

# Refinements and Generalizations of the Shannon Lower Bound via Extensions of the Kraft Inequality

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering  
Technion - Israel Institute of Technology  
Technion City, Haifa 32000, ISRAEL  
E-mail: `merhav@ee.technion.ac.il`

## Abstract

We derive a few extended versions of the Kraft inequality for lossy compression, which pave the way to the derivation of several refinements and extensions of the well known Shannon lower bound in a variety of instances of rate-distortion coding. These refinements and extensions include sharper bounds for one-to-one codes and  $D$ -semifaithful codes, a Shannon lower bound for distortion measures based on sliding-window functions, and an individual-sequence counterpart of the Shannon lower bound.

## 1 Introduction

The Shannon lower bound (SLB) is one of the most important analytic tools in rate-distortion theory because it provides a simple, explicit, and often very tight lower bound to the rate-distortion function for a wide class of sources and distortion measures, see, e.g., Subsection 3.4.1 of [1], Sections 4.3 and 4.6 of [2], and Problem 10.6 of [3]. Its significance lies in giving a universal benchmark that connects rate-distortion tradeoffs to the entropy or the differential entropy of the source (depending on whether the source has a discrete or continuous alphabet), thereby offering an intuitively transparent approximation in regimes where exact evaluation of the rate-distortion function is intractable. The SLB is particularly powerful at low distortion, where it frequently coincides with the exact rate-distortion function for smooth sources, and it serves as a foundation for many refinements and asymptotic approximations (e.g., high-resolution analysis, corrections for lattice quantizers, and recent non-asymptotic bounds). In fact, the literature contains reported

results on the asymptotic tightness of the SLB in the limit of low distortion under fairly mild regularity conditions [4], [5]. Because computing the rate-distortion function exactly is generally difficult, the SLB plays a central role in analysis, design, and performance assessment of lossy compression schemes. More recent studies include the finite block-length regime [6], [7] and further developments concerning the quadratic distortion function [8].

In [9] Campbell derived an extension of Kraft's inequality, that leads to the SLB in both the discrete- and the continuous alphabet cases. However, his results were claimed to apply to rate-distortion codes whose distortion level is defined by the distance between the two most distant source-space vectors that are mapped to the same codeword, namely, the diameter (as opposed to the radius) of the distortion ball centered at the reproduction vector. In [10], a similar Kraft inequality was derived for  $D$ -semifaithful codes, namely, codes that incur per-letter distortion that never exceeds  $D$  in the usual sense. An additional benefit of the derivation in [10] was that it could deliver also an  $O\left(\frac{\log n}{n}\right)$  redundancy term on top of the SLB at least for certain distortion measures for which there is an explicit expression of the cardinality (or the volume, in the continuous case) of a ball of normalized radius  $D$  in the source vector space.

In this work, we propose a few other extended versions of Kraft's inequality that together pave the way to several further refinements and generalizations of the SLB. In these extensions of Kraft's inequality, the idea is to upper bound the summation (or the integral, in the continuous case) of an exponentiated negative linear combination of the code-length and the distortion incurred by each and every vector in the source space. By contrasting an upper bound with a lower bound to this quantity, we obtain several refinements and extensions to the SLB, which apply to the following scenarios.

1. *One-to-one codes.* Instead of assuming that the reproduction vectors are represented by uniquely decodable (UD) codes, we relax this restriction and allow any one-to-one code in the level of  $n$ -vectors. The lower bound then becomes the SLB minus an  $O\left(\frac{\log n}{n}\right)$  term, similarly as in the lossless case derived by Rissanen [11].

2.  *$D$ -semifaithful codes.* Similarly as in [10], we consider  $D$ -semifaithful codes, but here we allow

several simultaneous distortion criteria. Using saddle-point integration, we show that the resulting lower bound is given by the SLB plus  $\frac{k \log n}{2n} + o\left(\frac{1}{n}\right)$ , where  $k$  is the “effective number” of distortion constraints. The concrete meaning of this term will be provided in the sequel.

3. *Sliding-window distortion functions.* In some applications, one might wish to shape the spectrum or the memory properties of the reconstruction error signal. This can be done by imposing additional distortion constraints defined by additive functions that operate on two or more consecutive samples of the error signal in sliding-window fashion. Our framework is capable of incorporating such distortion functions and allowing a derivation of generalized SLB for this case.

4. *Individual sequences and finite-state encoders.* By developing a generalized Kraft inequality, similar (but not identical) to the one by Ziv and Lempel [12] for finite-state encoders, we derive also an individual-sequence counterpart of the SLB, where the source entropy term is replaced by the Lempel-Ziv complexity.

As is well known, the classical SLB can be obtained significantly more simply and easily than going via the Kraft inequality. In particular, it is obtained by a straightforward direct manipulation of the mutual information. But it should be emphasized that the point of this article is not in a quest for a simpler proof of the SLB. The point is that the path that goes via the Kraft inequality leads to the above mentioned extensions and the refinements.

