Lossy Compression of Exponential and Laplacian Sources using Expansion Coding

Hongbo Si
The University of Texas at Austin
HongboSi@utexas.edu

O. Ozan Koyluoglu
The University of Arizona
ozan@email.arizona.edu

Sriram Vishwanath
The University of Texas at Austin
sriram@austin.utexas.edu

***Abstract*—A general method of source coding is proposed in this paper, which enables one to reduce the problem of compressing an analog (continuous-valued) source to a set of much simpler problems, compressing discrete sources. Specifically, the focus is on lossy compression of exponential and Laplace sources, which are represented as set of discrete variables through a finite alphabet expansion. Due to the decomposability property of such sources, the resulting random variables post expansion are independent and discrete. Thus, these variables can be considered as independent discrete source coding problems, and the original problem is reduced to coding over these sources with a total distortion constraint. Any feasible solution to this resulting optimization problem corresponds to an achievable rate distortion pair of the original continuous-valued source compression problem. Although finding the optimal solution for a given distortion is not a tractable task, we show that, via a heuristic choice, our expansion coding scheme still presents a good performance in the low distortion regime. Further, by adopting low-complexity codes designed for discrete source coding, the total coding complexity can be reduced for practical implementations.**

I. Introduction

The compression of continuous-valued sources remains one of the most well-studied (and practically valuable) research problems in information theory. Given the increased importance of voice, video and other multimedia, all of which are typically "analog" in nature, the value associated with low-complexity algorithms to compress continuous-valued data is likely to remain significant in the years to come.

For discrete-valued ("finite alphabet") sources, both the associated coding theorems and practical coding schemes are now well known. Trellis based quantizers [1] are the first to achieve the rate distortion tradeoff, but with encoding complexity scaling exponentially with the constraint length. A low density parity check (LDPC) ensemble, under suitable conditions on the ensemble structure is shown to achieve the rate distortion bound using an optimal decoder [2]. Recently, low density generator matrix (LDGM) codes, with suitably irregular degree distributions, are shown to empirically perform close to the Shannon rate-distortion bound with message-passing algorithms [3]. More recently, polar codes [4][5] are the first provably rate distortion limit achieving codes with low encoding and decoding complexity [6].

In the case of analog sources, although both practical coding schemes as well as theoretical analysis are heavily studied, very limited literature exists that connects theory with low-complexity codes in practice. The most relevant literature in this context is on lattice compression and its low-density constructions [7], although this literature is somewhat limited in scope and application. In the domains of image compression and speech coding, Laplacian and exponential distributions are widely adopted as natural models of correlation between pixels and amplitude of voice [8]. Exponential distribution is also fundamental in characterizing continuous-time Markov processes [9]. Many schemes have been proposed, such as entropy constrained scalar quantization (ECSQ) [10], vector quantization (VQ) [11], Markov chain Monte Carlo (MCMC) based approach [12], however, they are still suboptimal in rate with respect to Shannon limit [12]. Our general understanding of low-complexity coding schemes with good rates, in particular for the low-distortion regime, remains limited.

In this paper, we present an expansion coding scheme for both exponential and Laplacian sources. Previously, our work in [13] considers expansion coding for the dual problem, i.e. the channel coding case, where coding over exponential noise channels are converted to coding over a set of parallel (and independent) discrete channels. For source coding, we utilize a similar approach: expanding exponential and Laplace sources into binary sequences, and coding over the resulting set of parallel discrete sources. We show that the achievable rates from this scheme approach the rate distortion limit in the low distortion regime.

The rest of paper is organized as follows. The next section describes the decomposability of exponential distribution. In Section III and IV, we present the main results of this paper, expansion coding schemes for exponential and Laplacian sources, respectively. The paper concludes with a discussion section. Proofs to the main theorems are given in the appendix. (Detailed proofs are provided in [14].)

II. Decomposability of Exponential Distribution

The intuition underlying expansion coding originates from the decomposability property of exponential random variables, which can be expressed as a summation of a set of independent discrete-valued random variables. The following lemma summarizes this phenomenon:

Lemma 1 ([15]). *Let B_l's be independent Bernoulli random variables, and their distributions are given by parameters $b_l \triangleq Pr\{B_l = 1\}$. Then, the random variable $B = \sum_{l=-\infty}^{\infty} 2^l B_l$*

is exponentially distributed with mean λ^{-1}, if and only if the choice of b_l is given by $b_l = 1/(1+e^{\lambda 2^l})$.

