsymbol{E_1}$, where $\boldsymbol{R_1 \ge R_P(D_1, E_1)}$.

With the same logic as before, it is also clear that the achieved error exponent at layer i is at least E_i, for $i = 1, 2$. The achieved asymptotic rate at the second layer in this case is given by

$$R_2 = \sup_{Q:\,\mathcal{D}(Q\|P)\le E_2} R_Q(D_1,\,D_2,\,R_1) \quad (10)$$

since we only expend bits at the second layer for the types Q satisfying $\mathcal{D}(Q\|P) \le E_2$, and the required rate for those types is given by $R_Q(D_1,\,D_2,\,R_1)$.

Remarks:

1) Observe from (9) and (10) that the achieved second-layer rate R_2 in both cases is bounded below by $R_P(D_2,\,E_2)$, as expected.

2) It should be noted that the above scalable source coder construction is source dependent (it relies on the knowledge of the source pmf P) in distinction with the *universal* constructions employed in [9], [6].

Combining (9) and (10), we observe that if

$$\begin{aligned} R_2 &\ge R_P^*(D_1, D_2, E_1, E_2, R_1) \\ &\triangleq \max\left\{\sup_{Q:\,\mathcal{D}(Q\|P)\le \min(E_1,E_2)} R_Q(D_1, D_2, R_1), R_P(D_2, E_2)\right\} \end{aligned} \quad (11)$$

then $(R_1,\,R_2,\,E_1,\,E_2,\,D_1,\,D_2)$ is achievable. For the case $E_1 < E_2$, $R_P^*(D_1,\,D_2,\,E_1,\,E_2,\,R_1)$ immediately reduces to (9). For $E_1 \ge E_2$, we use (8) with E_2 playing the role of E_1, i.e.,

$$\sup_{Q:\,\mathcal{D}(Q\|P)\le E_2} R_Q(D_1,\,D_2,\,R_1) \ge \sup_{Q:\,\mathcal{D}(Q\|P)\le E_2} R_Q(D_2) \quad (12)$$

to conclude that $R_P^*(D_1,\,D_2,\,E_1,\,E_2,\,R_1)$ is indeed equivalent to (10). A fairly general curve of $R_P^*(D_1,\,D_2,\,E_1,\,E_2,\,R_1)$, as a function of E_2 only, is plotted in Fig. 2. $R_P^*(D_1,\,D_2,\,E_1,\,E_2,\,R_1)$ is indicated as the bold curve which traces the curve of (10) for $E_2 \le E_1$. Note that by (12), $R_P^*(D_1,\,D_2,\,E_1,\,E_2,\,R_1)$ is above the curve of $R_P(D_2,\,E_2)$ in this range of E_2 values. On the other hand, as E_2 increases beyond E_1, according to (9), $R_P^*(D_1,\,D_2,\,E_1,\,E_2,\,R_1)$ stays at the constant value

$$\sup_{Q:\,\mathcal{D}(Q\|P)\le E_1} R_Q(D_1,\,D_2,\,R_1)$$

until it meets the curve of $R_P(D_2,\,E_2)$, which it traces afterwards.

For the purpose of the proof (in the next section) that $R_P^*(D_1, D_2, E_1, E_2, R_1)$ actually specifies the entire achievable region, it will be more convenient to work with infimum of $\mathcal{D}(Q\|P)$ over certain sets. More specifically, we seek a dual characterization of the same set of achievable $(R_1, R_2, E_1, E_2, D_1, D_2)$, in analogy to the nonscalable case where $R_P(D_1, E_1)$ and $E_P(D_1, R_1)$ constitute the dual characterization. The corresponding sufficient condition for achievability is given by

$$E_2 \le E_P^*(D_1, D_2, R_1, R_2, E_1) \triangleq \min\left\{ \inf_{\substack{Q:\, \mathcal{D}(Q\|P)\le E_1,\\ R_Q(D_1,D_2,R_1)>R_2}} \mathcal{D}(Q\|P),\ E_P(D_2, R_2) \right\} \quad (13)$$

with the standard convention that infimum over an empty set yields infinity. The best way to justify the duality between

$$E_P^*(D_1, D_2, R_1, R_2, E_1) \quad \text{and} \quad R_P^*(D_1, D_2, E_1, E_2, R_1)$$