The outline of the remaining part of this article is as follows. In Section 2, we establish some notation conventions (Subsection 2.1) and provide elementary background on the SLB (Subsection 2.2). In Section 3, we present and prove our extended Kraft inequality in several variations. In Section 4, we derive corresponding lower bounds, first, for UD lossless compression of the reproduction data, then for one-to-one compression thereof (Subsection 4.1), and finally, for  $D$ -semifaithful codes (Subsection 4.2). In Section 5, we address the case of sliding-window distortion functions. Finally, in Section 6, we first provide some background on finite-state compression of individual sequences and the LZ algorithm (Subsection 6.1) and then derive an individual-sequence counterpart of the SLB (Subsection 6.2).

## 2 Notation Conventions and Background

### 2.1 Notation Conventions

Throughout this paper, scalar random variables (RV's) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and their sample values, which will be denoted with same symbols superscripted by the dimension. Thus, for example,  $U^n$  ( $n$  – positive integer) will denote a random  $n$ -vector  $(U_1, \dots, U_n)$ , and  $u^n = (u_1, \dots, u_n)$  is a specific vector value in  $\mathcal{U}^n$ , the  $n$ -th Cartesian power of  $\mathcal{U}$ , which is the alphabet of each component of  $u^n$ . In some of our derivations below, there will be a need to refer to multiple copies of the vector  $u^n$ . In such cases, in order to avoid cumbersome subscripts and superscripts for indexing, we will use the alternative notation  $\mathbf{u}$  for  $u^n$ , and then the various copies will be denoted by  $\mathbf{u}_1, \mathbf{u}_2$ , etc. Returning to the first notation method,  $u_i^j$  and  $U_i^j$ , where  $i$  and  $j$  are integers and  $i \leq j$ , will designate segments  $(u_i, \dots, u_j)$  and  $(U_i, \dots, U_j)$ , respectively, where for  $i = 1$ , the subscript will be omitted (as above). For  $i > j$ ,  $u_i^j$  (or  $U_i^j$ ) will be understood as the null string. Logarithms and exponents, throughout this paper, will be understood to be taken to the base 2 unless specified otherwise. The indicator of an event  $\mathcal{A}$  will be denoted by  $\mathcal{I}\{\mathcal{A}\}$ , i.e.,  $\mathcal{I}\{\mathcal{A}\} = 1$  if  $\mathcal{A}$  occurs and  $\mathcal{I}\{\mathcal{A}\} = 0$  if not.

Sources and probability distributions associated with them will be denoted generically by the letter  $P$  subscripted by the name of the RV and its conditioning, if applicable, exactly like in ordinary textbook notation standards, e.g.,  $P_{U^n}(u^n)$  is the probability function of  $U^n$  at the point  $U^n = u^n$ ,  $P_{X|U^n}(x|u^n)$  is the conditional probability of  $X = x$  given  $U^n = u^n$ , and so on. Whenever clear from the context, these subscripts will be omitted. Information theoretic quantities, like entropies and mutual informations, will be denoted following the usual conventions of the information theory literature, e.g.,  $H(U^n)$ ,  $I(S; U^n|V^n)$ , and so on. The differential entropy of a continuous valued RV,  $U^n$ , will be denoted by  $h(U^n)$ . The expectation operator will be denoted by  $\mathbf{E}\{\cdot\}$  and the probability of an event  $\mathcal{A}$  will be denoted by  $\Pr\{\mathcal{A}\}$ .

It should be noted that our derivations will apply to both discrete-alphabet sources and to continuous alphabet sources. To avoid repetitions, we henceforth carry on under the assumption of a continuous alphabet source with the understanding that in the discrete-alphabet cases, integrations

over  $\mathcal{U}^n$  should simply be replaced by summations.

Let  $U^n = (U_1, U_2, \dots, U_n)$  denote a source vector, drawn from a stochastic process,  $P$ , whose alphabet is  $\mathcal{U}$ . The source vector is compressed by a lossy fixed-to-variable (F-V) source code, defined by an encoder  $\phi_n : \mathcal{U}^n \rightarrow \mathcal{B}_n \subset \{0, 1\}^*$  and a decoder  $\psi_n : \mathcal{B}_n \rightarrow \mathcal{V}_n \subseteq \mathcal{U}^n$ , where  $\mathcal{B}_n$  denotes a certain subset of the set of all binary variable-length strings,  $\{0, 1\}^*$ . Without loss of generality (and optimality), it is assumed that the encoder  $\phi_n$ , can be viewed as a cascade of a reproduction encoder (vector quantizer)  $\mathcal{U}^n \rightarrow \mathcal{V}_n$  and a uniquely decodable (UD) lossless code  $\mathcal{V}_n \rightarrow \mathcal{B}_n$ . In a certain part of our results this unique decodability assumption will be partially relaxed to become the less demanding assumption of a one-to-one mapping. Let  $L[\phi_n(u^n)]$  denote the length (in bits) of the compressed codeword,  $\phi_n(u^n)$ .