III. Exponential Source Coding

A. Problem Setup

Consider an i.i.d. exponential source sequence $X_1, \ldots, X_n$, i.e., omitting index i, each variable has a pdf given by

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \tag{1}$$

where λ^{-1} is the mean of X. Distortion measure of concern is the "one-sided error distortion", i.e.

$$d(x_{1:n}, \hat{x}_{1:n}) = \begin{cases} \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \hat{x}_i), & \text{if } x_{1:n} \succcurlyeq \hat{x}_{1:n}, \\ \infty, & \text{otherwise.} \end{cases} \tag{2}$$

This setup is equivalent to the one in [9], where another distortion measure is considered. (See [14].)

Lemma 2 ([9])**.** *The rate distortion function for an exponential source with the one-sided error distortion is given by*

$$R(D) = \begin{cases} -\log(\lambda D), & 0 \leq D \leq \frac{1}{\lambda}, \\ 0, & D > \frac{1}{\lambda}. \end{cases} \tag{3}$$

Moreover, the optimal conditional distribution to achieve the limit is given by

$$f^*_{X|\hat{X}}(x|\hat{x}) = \frac{1}{D} e^{-(x-\hat{x})/D}, \quad x \geq \hat{x} \geq 0. \tag{4}$$

B. Expansion Coding

Using Lemma 1, we reconstruct the exponential distribution by a set of discrete Bernoulli random variables. In particular, the expansion of exponential source over levels ranging from $-L_1$ to L_2 can be expressed as

$$X_i = \sum_{l=-L_1}^{L_2} 2^l X_{i,l}, \quad i = 1, 2, \ldots, n, \tag{5}$$

where $X_{i,l}$ are Bernoulli random variables with parameter

$$p_l \triangleq \Pr\{X_{i,l} = 1\} = \frac{1}{1+e^{\lambda 2^l}}. \tag{6}$$

This expansion perfectly approximates exponential source by letting $L_1, L_2 \to \infty$. Consider a similar expansion of the source estimate, i.e.

$$\hat{X}_i = \sum_{l=-L_1}^{L_2} 2^l \hat{X}_{i,l}, \quad i = 1, 2, \ldots, n, \tag{7}$$

where $\hat{X}_{i,l}$ is the resulting Bernoulli random variable with parameter $\hat{p}_l \triangleq \Pr\{\hat{X}_{i,l} = 1\}$. Utilizing the expansion method, the original problem of coding for a continuous source can be translated to a problem of coding for a set of independent binary sources. This transformation, although seemingly obvious, is valuable as one can utilize powerful coding schemes over discrete sources to achieve rate distortion limits with low complexity. In particular, we design two schemes for the binary source coding problem at each level.

1) Coding with one-sided distortion: We formulate each level as a binary source coding problem under the binary one-sided distortion constraint: $d_O(x_l, \hat{x}_l) = \mathbf{1}_{\{x_l > \hat{x}_l\}}$. Denoting the distortion at level l as d_l, an asymmetric test channel (Z-channel) from $\hat{X}_l$ to X_l can be constructed, where

$$\Pr\{X_l = 1 | \hat{X}_l = 0\} = \frac{d_l}{1 - p_l + d_l}.$$

Based on this, we have $p_l - \hat{p}_l = d_l$, and the achievable rate at a single level l is given by

$$R_l = H(p_l) - (1 - p_l + d_l) H\left(\frac{d_l}{1 - p_l + d_l}\right). \tag{8}$$

Due to the decomposability property as stated previously, the coding scheme provided is over a set of parallel discrete levels indexed by $l = -L_1, \ldots, L_2$. Thus, by adopting rate distortion limit achieving codes over each level, expansion coding scheme readily achieves the following result:

Theorem 3. *For an exponential source, expansion coding achieves the rate distortion pair given by*

$$R^{(1)} = \sum_{l=-L_1}^{L_2} R_l, \tag{9}$$

$$D^{(1)} = \sum_{l=-L_1}^{L_2} 2^l d_l + 2^{-L_2}/\lambda^2 + 2^{-L_1}, \tag{10}$$

for any $L_1, L_2 > 0$, and $d_l \in [0, 0.5]$ for $l \in \{-L_1, \cdots, L_2\}$, where p_l is given by (6)*.*

Note that, the last two terms in (10) are resulting from the truncation and vanish in the limit of large number of levels.