is perhaps through Fig. 2. Note that the equality

$$\inf_{\substack{Q:\, \mathcal{D}(Q\|P)\le E_1,\\ R_Q(D_1,D_2,R_1)>R_2}} \mathcal{D}(Q\|P) = \inf_{Q:\, R_Q(D_1,D_2,R_1)>R_2} \mathcal{D}(Q\|P) \quad (14)$$

holds whenever the right-hand side is lower than or equal to E_1. Otherwise, the left-hand side yields infinity. Another observation is that when (14) holds, the resultant infimum is lower than or equal to $E_P(D_2, R_2)$. Therefore, the first infimum in (13) fits the first part of the bold curve in Fig. 2 where we potentially have

$$R_P^*(D_1, D_2, E_1, E_2, R_1) > R_P(D_2, E_2).$$

Obviously, the second component in the minimization of (13), $E_P(D_2, R_2)$, corresponds to the second part of the curve in Fig. 2, i.e., where

$$R_P^*(D_1, D_2, E_1, E_2, R_1) = R_P(D_2, E_2).$$

Characterization of achievable 6-tuples through $E_P^*(D_1, D_2, R_1, R_2, E_1)$ has the additional advantage that we can now compare the result with the second-layer error exponent obtained in [6], where the first layer is assumed to achieve the optimal exponent $E_1 = E_P(D_1, R_1)$. In [6], the formula for the best second-layer exponent was given as the right-hand side of (14). $E_P^*(D_1, D_2, R_1, R_2, E_1)$, on the other hand, promises a potentially larger (and in fact optimal) second-layer exponent if the second-layer rate R_2 is large enough. It is noteworthy that in order to achieve the optimal result here we had recourse to a source-dependent strategy.

IV. Necessary Conditions for Successive Achievability

We derive necessary conditions for successive achievability of $(R_1, R_2, E_1, E_2, D_1, D_2)$ by extending the approach of Marton [9]. For any n-block coding strategy, characterized by encoding and decoding functions f_1, f_2, g_1, and g_2, we introduce the notation

$$\mathcal{U}_1(f_1, g_1) = \{x^n : d_1(x^n, g_1(f_1(x^n))) > D_1\} \quad (15)$$

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

$$\mathcal{U}_2(f_1, f_2, g_2) = \{x^n : d_2(x^n, g_2(f_1(x^n), f_2(x^n))) > D_2\}. \quad (16)$$

Here, we dropped the superscript (n) from f_i and g_i for the sake of notational simplicity. We proceed to state the main theorem of this section.

Theorem 2: Given a discrete memoryless source P with $P(x) > 0$ for all $x \in \mathcal{X}$, let $R_1 \ge R_P(D_1, E_1)$. An n-block coding strategy, characterized by encoding functions $f_i : \mathcal{X}^n \longrightarrow \mathcal{M}_i^{(n)}$ for $i = 1, 2$, and decoding functions $g_1 : \mathcal{M}_1^{(n)} \longrightarrow \mathcal{Y}_1^n$ and $g_2 : \mathcal{M}_1^{(n)} \times \mathcal{M}_2^{(n)} \longrightarrow \mathcal{Y}_2^n$, satisfies

$$\frac{1}{n}\log\left|\mathcal{M}_1^{(n)}\right| \le R_1 + \delta$$
$$\frac{1}{n}\log\left|\mathcal{M}_2^{(n)}\right| \le R_2 - R_1 + \delta$$
$$-\frac{1}{n}\log P^n\left(\mathcal{U}_1(f_1, g_1)\right) \ge E_1 - \epsilon \quad (17)$$

for any $\epsilon > 0$ and $\delta > 0$, and for large enough n, only if

$$\limsup_{n\to\infty} -\frac{1}{n}\log P^n\left(\mathcal{U}_2(f_1, f_2, g_2)\right) \le E_P^*(D_1, D_2, R_1, R_2, E_1). \quad (18)$$

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

$$E_P(D_1, D_2, R_1, R_2, E_1) = E_P^*(D_1, D_2, R_1, R_2, E_1) \quad (19)$$
$$R_P(D_1, D_2, E_1, E_2, R_1) = R_P^*(D_1, D_2, E_1, E_2, R_1). \quad (20)$$