Assuming that  $\mathcal{U}$  is a group with certain addition and subtraction operations (e.g., modulo- $K$  addition/subtraction for  $\mathcal{U} = \{0, 1, \dots, K-1\}$  for a finite positive integer  $K$ , or ordinary addition/subtraction for  $\mathcal{U} = \mathbb{R}$ ) we will focus on additive difference distortion measures, where the distortion between the source vector  $u^n \in \mathcal{U}^n$  and the reproduction vector  $v^n = \psi_n(\phi_n(u^n)) \in \mathcal{V}_n$  will be given by

$$d(u^n, v^n) = \sum_{i=1}^n d(u_i, v_i) = \sum_{i=1}^n \rho(u_i - v_i), \quad (1)$$

where  $\rho(z)$ ,  $z \in \mathcal{U}$  is a certain non-negative function, which vanishes if and only if  $z = 0$ , for example,  $\rho(z) = |z|$ , or  $\rho(z) = z^2$ , etc. For the sake of convenience, we will sometimes denote  $d(u^n, v^n)$  by  $\rho(u^n - v^n)$ , which for a given encoder-decoder pair, is also  $\rho(u^n - \psi_n(\phi_n(u^n)))$ . Given an encoder-decoder pair,  $(\phi_n, \psi_n)$ , the expected distortion,  $\sum_{i=1}^n \mathbf{E}\{\rho(U_i - V_i)\}$  will be constrained to be less than or equal to  $nD$ , where  $D > 0$  designates the per-letter distortion level allowed. In certain parts of our derivations, the more restrictive pointwise distortion constraint will be imposed, i.e.,  $\rho(u^n - \psi_n(\phi_n(u^n))) \leq nD$  for all  $u^n \in \mathcal{U}^n$ . In other parts, more than one difference distortion measure will play a role, and accordingly, more than one distortion constraint will be imposed. Given  $k$  difference distortion functions,  $\rho_j(\cdot)$ ,  $j = 1, 2, \dots, k$ , we then require  $\mathbf{E}\{\rho_j(U^n - \psi_n(\phi_n(U^n)))\} \leq nD_j$  (or  $\max_{u^n \in \mathcal{U}^n} \rho_j(u^n - \psi_n(\phi_n(u^n))) \leq nD_j$ ) for all  $j = 1, 2, \dots, k$ . In another part of our results, we also allow distortion functions to be additive sliding-window functions operating on  $m$  consecutive symbols of the difference  $z^n = u^n - v^n$ . I.e.,

$$\rho(z^n) = \sum_{i=m}^n \rho(z_{i-m+1}^i). \quad (2)$$

This corresponds to situations where we wish to shape not only the ‘intensity’ of the error signal,  $z^n$ , but also its memory properties, for example, the correlations between consecutive symbols of  $z^n$ . We will elaborate more on this in Section 5.

Similarly as with the  $u^n$ , for which we adopt the alternative notation  $\mathbf{u}$ , the same will apply to  $v^n$ , which will be denoted also by  $\mathbf{v}$ , along with its multiple copies  $\mathbf{v}_1$ ,  $\mathbf{v}_2$ , and so on.

## 2.2 Background

For a continuous-alphabet memoryless source  $P$ , the SLB is given by

$$R(D) \geq R_{\text{SLB}}(D) \triangleq h(U) - \Phi(D), \quad (3)$$

where  $h(U)$  is the differential entropy of a single symbol  $U$  and

$$\Phi(D) \triangleq \sup_{\{Z: \mathbf{E}\{\rho(Z)\} \leq D\}} h(Z) = \inf_{\beta \geq 0} \left\{ \beta D + \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right] \right\}, \quad (4)$$

assuming that  $\int_{\mathcal{U}} 2^{-\beta \rho(z)} dz < \infty$  for some  $\beta > 0$ . The equivalence between the two expressions of  $\Phi(D)$  can be easily shown using standard techniques. For the sake of completeness, we prove this equivalence in the appendix (see also Section 4.3.1 in [1] and Sections 4.3 and 4.6 in [2]). The advantage of the second formula of  $\Phi(D)$  is that it involves optimization over one parameter only, as opposed to variational calculus in the first formula. Clearly, if the source is discrete rather than continuous, the differential entropy,  $h(U)$ , should be replaced by ordinary entropy,  $H(U)$ , and the integration over  $\mathcal{U}$  should be replaced by summation, as indicated above.

For a source with memory, the SLB is given by

$$R_n(D) \triangleq \min_{\{P_{V^n|U^n}: \mathbf{E}\{\rho(U^n - V^n)\} \leq nD\}} \frac{I(U^n; V^n)}{n} \geq \frac{h(U^n)}{n} - \Phi(D), \quad (5)$$

which in the limit of large  $n$  becomes

$$R(D) \geq \lim_{n \rightarrow \infty} \frac{h(U^n)}{n} - \Phi(D) \triangleq \bar{h}(U^\infty) - \Phi(D), \quad (6)$$

where  $\bar{h}(U^\infty)$  is the differential entropy rate of the source. In all cases, the function  $\Phi(D)$  remains as in (4).