2) Successive encoding and decoding: Note that it is not necessary to make sure $X_l \geq \hat{X}_l$ for every l to guarantee $X \geq \hat{X}$. To this end, we introduce successive coding scheme, where encoding and decoding start from the highest level L_2 to the lowest. At a certain level, if all higher levels are encoded as equal to the source, then we must model this level as binary source coding with the one-sided distortion. Otherwise, we formulate this level as binary source coding with the symmetric distortion. In particular for the latter case, the distortion of concern is Hamming distortion, i.e. $d_H(x_l, \hat{x}_l) = \mathbf{1}_{\{x_l \neq \hat{x}_l\}}$. Denoting the equivalent distortion at level l as d_l, i.e. $\mathbb{E}[X_l - \hat{X}_l] = d_l$, then the symmetric test channel from $\hat{X}_l$ to X_l is modeled as

$$\Pr\{X_l = 1 | \hat{X}_l = 0\} = \Pr\{X_l = 0 | \hat{X}_l = 1\} = \frac{d_l}{1 - 2p_l + 2d_l}.$$

Hence, the achievable rate at level l is given by

$$\bar{R}_l = H(p_l) - H\left(\frac{d_l}{1 - 2p_l + 2d_l}\right). \tag{11}$$

Based on these, we have the following achievable result:

Theorem 4. *For an exponential source, applying successive coding, expansion coding achieves the rate distortion pairs*

$$R^{(2)} = \sum_{l=-L_1}^{L_2} \left[q_l R_l + (1-q_l)\,\bar{R}_l\right], \tag{12}$$

$$D^{(2)} = \sum_{l=-L_1}^{L_2} 2^l d_l + 2^{-L_2}/\lambda^2 + 2^{-L_1}, \tag{13}$$

for any $L_1, L_2 > 0$, and $d_l \in [0, 0.5]$ for $l \in \{-L_1, \cdots, L_2\}$. Here, p_l is given by (6)*, $q_{L_2} = 1$, and $q_l = \prod_{k=l+1}^{L_2}(1-d_k)$ for $l \le L_2 - 1$.*

In this sense, the achievable pairs in both theorems are given by optimization problems over a set of parameters $\{d_{-L_1}, \ldots, d_{L_2}\}$. However, the problems are not convex, so an effective theoretical analysis may not be performed here for the optimal solution. But, by a heuristic choice of d_l, we can still get a good performance. Inspired by the fact that the optimal scheme models noise as exponential with parameter $1/D$ in the test channel, we design d_l as the expansion parameter from this distribution, i.e.

$$d_l = \frac{1}{1+e^{2^l/D}}. \tag{14}$$

We note that higher levels get higher priority and lower distortion with this choice, which is consistent with the intuition. This choice of d_l may not guarantee any optimality, although simulation results imply that it can be a local optimum. In the following, we show that the proposed expansion coding scheme achieves within a constant gap to the rate distortion function (at each distortion value).

Theorem 5. *For any $D \in [0, 1/\lambda]$, there exists a constant $c > 0$, such that for $L_1, L_2 > -\log(\lambda D)$, the achievable rate pairs obtained from expansion coding schemes are both within c bit gap to Shannon rate distortion function, i.e.*

$$R^{(1)} - R(D^{(1)}) \le c, \quad R^{(2)} - R(D^{(2)}) \le c,$$

where $D^{(1)}$ and $D^{(2)}$ are given by (10) *and* (13) *respectively, with a choice of d_l as in* (14)*.*

Proof: See Appendix A, and for a detailed proof please refer to [14]. ■

C. Numerical Results

Numerical results showing achievable rates along with the rate distortion limit are plotted in Fig. 1. It is evident that both forms of expansion coding perform within a constant gap of the limit. Theorem 5 showcases that this gap is bounded by a constant. Here, numerical results show that the gap is not necessarily as wide as predicted by the analysis. In fact, the gap is numerically found to correspond to 0.43 bits and 0.24 bits for each coding scheme respectively.