To prove the theorem we make use of two lemmas. Lemma 2 states that when the first-layer coding achieves the error exponent E_1, then the probability of the "uncovered" subset $\mathcal{U}_1(f_1, g_1)$ vanishes not only for the actual distribution P, but for all distributions Q close enough to P, i.e., all Q that satisfy $\mathcal{D}(Q\|P) < E_1$. On the other hand, according to Lemma 3, for any distribution $Q \in \mathcal{Q}$, where $\mathcal{Q}$ is the set on which the minimization of $\mathcal{D}(Q\|P)$ is performed in order to compute $E_P^*(D_1, D_2, R_1, R_2, E_1)$, the probability of the subset $\mathcal{U}_2(f_1, f_2, g_2)$ is asymptotically bounded away from 0.

Lemma 2: If a coding strategy satisfies (17), then for any $\gamma > 0$ and for any probability distribution Q such that $\mathcal{D}(Q\|P) < E_1 - \gamma$, it also satisfies

$$Q^n\left(\mathcal{U}_1(f_1, g_1)\right) \longrightarrow 0$$

as $n \to \infty$.

Proof: Let

$$G_n \triangleq \left\{x^n : \left|\frac{1}{n}\log\frac{Q^n(x^n)}{P^n(x^n)} - \mathcal{D}(Q\|P)\right| < \eta\right\} \quad (21)$$

where $0 < \eta < E_1 - \gamma - \mathcal{D}(Q\|P)$. By the weak law of large numbers, $Q^n(G_n) \longrightarrow 1$ as $n \longrightarrow \infty$. Next

$$Q^n\left(\mathcal{U}_1(f_1, g_1)\right) = \sum_{x^n \in \mathcal{U}_\infty} (f_1, g_1)\, Q^n(x^n) \le \sum_{x^n \notin G_n} Q^n(x^n) + \sum_{x^n \in G_n \cap \mathcal{U}_1} (f_1, g_1)\, Q^n(x^n). \quad (22)$$

The first term tends to zero by the weak law of large numbers

$$\sum_{x^n \notin G_n} Q^n(x^n) = 1 - Q^n(G_n) \longrightarrow 0. \quad (23)$$

The second term satisfies

$$\sum_{x^n \in G_n \cap \mathcal{U}_1} (f_1, g_1)\, Q^n(x^n) = \sum_{x^n \in G_n \cap \mathcal{U}_1} (f_1, g_1)\, P^n(x^n) \exp\left\{\log\frac{Q^n(x^n)}{P^n(x^n)}\right\} < P^n\left(\mathcal{U}_1(f_1, g_1)\right) \exp\left\{n[\mathcal{D}(Q\|P) + \eta]\right\}$$

where the inequality follows from (21). We combine this with (17) using $\epsilon < \gamma$ and large enough n to obtain

$$\begin{aligned}\sum_{x^n \in G_n \cap \mathcal{U}_1} (f_1, g_1)\, Q^n(x^n) &< \exp\{n[\mathcal{D}(Q\|P)+\eta-E_1+\epsilon]\} \\ &< \exp\{n[\epsilon-\gamma]\} \\ &\longrightarrow 0. \end{aligned} \tag{24}$$

From (22)–(24), we conclude that $Q^n(\mathcal{U}_1(f_1, g_1)) \longrightarrow 0$ as $n \longrightarrow \infty$. □

Lemma 3: For $\gamma > 0$, let $\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2$, where

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

If $Q \in \mathcal{Q}$, then there exists a number $\alpha(Q, D_1, D_2, R_1, R_2) > 0$ such that for sufficiently large n

$$Q^n(\mathcal{U}_2(f_1, f_2, g_2)) \geq \alpha(Q, D_1, D_2, R_1, R_2) \tag{25}$$

for all coding strategies (f_1, f_2, g_2) satisfying (17).