### 3 Extended Kraft Inequalities

For a given encoder-decoder pair,  $(\phi_n, \psi_n)$ , and parameters  $\alpha > 1$  and  $\beta \geq 0$ , define the *extended Kraft integral* (or, the extended Kraft sum, in the discrete-alphabet case) as:

$$Z^n(\alpha, \beta) \triangleq \int_{\mathcal{U}^n} \exp_2\{-\alpha L[\phi_n(u^n)] - \beta \rho(u^n - \psi_n(\phi_n(u^n)))\} du^n. \quad (7)$$

*Lemma 1.* Let  $(\phi_n, \psi_n)$  induce a UD lossless compression of  $v^n = \psi_n(\phi_n(u^n))$ . Then, for every  $\alpha > 1$  and  $\beta \geq 0$ ,

$$Z^n(\alpha, \beta) \leq \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right]^n. \quad (8)$$

*Proof of Lemma 1.* Let us examine the expression of  $[Z_n(\alpha, \beta)]^k$  for an arbitrary positive integer  $k$ .

$$\begin{aligned} [Z^n(\alpha, \beta)]^k &= \left[ \int_{\mathcal{U}^n} \exp_2\{-\alpha L[\phi_n(\mathbf{u})] - \beta \rho(\mathbf{u} - \psi_n(\phi_n(\mathbf{u})))\} d\mathbf{u} \right]^k \\ &= \int_{\mathcal{U}^n} d\mathbf{u}_1 \cdots \int_{\mathcal{U}^n} d\mathbf{u}_k \cdot \exp_2 \left\{ -\alpha \sum_{i=1}^k L[\phi_n(\mathbf{u}_i)] - \beta \sum_{i=1}^k \rho(\mathbf{u}_i - \psi_n(\phi_n(\mathbf{u}_i))) \right\} \\ &\leq \sum_{\ell=1}^{\infty} \sum_{\{\{\mathbf{v}_i\}_{i=1}^k: \sum_{i=1}^k L(\mathbf{v}_i) = \ell\}} \int_{\{\mathbf{u}_1: \psi_n(\phi_n(\mathbf{u}_1)) = \mathbf{v}_1\}} d\mathbf{u}_1 \cdots \int_{\{\mathbf{u}_k: \psi_n(\phi_n(\mathbf{u}_k)) = \mathbf{v}_k\}} d\mathbf{u}_k \times \\ &\quad \exp_2 \left\{ -\alpha \ell - \beta \sum_{i=1}^k \rho(\mathbf{u}_i - \mathbf{v}_i) \right\} \\ &\leq \sum_{\ell=1}^{\infty} 2^{-\alpha \ell} \sum_{\{\{\mathbf{v}_i\}_{i=1}^k: \sum_{i=1}^k L(\mathbf{v}_i) = \ell\}} \int_{\mathcal{U}^n} dz_1 \cdots \int_{\mathcal{U}^n} dz_k \exp_2 \left\{ -\beta \sum_{i=1}^k \rho(\mathbf{z}_i) \right\} \\ &\leq \sum_{\ell=1}^{\infty} 2^{-\alpha \ell} \cdot 2^\ell \left[ \int_{\mathcal{U}^n} dz^n \exp_2\{-\beta \rho(z^n)\} \right]^k \\ &= \left[ \int_{\mathcal{U}^n} dz^n \exp_2\{-\beta \rho(z^n)\} \right]^k \cdot \sum_{\ell=1}^{\infty} 2^{-(\alpha-1)\ell} \\ &= \frac{[\int_{\mathcal{U}} dz \exp_2\{-\beta \rho(z)\}]^{nk}}{2^{\alpha-1} - 1}, \end{aligned} \quad (9)$$

and so,

$$Z^n(\alpha, \beta) \leq \frac{[\int_{\mathcal{U}} dz \exp_2\{-\beta \rho(z)\}]^n}{(2^{\alpha-1} - 1)^{1/k}}, \quad (10)$$

which upon taking the limit  $k \rightarrow \infty$ , becomes

$$Z^n(\alpha, \beta) \leq \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right]^n, \quad (11)$$

completing the proof of Lemma 1.

For one-to-one codes, instead of taking the limit of  $k \rightarrow \infty$  in the proof of Lemma 1, we simply set  $k = 1$ , since the one-to-one property is merely imposed in the level of a single  $n$ -block rather than in the level of concatenations of blocks, as in UD codes. This yields the following variation of Lemma 1.

*Lemma 2.* Let  $(\phi_n, \psi_n)$  induce a one-to-one mapping between  $\psi_n(u^n)$  and  $v^n = \psi_n(\phi_n(u^n))$ . Then, for every  $\alpha > 1$  and  $\beta \geq 0$ ,

$$Z_{1-1}^n(\alpha, \beta) \leq \frac{[\int_{\mathcal{U}} 2^{-\beta \rho(z)} dz]^n}{2^{\alpha-1} - 1}, \quad (12)$$

where  $Z_{1-1}^n(\alpha, \beta)$  is defined exactly as  $Z^n(\alpha, \beta)$ , except that it may apply to the larger class of one-to-one codes, rather than UD codes.

When  $\alpha \in (1, 2)$ , the denominator,  $2^{\alpha-1} - 1$ , is smaller than unity, and then the upper bound to  $Z_{1-1}^n(\alpha, \beta)$  in Lemma 2 is larger than that of Lemma 1. Indeed, the interesting region for selecting the values of  $\alpha$  is in the vicinity of unity, where  $2^{\alpha-1} - 1 < 1$ .

Returning to the class of UD lossless encodings of  $v^n = \psi_n(\phi_n(u^n))$ , two additional variations of the above extended Kraft inequality can be considered. The first pertains to  $D$ -semifaithful codes, namely, codes for which the distortion is restricted to never exceed  $nD$ , pointwise, and not merely in expectation, that is,  $\max_{u^n \in \mathcal{U}^n} \rho(u^n - \psi_n(\phi_n(u^n))) \leq nD$ , where  $D$  is the allowed per-letter distortion. The second variation is associated with fixed-rate codes, i.e., codes for which  $L[\phi_n(u^n)] = nR$  for all  $u^n \in \mathcal{U}^n$ , where  $R > 0$  is the allowed coding rate.