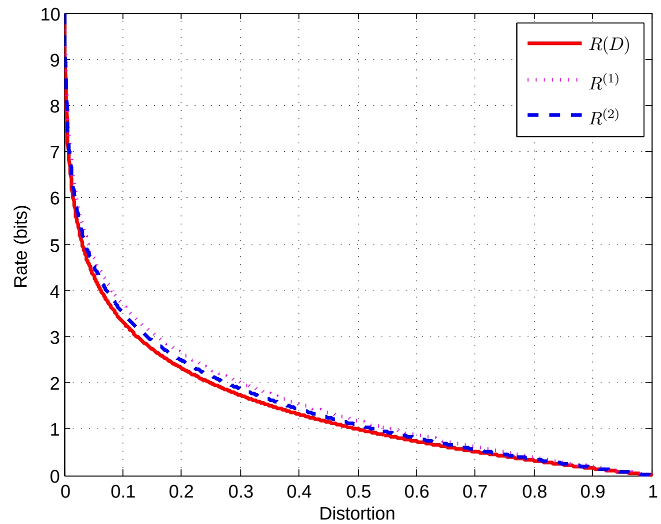

Fig. 1: **Achievable rate distortion pairs using expansion coding for exponential source with the one-sided error distortion.** We set $\lambda = 1$, and $L_1 = L_2 = -\log D + 5$. $R(D)$ (red-solid) is the rate distortion limit; $R^{(1)}$ (purple-dotted) is the achievable rates curve given by Theorem 3; $R^{(2)}$ (blue-dashed) is the achievable rates curve given by Theorem 4.

IV. Laplacian Source Coding

A. Problem Setup

In this section, we focus on Laplacian source coding. Consider an i.i.d. Laplacian source sequence $X_1, \ldots, X_n$, i.e., omitting index i, each variable has a pdf given by

$$f_X(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad x \in \mathbb{R}, \tag{15}$$

where $2/\lambda^2$ is the variance of X. Distortion measure here is the absolute value error distortion, i.e.

$$d(x_{1:n}, \hat{x}_{1:n}) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \hat{x}_i|. \tag{16}$$

Lemma 6 ([16]). *The rate distortion function for a Laplacian source with the absolute error distortion is given by*

$$R(D) = \begin{cases} -\log(\lambda D), & 0 \le D \le \frac{1}{\lambda}, \\ 0, & D > \frac{1}{\lambda}. \end{cases} \tag{17}$$

Moreover, the optimal conditional distribution is

$$f^*_{X|\hat{X}}(x|\hat{x}) = \frac{1}{2D} e^{-|x-\hat{x}|/D}, \quad x, \hat{x} \in \mathbb{R}. \tag{18}$$

B. Expansion Coding

By noting that Laplacian is two-sided exponential, the expansion of source and estimate over levels ranging from $-L_1$ to L_2 can be expressed as

$$X_i = X_i^{\text{sign}} \sum_{l=-L_1}^{L_2} 2^l X_{i,l}, \quad i = 1, 2, \ldots, n, \tag{19}$$

$$\hat{X}_i = \hat{X}_i^{\text{sign}} \sum_{l=-L_1}^{L_2} 2^l \hat{X}_{i,l}, \quad i = 1, 2, \ldots, n, \tag{20}$$

where X_i^{sign} and $\hat{X}_i^{\text{sign}}$ represent the sign of X_i and $\hat{X}_i$ correspondingly, both uniformly distributed from $\{-1,+1\}$.