Proof: If $Q \in \mathcal{Q}_2$, then the result follows immediately from [9, proof of Theorem 1]. Suppose $Q \notin \mathcal{Q}_2$, and let $R_2 < R_Q(D_1, D_2, R_1)$. Recall that $R_Q(D_1, D_2, R_1)$ is a continuous function in all of its arguments, and a nonincreasing function in D_1 and D_2. Thus, we can pick small enough $\delta > 0$ such that

$$R_2 + 2\delta < R_Q(D_1, D_2, R_1 + \delta).$$

There also exist $\zeta > 0$ and $D_2^* > D_2$ such that

$$R_2 + 2\delta < R_Q(D_1 + \zeta, D_2, R_1 + \delta) \tag{26}$$

and

$$R_2 + 2\delta = R_Q(D_1 + \zeta, D_2^*, R_1 + \delta). \tag{27}$$

Further, if $Q \in \mathcal{Q}_1$, i.e., if $\mathcal{D}(Q\|P) < E_1 - \gamma$ is also satisfied, then, from Lemma 2, for large enough n, we have

$$\begin{aligned}D_1(f_1, g_1) &\triangleq \mathbf{E}_{Q^n}\{d_1(X^n, g_1(f_1(X^n)))\} \\ &\leq [1 - Q^n(\mathcal{U}_1(f_1, g_1))]\, D_1 + Q^n(\mathcal{U}_1(f_1, g_1))\, d_{1,\max} \\ &\leq D_1 + \zeta,\end{aligned}$$

where $d_{1,\max} = \max_{x,y_1} d_1(x, y_1) < \infty$. By arguments similar to the standard proof of the weak converse to the rate-distortion theorem (e.g., [2, Theorem 13.2.1]), for n large enough to satisfy $D_1(f_1, g_1) \leq D_1 + \zeta$ and (17), it follows that

$$R_2 + 2\delta \geq R_Q(D_1 + \zeta, D_2(f_1, f_2, g_2), R_1 + \delta) \tag{28}$$

where

$$D_2(f_1, f_2, g_2) \triangleq \mathbf{E}_{Q^n}\{d_2(X^n, g_2(f_1(X^n), f_2(X^n)))\}.$$

Hence we have

$$\begin{aligned}R_Q(D_1 + \zeta, D_2, R_1 + \delta) &> R_Q(D_1 + \zeta, D_2^*, R_1 + \delta) \\ &\geq R_Q(D_1 + \zeta, D_2(f_1, f_2, g_2), R_1 + \delta)\end{aligned}$$

which, in turn, implies from the monotonicity of $R_Q(\cdot, \cdot, \cdot)$ in the second argument that

$$D_2 < D_2^* \leq D_2(f_1, f_2, g_2). \tag{29}$$

On the other hand, we have

$$\begin{aligned}D_2(f_1, f_2, g_2) \leq\, &[1 - Q^n(\mathcal{U}_2(f_1, f_2, g_2))]\, D_2 \\ &+ Q^n(\mathcal{U}_2(f_1, f_2, g_2))\, d_{2,\max}\end{aligned}$$

where $d_{2,\max} = \max_{x,y_2} d_2(x, y_2) < \infty$. Therefore,

$$Q^n(\mathcal{U}_2(f_1, f_2, g_2)) \geq \frac{D_2(f_1, f_2, g_2) - D_2}{d_{2,\max} - D_2} \geq \frac{D_2^* - D_2}{d_{2,\max} - D_2} > 0$$

where the last two inequalities employ (29). The proof is completed by setting

$$\alpha(Q, D_1, D_2, R_1, R_2) = \frac{D_2^* - D_2}{d_{2,\max} - D_2}.$$

Note from (27) that α depends on Q, D_1, R_1, and R_2, through D_2^*. □

We close this section with the proof of Theorem 2.

Proof [Theorem 2]: Pick any $Q \in \mathcal{Q}$, where $\mathcal{Q}$ is as defined in Lemma 3. Let G_n be defined as in Lemma 2 with η unspecified. For sufficiently large n, the weak law of large numbers ensures that

$$Q^n(G_n) > 1 - \tfrac{1}{2}\,\alpha(Q, D_1, D_2, R_1, R_2). \tag{30}$$

We next consider the error probability at the second layer

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

for sufficiently large n, where the last inequality follows from (25) and (30). This implies that

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

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 $\gamma, \eta \longrightarrow 0$. Note from (13) that

$$\begin{aligned}&E_P^*(D_1, D_2, R_1, R_2, E_1) \\ &\quad= \min\left\{\inf_{Q \in \mathcal{Q}_1} \mathcal{D}(Q\|P),\ \inf_{Q \in \mathcal{Q}_2} \mathcal{D}(Q\|P)\right\} \\ &\quad= \inf_{Q \in \mathcal{Q}} \mathcal{D}(Q\|P).\end{aligned}$$