For  $D$ -semifaithful codes, let us re-define the integrand of the extended Kraft integral by replacing the term  $\beta \rho(u^n - \psi_n(\phi_n(u^n)))$  at the exponent with the function  $W(\rho(u^n - \psi_n(\phi_n(u^n))) - nD)$ , where  $W[\cdot]$  is the infinite well function (IWF),

$$W(t) \triangleq \begin{cases} 0 & t \leq 0 \\ \infty & t > 0 \end{cases} \quad (13)$$

This causes the integrand of the extended Kraft integral to vanish wherever the distortion constraint



is violated, and then the extended Kraft integral becomes

$$Z_{\text{D-sf}}^n(\alpha) \triangleq \int_{\mathcal{S}_n(D)} \exp_2\{-\alpha L[\phi_n(u^n)]\} du^n, \quad (14)$$

where  $\mathcal{S}_n(D) = \{u^n : \rho(u^n - \psi_n(\phi_n(u^n))) \leq nD\}$ . In this case, a simple modification of the proof of Lemma 1 yields the following version of the extended Kraft inequality (see also [10]):

*Lemma 3.* Let  $(\phi_n, \psi_n)$  be a  $D$ -semifaithful code that comprises UD lossless encoding of  $v^n = \psi_n(\phi_n(u^n))$ . Then, for every  $\alpha > 1$ ,

$$Z_{\text{D-sf}}^n(\alpha) \leq \text{Vol}\{z^n : \rho(z^n) \leq nD\}. \quad (15)$$

Here too, if the UD property is replaced by one-to-one property, then the right-hand side should be divided by  $2^{\alpha-1} - 1$ .

Finally, for fixed-rate codes, let us replace the term  $L[\phi_n(u^n)]$  of the Kraft integrand by  $nR$  and return the second term therein to be  $\beta\rho(u^n - \psi_n(\phi_n(u^n)))$ . By similar manipulations of the proof of Lemma 1, we find that for fixed rate codes, the Kraft integral becomes

$$Z_{\text{fr}}^n(\alpha, \beta) \triangleq \int_{\mathcal{U}^n} \exp_2\{-\alpha nR - \beta\rho(u^n - \psi_n(\phi_n(u^n)))\} du^n, \quad (16)$$

with the following version of the extended Kraft inequality:

*Lemma 4.* Let  $(\phi_n, \psi_n)$  be a rate- $R$  fixed-rate code. Then, for every  $\alpha \geq 0$  and  $\beta \geq 0$ ,

$$Z_{\text{fr}}^n(\alpha, \beta) \leq 2^{n(1-\alpha)R} \cdot \left[ \int_{\mathcal{U}} 2^{-\beta\rho(z)} dz \right]^n. \quad (17)$$

## 4 Lower Bounds

To fix ideas, we first demonstrate how the classical SLB is obtained from Lemma 1. Consider an arbitrary encoder-decoder pair,  $(\phi_n, \psi_n)$ , in which the reproduction vector,  $v^n = \psi_n(\phi_n(u^n))$  is losslessly compressed by a UD code. Then, following Lemma 1, we have the following chain of inequalities:

$$\left[ \int_{\mathcal{U}} 2^{-\beta\rho(z)} dz \right]^n \stackrel{\text{(a)}}{\geq} Z_n(\alpha, \beta)$$

$$\begin{aligned}
&= \int_{\mathcal{U}^n} \exp_2\{-\alpha L[\phi_n(u^n)] - \beta \rho(u^n - \psi_n(\phi_n(u^n)))\} du^n \\
&= \int_{\mathcal{U}^n} P(u^n) \exp_2\{-\alpha L[\phi_n(u^n)] - \beta \rho(u^n - \psi_n(\phi_n(u^n))) - \log P(u^n)\} du^n \\
&= \mathbf{E}\{\exp_2\{-\alpha L[\phi_n(U^n)] - \beta \rho(U^n - \psi_n(\phi_n(U^n))) - \log P(U^n)\}\} \\
&\stackrel{(b)}{\geq} \exp_2[-\alpha \mathbf{E}\{L[\phi_n(U^n)]\} - \beta \mathbf{E}\{\rho(U^n - \psi_n(\phi_n(U^n)))\} - \mathbf{E}\{\log P(U^n)\}] \\
&= \exp_2[-\alpha \mathbf{E}\{L[\phi_n(U^n)]\} - \beta \mathbf{E}\{\rho(U^n - \psi_n(\phi_n(U^n)))\} + h(U^n)], \quad (18)
\end{aligned}$$

where (a) stems from Lemma 1 and (b) is due to Jensen's inequality applied to the convex function,  $f(t) = 2^t$ . It follows that

$$\alpha \mathbf{E}\{L[\phi_n(U^n)]\} + \beta \mathbf{E}\{\rho(U^n - \psi_n(\phi_n(U^n)))\} \geq h(U^n) - n \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right]. \quad (19)$$