By performing expansion, we reduce the original problem to coding for a set of independent binary sources. However, particularly for Laplacian case, we let $X^{\text{sign}} = \hat{X}^{\text{sign}}$, i.e., using 1 bit to perfectly recover the sign bit, and then for other levels, we formulate each as a binary source coding problem with Hamming distortion. In particular, for level l, we design a symmetric test channel from $\hat{X}_l$ to X_l, where the cross probability is given by $d_l = (p_l - \hat{p}_l)/(1-2\hat{p}_l)$. Hence, the achievable rate at level l is given by

$$R_l = H(p_l) - H(d_l). \tag{21}$$

To this end, we have the following result:

Theorem 7. *For Laplacian source X, expansion coding, where the estimate $\hat{X}$ is constructed as in* (20)*, achieves the rate distortion pair (R, D) with*

$$R = 1 + \sum_{l=-L_1}^{L_2} \left[H(p_l) - H(d_l) \right], \tag{22}$$

for any $L_1, L_2 > 0$ and d_l such that $\mathbb{E}[|X - \hat{X}|] \le D$.

The absolute value error distortion $\mathbb{E}[|X - \hat{X}|]$ cannot be written as a simple weighted sum of Hamming distortions from each level. In fact, we have to use an induction method to characterize the complicated relation. Denote $\mathcal{D}_k \triangleq \mathbb{E}\left[\left|\sum_{l=-L_1}^{k} 2^l (X_l - \hat{X}_l)\right|\right]$ for any $-L_1 \le k \le L_2$, which represents the accumulative absolute value distortion up to level k.

- Initialization: at level $-L_1$, $\mathcal{D}_{-L_1} = 2^{-L_1} d_{-L_1}$.
- Induction: for levels $-L_1 + 1 \le k \le L_2$,

$$\begin{aligned}\mathcal{D}_k &= \mathcal{D}_{k-1}(1-d_k) + 2^k d_k \\ &\quad + \frac{d_k(1-2p_k)}{1-2d_k} \sum_{l=-L_1}^{k-1} \frac{2^l d_l (1-2p_l)}{1-2d_l}.\end{aligned} \tag{23}$$

To this end, the expansion based coding scheme can be clearly expressed as an optimization problem over variables $\{d_{-L_1}, \ldots, d_{L_2}\}$, but not convex. Here, we step back to heuristically choose the values of d_ls. More precisely, for an aiming distortion D, d_l is chosen as

$$d_l = \frac{1}{1 + e^{2^l/D}}. \tag{24}$$

Then, by Theorem 7 and the iterative algorithm to calculate the real distortion $\mathcal{D}_{L_2}$, the rate distortion pair $(R^{(1)}, D^{(1)})$ is achievable, where

$$R^{(1)} = 1 + \sum_{l=-L_1}^{L_2} \left[H(p_l) - H(d_l) \right], \quad D^{(1)} = \mathcal{D}_{L_2}.$$

Evidently, this coding scheme may not behave well at high distortion regime, since $R^{(1)}$ is at least 1 bit. For high distortion, precisely compressing the sign bit seems inefficient. To this end, a time sharing scheme is utilized to reduce the coding rate. More precisely, for any $\alpha \in [0,1]$, we compress α fraction of source sequences into codeword 0, then the following rate distortion pair is found to be achievable:

$$R^{(2)} = (1-\alpha)R^{(1)}, \quad D^{(2)} = (1-\alpha)D^{(1)} + \alpha/\lambda.$$

Here, we provide an upper bound on rate distortion gap of expansion coding scheme for Laplacian source.

Theorem 8. *For any $D \in [0, 1/\lambda]$, with a choice of d_l in* (24) *and $L_1, L_2 > -\log(\lambda D)$, the achievable rate distortion pairs $(R^{(1)}, D^{(1)})$ and $(R^{(2)}, D^{(2)})$ are within a constant (c bits) gap to Shannon rate distortion function, i.e.*

$$R^{(1)} - R(D^{(1)}) \le c, \quad R^{(2)} - R(D^{(2)}) \le c.$$

Proof: See [14]. ■

C. Numerical Results

Since the calculation of $D^{(1)}$ from d_ls is non-trivial, it is hard to characterize the extent to which the overall distortion is overestimated by the bound. Using numerical analyses, we find this gap to be 0.52 bits in the low distortion regime (shown in Fig. 2).