□

V. Successive Refinability

A very desirable feature of scalable source coding is *successive refinability*. The source is said to be successively refinable if there exists a scalable coding strategy which achieves the desired output quality by expending only the rate needed in a nonscalable scenario, at each layer. In the rate-distortion sense, this notion was introduced in [7], [8], and [4], which also provided the necessary and sufficient conditions to be satisfied by the conditional distributions $W_1(y_1|x)$ and $W_2(y_2|x)$ that achieve $R_P(D_1)$ and $R_P(D_2)$, respectively, in (1).

Successive refinement in rate–reliability–distortion analysis was first discussed in [6]. The analysis in [6] was concerned with whether the nonscalable coding exponent $E_P(D_2, R_2)$ coincides with the exponent of the event that either the first layer or the second layer introduces error, when the first-layer exponent is fixed at $E_P(D_1, R_1)$. Since we discuss the tradeoff between the separately defined error exponents E_1 and E_2, we are naturally interested in the notion of successive refinement defined below.

Definition 5: The source P is successively refinable from (D_1, E_1) to (D_2, E_2), if the 6-tuple

$$(R_P(D_1, E_1), R_P(D_2, E_2), E_1, E_2, D_1, D_2)$$

with $R_P(D_2, E_2) \geq R_P(D_1, E_1)$ is successively achievable.

This definition, which is repeated here for convenience and notational adjustment, first appeared in [5], where the authors analyzed the case $E_2 \geq E_1$. An equivalent definition in terms of Marton's error exponent function is as follows.

Definition 6: The source P is successively refinable from (D_1, R_1) to (D_2, R_2) for $R_2 \geq R_1$, if the 6-tuple

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

is successively achievable.

It is clear from our characterization of the achievable 6-tuples $(R_1, R_2, E_1, E_2, D_1, D_2)$ that the source is successively refinable in the rate–reliability–distortion sense if and only if

$$R_P(D_2, E_2) = R_P(D_1, D_2, E_1, E_2, R_P(D_1, E_1)) \tag{31}$$

or, equivalently, if and only if

$$E_P(D_2, R_2) = E_P(D_1, D_2, R_1, R_2, E_P(D_1, R_1)). \tag{32}$$

In this section, we will present necessary and sufficient conditions for (31) to hold.

Comparing (31) and (11), one can state that the source P is successively refinable from (D_1, E_1) to (D_2, E_2) if and only if

$$R_P(D_2, E_2) \geq \sup_{Q:\, \mathcal{D}(Q\|P) \leq \min(E_1, E_2)} R_Q(D_1, D_2, R_P(D_1, E_1)). \tag{33}$$

If $E_2 \leq E_1$, using the definition of $R_P(D_2, E_2)$, the condition (33) can be recast as

$$\sup_{Q:\, \mathcal{D}(Q\|P) \leq E_2} R_Q(D_2) \geq \sup_{Q:\, \mathcal{D}(Q\|P) \leq E_2} R_Q(D_1, D_2, R_P(D_1, E_1)). \tag{34}$$

Since $R_Q(D_2) \leq R_Q(D_1, D_2, R_P(D_1, E_1))$ for all Q, the preceding inequality is satisfied if and only if

$$R_{Q^*}(D_2) = R_{Q^*}(D_1, D_2, R_P(D_1, E_1)) \tag{35}$$

for the distribution Q^* achieving the supremum on the right-hand side of (34). (In fact, (34) can only be satisfied with equality, therefore Q^* also achieves the supremum on the left-hand side.) From (1) and (2), it follows that (35) is satisfied if and only if there exists a conditional pmf $W(y_1, y_2|x)$ such that $X - Y_2 - Y_1$ forms a Markov chain, and

$$\begin{aligned} &\mathbf{I}_{Q^*W}(X; Y_1) \leq R_P(D_1, E_1) \\ &\mathbf{I}_{Q^*W}(X; Y_2) = R_{Q^*}(D_2) = R_P(D_2, E_2) \\ &\mathbf{E}_{Q^*W}\, d_1(X, Y_1) \leq D_1 \\ &\mathbf{E}_{Q^*W}\, d_2(X, Y_2) \leq D_2. \end{aligned}$$