Since this holds true for every  $\alpha > 1$  while the right-hand side is independent of  $\alpha$ , we may take the infimum of the left-hand side in the range  $\alpha > 1$  to obtain, after normalization by  $n$ ,

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} + \beta \cdot \frac{\mathbf{E}\{\rho(U^n - \psi_n(\phi_n(U^n)))\}}{n} \geq \frac{h(U^n)}{n} - \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right]. \quad (20)$$

At this point, there are several possible perspectives that can be adopted regarding eq. (20). The first is, of course, to view this as a lower bound to the Lagrangian of rate and distortion. The second is to impose an expected distortion constraint,  $\mathbf{E}\{\rho(U^n - \psi_n(\phi_n(U^n)))\} \leq nD$ , and then, after normalization by  $n$ , (20) would yield a lower bound to the expected rate according to

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \frac{h(U^n)}{n} - \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right] - \beta D. \quad (21)$$

Since this holds true for every  $\beta \geq 0$  while the left-hand side is independent of  $\beta$ , we may maximize the right-hand side over  $\beta \geq 0$ , to obtain

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \frac{h(U^n)}{n} - \inf_{\beta \geq 0} \left\{ \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right] + \beta D \right\} = \frac{h(U^n)}{n} - \Phi(D), \quad (22)$$

thus recovering the classical SLB. In case of multiple distortion constraints, say,

$$\mathbf{E}\{\rho_j(U^n - \psi_n(\phi_n(U^n)))\} \leq nD_j, \quad j = 1, 2, \dots, k, \quad (23)$$

the above rate lower bound continues to apply, provided that  $\beta$ ,  $D$  and  $\rho(z)$  are redefined as  $k$ -dimensional vectors,  $\beta = (\beta_1, \dots, \beta_k)$ ,  $D = (D_1, \dots, D_k)$ , and  $\rho(z) = (\rho_1(z), \dots, \rho_k(z))$ , and

accordingly,  $\beta D$  and  $\beta \rho(z)$  are understood to be inner products. The infimum over  $\beta$  is then defined across  $[0, \infty)^k$ .

Returning to eq. (20) and to a single distortion criterion, we may alternatively apply a rate constraint  $\mathbf{E}\{L[\phi_n(U^n)]\} \leq nR$  and then obtain a lower bound to the expected distortion:

$$\begin{aligned} \frac{\mathbf{E}\{\rho(U^n - \psi_n(\phi_n(U^n)))\}}{n} &\geq \sup_{\beta \geq 0} \frac{1}{\beta} \left( \frac{h(U^n)}{n} - \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right] - R \right) \\ &= \sup_{\gamma \geq 0} \gamma \left( \frac{h(U^n)}{n} - \log \left[ \int_{\mathcal{U}} 2^{-\rho(z)/\gamma} dz \right] - R \right), \end{aligned} \quad (24)$$

which is the distortion-rate counterpart of the SLB.

Having established the classical SLB for general variable-rate codes with UD lossless compression of the reproduction vectors, we now carry on to derive some refinements and extensions associated with the other types of codes that we mentioned. The idea would be to apply the same chain of inequalities as in (18), but to invoke Lemma 2 or Lemma 3, or Lemma 4, according to the relevant class of codes, instead of Lemma 1. We next implement this plan for one-to-one codes and for  $D$ -semifaithful codes. The same methodology can be applied to fixed-rate codes, but will not delve into this here.

#### 4.1 One-to-One Codes

For one-to-one codes, we repeat the same derivation as in (18) by invoking Lemma 2 instead of Lemma 1. This yields the following modified version of eq. (19):

$$\alpha \mathbf{E}\{L[\phi_n(U^n)]\} + \beta \mathbf{E}\{\rho(U^n - \psi_n(\phi_n(U^n)))\} \geq h(U^n) - n \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right] + \log(2^{\alpha-1} - 1). \quad (25)$$

Applying the the average distortion constraint, optimizing over  $\beta$ , and normalizing by  $n$ , we obtain

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \sup_{\alpha \geq 1} \left\{ \frac{h(U^n)/n - \Phi(D)}{\alpha} + \frac{\log(2^{\alpha-1} - 1)}{\alpha n} \right\}. \quad (26)$$

The maximization of the right-hand side w.r.t.  $\alpha$  does not seem to lend itself to a closed form expression, but by selecting  $\alpha = \alpha_n = 1 + c \frac{\log n}{n}$  ( $c$  being an arbitrary positive constant), it is readily seen that the resulting lower bound becomes

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \frac{h(U^n)}{n} - \Phi(D) - O\left(\frac{\log n}{n}\right), \quad (27)$$

which is in agreement with the subtraction of an  $O((\log n)/n)$  below the entropy for lossless one-to-one codes [11]. This reduction of  $O((\log n)/n)$  is due to the fact that the class of one-to-one codes is broader than the class of UD codes, and therefore, the former codes are potentially more capable for a given finite  $n$ , albeit the difference fades away as  $n$  grows.