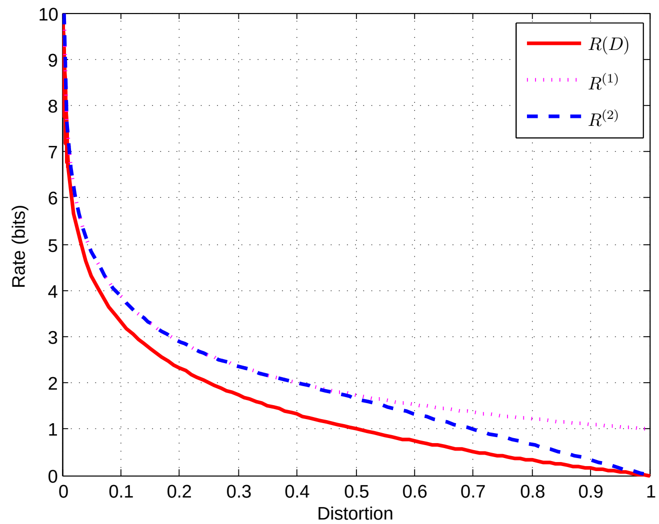

Fig. 2: **Achievable rate distortion pairs using expansion coding for Laplacian source with the absolute value error distortion.** We set $\lambda = 1$, and $L_1 = L_2 = -\log D + 5$. $R(D)$ (red-solid) is the rate distortion limit; $R^{(1)}$ (purple-dotted) is the achievable rates curve using expansion coding: and $R^{(2)}$ (blue-dashed) is the achievable rates curve using expansion coding and time sharing.

V. Discussion

Expansion coding enables the construction of "good" lossy compression codes for exponential and Laplacian sources using parallel discrete-valued source codes. Theoretical analyses and numerical results illustrate that expansion coding performs within a constant gap to the rate distortion limit, and therefore, approaches the limit in ratio, in the low distortion regime.

One significant benefit from expansion coding is the coding complexity. As indicated in the theoretical analysis and simulation results, approximately $-2\log(\lambda D)$ number of levels are sufficient for the coding scheme. Thus, by choosing "good" low complexity codes within each level (such as source coding with polar codes [4], [6]), the overall complexity of the proposed coding scheme can be easily characterized, resulting in a low-complexity code for the original continuous-valued source coding problem.

Although the paper focuses primarily on binary expansion case, our results can be generalized to q-ary expansion case, with similar performance guarantees. Moreover, we focus on exponential and Laplacian sources due to their decomposable property. As we can imagine, all decomposable distributions can be treated in a similar way. Even for indecomposable distributions, such as Gaussian, the expansion coding scheme presents a means of developing low-complexity coding schemes, although with a larger gap due to dependence between levels. The study on applications of expansion coding for general sources will be reported elsewhere.

References

[1] A. J. Viterbi and J. K. Omura, "Trellis encoding of memoryless disctrete-time sources with a fidelity criterion," *IEEE Trans. on Inf. Theory*, vol. 20, no. 3, pp. 325–332, May 1974.

[2] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by LDPC codes," *IEEE Trans. on Inf. Theory*, vol. 49, no. 9, pp. 2225–2229, Sept. 2003.

[3] M. J. Wainwright, E. Maneva, and E. Martinian, "Lossy source compression using low-density generator matrix codes: analysis and algorithms," *IEEE Trans. on Inf. Theory*, vol. 56, no. 3, pp. 1351–1368, Mar. 2010.

[4] E. Arıkan, "Channel polarization: A Method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. on Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[5] E. Arıkan, "Source polarization," in *Proc. 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, Austin, Texas, Jun. 2010.

[6] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. on Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.

[7] R. Zamir, "Lattices are everywhere," in *Proc. 2009 IEEE Information Theory and Applications Workshop (ITA 2009)*, San Diego, California, Feb. 2009.

[8] R. G. Gallager, *Information theory and reliable communication*. Wiley, 1968.

[9] S. Verdú, "The exponential distribution in information theory," *Problems of Information Transmission*, vol. 32, no. 1, pp. 86–95, Jan. 1996.

[10] T. Berger, "Optimum quantizers and permutation codes," *IEEE Trans. on Inf. Theory*, vol. 18, pp. 759–765, Nov. 1972.

[11] A. Gersho and R. M. Gray, *Vector quantization and signal compression.* Springer, 1993.