If, on the other hand, $E_2 \geq E_1$, then (33) becomes

$$\sup_{Q:\, \mathcal{D}(Q\|P) \leq E_2} R_Q(D_2) \geq \hat{R}_2 \tag{36}$$

where

$$\hat{R}_2 \triangleq \sup_{Q:\, \mathcal{D}(Q\|P) \leq E_1} R_Q(D_1, D_2, R_P(D_1, E_1)).$$

Note that $\hat{R}_2$ does not depend on E_2 and the left-hand side of (36) is monotonically nondecreasing in E_2. Therefore, unless

$$\hat{R}_2 > R^0(D_2) = \max_Q R_Q(D_2)$$

there exists a second-layer exponent threshold $\hat{E}_2 \geq E_1$ such that (36) is satisfied if and only if $E_2 \geq \hat{E}_2$. In fact

$$\hat{E}_2 = E_P(D_2, \hat{R}_2).$$

Pictorially, $\hat{R}_2$ corresponds to the ordinate of the straight bold line segment in Fig. 2, and $\hat{E}_2$ is the abscissa of the right endpoint of the same line segment.[4] If $\hat{R}_2 > R^0(D_2)$, successive refinement is not possible for any $E_2 \geq E_1$.

Note that our successive refinability conditions are fundamentally different from those given in [5]. Our result is surprising in that for $E_2 \geq \hat{E}_2$, successive refinability is granted without any further condition, whereas the conditions in [5] for $E_2 \geq E_1$ are somewhat reminiscent of those we derived for the case $E_2 \leq E_1$.

A special case of the above discussion is when $E_1, E_2 \longrightarrow 0$. Then $R_P(D_i, E_i) = R(D_i)$, for $i = 1, 2$. In that case, (33) reduces to

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

which can only be satisfied with equality. The necessary and sufficient conditions for equality were provided in [8], [4].

VI. Conclusion

We characterized the region of all achievable 6-tuples $(R_1, R_2, E_1, E_2, D_1, D_2)$ for the scalable source coding scenario. Given source P, the characterization is in terms of the information divergence $\mathcal{D}(Q\|P)$ and the rate-distortion functions $R_Q(D_2)$ and $R_Q(D_1, D_2, R_1)$, with respect to all possible sources Q. 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 $R_1 = R_P(D_1, E_1)$ and $R_2 = R_P(D_2, E_2)$ for all second-layer error exponents E_2 above a specified threshold $\hat{E}_2$.

Acknowledgment

The authors wish to thank the anonymous referees and the Associate Editor for their constructive comments and perspectives, which helped to improve the quality and the readability of this correspondence. Special thanks to the referee who suggested the inclusion of Fig. 1.

References

[1] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 721–726, Nov. 1985.

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

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

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

[5] E. A. Haroutunian and A. N. Harutyunyan, "Successive refinement of information with reliability criterion," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, June 2000, p. 205.

[6] A. Kanlis and P. Narayan, "Error exponents for successive refinement by partitioning," *IEEE Trans. Inform. Theory*, vol. 42, pp. 275–282, Jan. 1996.

[7] V. Koshelev, "Hierarchical coding of discrete sources," *Probl. Pered. Inform.*, vol. 16, no. 3, pp. 31–49, 1980.

[8] ——, "An evaluation of the average distortion for discrete schemes of sequential approximation," *Probl. Pered. Inform.*, vol. 17, no. 3, pp. 20–33, 1981.

[9] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 197–199, Mar. 1974.

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

[4] $\boldsymbol{E_P(D_2, \hat{R}_2)}$, as a function of $\boldsymbol{\hat{R}_2}$ may have countably many discontinuities, as discussed in [9], [1]. Therefore, a more precise formula is $\boldsymbol{\hat{E}_2 = \lim_{\delta \to 0} E_P(D_2, \hat{R}_2 - \delta)}$. If, $\boldsymbol{E_P(D_2, \hat{R}_2)}$ is indeed discontinuous at $\boldsymbol{\hat{R}_2}$, then $\boldsymbol{\hat{E}_2}$ is the minimum $\boldsymbol{E_2}$ such that $\boldsymbol{R_P(D_2, E_2)}$ and $\boldsymbol{\hat{R}_2}$ coincide.