## 4.2 $D$ -Semifaithful Codes

For  $D$ -semifaithful codes with UD correspondence between the compressed bit-stream and the reproduction, we obtain from Lemma 3, combined with a derivation like (18):

$$\mathbf{E}\{L[\phi_n(U^n)]\} \geq h(U^n) - \log \text{Vol}\{\mathcal{S}_n(D)\}, \quad (28)$$

which is line with the results of [10], and the remaining issue becomes the assessment of the log-volume of  $\mathcal{S}_n(D)$ , or an evaluation of a good upper bound to this quantity. One way to proceed is to apply a Chernoff bound, as was suggested by Campbell [9]. For  $\mathcal{U} = \mathbb{R}$ , this amounts to the following derivation:

$$\begin{aligned} \log \text{Vol}\{\mathcal{S}_n(D)\} &= \log \text{Vol}\{\mathbf{z} : \rho(\mathbf{z}) \leq nD\} \\ &= \log \left[ \int_{\mathbb{R}^n} \mathcal{I}\{\rho(\mathbf{z}) \leq nD\} d\mathbf{z} \right] \\ &\leq \log \left( \inf_{\beta \geq 0} \int_{\mathbb{R}^n} \exp_2\{\beta[nD - \rho(\mathbf{z})]\} d\mathbf{z} \right) \\ &= n \cdot \inf_{\beta \geq 0} \left\{ \beta D + \log \left( \int_{\mathbb{R}} 2^{-\beta \rho(z)} dz \right) \right\} \\ &= n\Phi(D), \end{aligned} \quad (29)$$

and we are back to the ordinary SLB,

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \frac{h(U^n)}{n} - \Phi(D), \quad (30)$$

exactly as we had for UD lossless compression of the reproduction data, and once again, this derivation extends straightforwardly to the case of multiple simultaneous distortion constraints by considering  $\beta$ ,  $D$ , and  $\rho(\cdot)$  to be vectors rather than scalars, as mentioned before.

However, since the class of  $D$ -semifaithful codes is narrower (and hence more limited to a certain extent) than the class of codes that merely comply with an average distortion constraint, it is conceivable to expect a somewhat tighter (larger) lower bound. Indeed, this turns out to be

the case if the log-volume of  $\mathcal{S}_n(D)$  is estimated using a more sophisticated analysis tool, namely, the saddle-point method (see, e.g., Chapter 5 in [13] and Chapter 3 in [14] as well as its extension to the multivariate case [15]).

To apply the saddle-point method, the idea is to represent the indicator function at the integrand of the second line in (29) as  $\mathcal{I}\{\rho(\mathbf{z}) \leq nD\} = u(nD - \rho(\mathbf{z}))$ , where  $u(t)$  is the unit step function, defined as

$$u(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \quad (31)$$

which in turn is represented as the inverse Laplace transform of the complex function  $U(s) = \frac{1}{s}$ , i.e.,

$$u(t) = \frac{1}{2\pi i} \int_{\text{Re}\{s\}=c} \frac{e^{st} ds}{s}, \quad (32)$$

where  $i \triangleq \sqrt{-1}$  and  $c$  is an arbitrary positive real. It follows that

$$\begin{aligned} \text{Vol}\{S_n(D)\} &= \int_{\mathbb{R}^n} u(nD - \rho(\mathbf{z})) d\mathbf{z} \\ &= \int_{\mathbb{R}^n} \frac{d\mathbf{z}}{2\pi i} \int_{\text{Re}\{s\}=c} \frac{e^{s(nD - \rho(\mathbf{z}))} ds}{s} \\ &= \frac{1}{2\pi i} \int_{\text{Re}\{s\}=c} \frac{e^{snD} ds}{s} \int_{\mathbb{R}^n} e^{-s\rho(\mathbf{z})} d\mathbf{z} \\ &= \frac{1}{2\pi i} \int_{\text{Re}\{s\}=c} \frac{e^{snD} ds}{s} \int_{\mathbb{R}^n} \exp\left\{-s \sum_{t=1}^n \rho(z_t)\right\} d\mathbf{z} \\ &= \frac{1}{2\pi i} \int_{\text{Re}\{s\}=c} \frac{e^{snD} ds}{s} \left[ \int_{\mathbb{R}} e^{-s\rho(z)} dz \right]^n \\ &= \frac{1}{2\pi i} \int_{\text{Re}\{s\}=c} \frac{ds}{s} \exp\left\{n \left[ sD + \ln \left( \int_{\mathbb{R}} e^{-s\rho(z)} dz \right) \right]\right\}. \end{aligned} \quad (33)$$

This path integral in the complex plane complies with the general form  $\int_A^B g(s) e^{nf(s)} ds$ , which under certain regularity conditions, can be approximated for large  $n$  (see, e.g., eq. (5.7.2) in [13]) according to:

$$\int_A^B g(s) e^{nf(s)} ds = e^{i\theta} \sqrt{\frac{2\pi}{n|f''(s_*)|}} \cdot g(s_*) e^{nf(s_*)} \cdot \left[ 1 + O\left(\frac{1}{n}\right) \right], \quad (34)$$

provided that the functions  $f$  and  $g$  are independent of  $n$  and analytic within some connected region  $\mathcal{D}$  that includes  $A$  and  $B$  (which are also independent of  $n$ ). Here,  $s_* \in \mathcal{D}$  is a saddle-point, i.e., a point where  $f'(s_*) = 0$ ,  $f''(s_*) > 0$  and  $g(s_*) \neq 0$ . The angle  $\theta$  is called the axis and is given by  $\theta = (\pi - \arg\{f''(s_*)\})/2$ . In our case,  $f(s) = sD + \ln \left[ \int_{\mathbb{R}} e^{-s\rho(z)} dz \right]$ ,  $g(s) = \frac{1}{s}$ ,  $\theta = \frac{\pi}{2}$ , and  $s_*$  is the