[12] D. Baron, and T. Weissman, "An MCMC approach to universal lossy compression of analog sources," *IEEE Trans. on Signal Processing*, vol. 60, no. 10, pp. 5230–5240, Oct. 2012.

[13] O. O. Koyluoglu, K. Appaiah, H. Si, and S. Vishwanath, "Expansion coding: Achieving the capacity of an AEN channel," in *Proc. 2012 IEEE International Symposium on Information Theory (ISIT 2012)*, Boston, Massachusetts, Jul. 2012.

[14] H. Si, O. O. Koyluoglu, and S. Vishwanath, "Lossy compression of exponential and laplacian sources using expanision coding," arXiv:1308.2338.

[15] G. Marsaglia, "Random variables with independent binary digits," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1922–1929, Dec. 1971.

[16] W. H. Equitz, and T. M. Cover, "Successive reinforcement of information," *IEEE Trans. on Inf. Theory*, vol. 37, no. 2, pp. 269–274, Mar. 1991.

Appendix A
Proof to Theorem 5

Without loss of generality, we assume $\lambda = 1$ for simplicity in the proof. By noting that d_l is also the parameter of expanded exponential distribution at level l, but with a different mean, we have

$$d_l = \frac{1}{1+e^{2^l/D}} = \frac{1}{1+e^{2^{l+\gamma}}} = p_{l+\gamma}, \tag{25}$$

where $\gamma \triangleq -\log D$. This result shows values of d_l are right-shifted version of p_l by γ positions. Using this fact, we have

$$\begin{aligned}\sum_{l=-L_1}^{L_2} [H(p_l) - H(d_l)] &= \sum_{l=-L_1}^{L_2} H(p_l) - \sum_{l=-L_1+\gamma}^{L_2+\gamma} H(p_l)\\ &\le \sum_{l=-L_1}^{L_1+\gamma-1} H(p_l) \le \gamma = -\log D. \end{aligned}\tag{26}$$

Moreover, it could be proved that [14], for $l \le -\gamma$,

$$H(d_l) - (1-p_l+d_l)H\left(\frac{d_l}{1-p_l+d_l}\right) \le \log e \cdot 2^{l+\gamma},$$

and for $l \le -\gamma$,

$$H(d_l) - (1-p_l+d_l)H\left(\frac{d_l}{1-p_l+d_l}\right) \le \log e \cdot 2^{-l-\gamma+1}.$$

Hence, combining these with (26), we have

$$\begin{aligned}R^{(1)} &= \sum_{l=-L_1}^{L_2}\left[H(p_l) - (1-p_l+d_l)H\left(\frac{d_l}{1-p_l+d_l}\right)\right]\\ &\le \sum_{l=-L_1}^{L_2} [H(p_l) - H(d_l)]\\ &\quad + \sum_{l=-L_1}^{-\gamma} \log e\cdot 2^{l+\gamma} + \sum_{-\gamma+1}^{L_2} \log e \cdot 2^{-l-\gamma+1}\\ &\le -\log D + 4\log e. \end{aligned}\tag{27}$$

Finally, using (10) and the convexity of log function, we obtain

$$R(D) \le R(D^{(1)}) + \log e\cdot(2^{-L_2}+2^{-L_1})/D \le R(D^{(1)}) + 2\log e,$$

where theorem assumptions $L_1, L_2 > -\log D$ are utilized in the last inequality. Relating this to (27), we have

$$R^{(1)} \le R(D) + 4\log e \le R(D^{(1)}) + 6\log e.$$

For the other part of the theorem, observe that

$$H\left(\frac{d_l}{1-2p_l+2d_l}\right) \ge (1-p_l+d_l)H\left(\frac{d_l}{1-p_l+d_l}\right).$$

Hence, for any $-L_1 \le l \le L_1$, we have $\bar{R}_l \le R_l$. Thus, by noting $R^{(2)}$ is a convex combination of $\bar{R}_l$ and R_l at each level, we have $R^{(2)} \le R^{(1)}$. Combing with the observation that $D^{(1)} = D^{(2)}$, we have $R^{(2)} \le R(D^{(2)}) + 6\log e$.