point at which  $f'$  vanishes, which is assumed strictly positive. But this point of zero derivative is also the point that minimizes the convex function  $f$  across the positive reals. It follows then that  $\text{Vol}\{S_n(D)\}$  can be approximated as follows:

$$\text{Vol}\{S_n(D)\} = \frac{1}{s_\star \sqrt{2\pi n |f''(s_\star)|}} \cdot \exp \left\{ n \left( s_\star D + \ln \left[ \int_{\mathbb{R}} e^{-s_\star \rho(z)} dz \right] \right) \right\} \cdot \left[ 1 + O \left( \frac{1}{n} \right) \right]. \quad (35)$$

As for the exponential factor, observe that

$$\begin{aligned} & \exp \left\{ n \left( s_\star D + \ln \left[ \int_{\mathbb{R}} e^{-s_\star \rho(z)} dz \right] \right) \right\} \\ &= \exp \left\{ n \cdot \inf_{s \geq 0} \left( s D + \ln \left[ \int_{\mathbb{R}} e^{-s \rho(z)} dz \right] \right) \right\} \\ &= \exp_2 \left\{ n (\log_2 e) \cdot \inf_{s \geq 0} \left( s D + \frac{1}{\log_2 e} \cdot \log_2 \left[ \int_{\mathbb{R}} 2^{-s \rho(z) \log_2 e} dz \right] \right) \right\} \\ &= \exp_2 \left\{ n \inf_{s \geq 0} \left( s D \log_2 e + \log_2 \left[ \int_{\mathbb{R}} 2^{-s \rho(z) \log_2 e} dz \right] \right) \right\} \\ &= \exp_2 \left\{ n \inf_{\beta \geq 0} \left( \beta D + \log_2 \left[ \int_{\mathbb{R}} 2^{-\beta \rho(z)} dz \right] \right) \right\} \\ &= 2^{n\Phi(D)}, \end{aligned} \quad (36)$$

and so,

$$\text{Vol}\{S_n(D)\} = \frac{2^{n\Phi(D)}}{s_\star \sqrt{2\pi n |f''(s_\star)|}} \cdot \left[ 1 + O \left( \frac{1}{n} \right) \right], \quad (37)$$

which yields

$$\frac{\log \text{Vol}\{S_n(D)\}}{n} = \Phi(D) - \frac{\log n}{2n} - O \left( \frac{1}{n} \right), \quad (38)$$

and then, following eq. (28), we end up with

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \frac{h(U^n)}{n} - \Phi(D) + \frac{\log n}{2n} + O \left( \frac{1}{n} \right). \quad (39)$$

We therefore observe that the SLB for  $D$ -semifaithful codes is given by the ordinary SLB plus redundancy whose leading term is  $\frac{\log n}{2n}$ . This is in agreement with findings of [10], where the derivations corresponded to special cases for which simple geometric and/or combinatorial considerations facilitated the accurate characterization of  $\log \text{Vol}\{S_n(D)\}$ , but here the conclusion is more general. Note that we could have been even more precise and specify also the  $O \left( \frac{1}{n} \right)$  term to be  $\frac{1}{n} \cdot \log[s_\star \sqrt{2\pi |f''(s_\star)|}] + O \left( \frac{1}{n^2} \right)$ , but this is, of course, less important.

When  $k$  simultaneous distortion constraints are imposed pointwise, i.e.,

$$\max_{u^n \in \mathbb{R}^n} \rho_j(u^n - \psi_n(\phi_n(u^n))) \leq nD_j, \quad j = 1, 2, \dots, k, \quad (40)$$

then using the multivariate saddle-point method [15], the above derivation extends with the vector version of the definition of  $\Phi(D)$  replacing the scalar one, and the pre-exponential factor of the saddle-point approximation becomes

$$\frac{(2\pi/n)^{k/2}}{Q\sqrt{\det\{\text{Hess}\{f\}|_{s_*}\}}},$$

where  $Q$  is the product of the components of the  $k$ -dimensional vector  $s_*$  – the point at which  $\nabla f(s) = 0$ , where  $f(s) = s \cdot D + \ln \left[ \int_{\mathbb{R}} e^{-s \cdot \rho(z)} dz \right]$  is defined such that  $s$ ,  $D$  and  $\rho(\cdot)$  are  $k$ -dimensional vectors with  $s \cdot D$  and  $s \cdot \rho(z)$  being inner products, as was defined before. Here,  $\det\{\text{Hess}\{f\}|_{s_*}\}$  is the determinant of the Hessian of  $f$ , computed at  $s = s_*$ . It is assumed, of course, that the  $k \times k$  matrix  $\text{Hess}\{f\}|_{s_*}$  is non-singular, otherwise there might be redundant (inactive) constraints, which should be removed from the calculation. In this case, the redundancy on top of the ordinary SLB is  $\frac{k' \log n}{2n} + O\left(\frac{1}{n}\right)$ , i.e.,

$$\frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \frac{h(U^n)}{n} - \Phi(D) + \frac{k' \log n}{2n} + O\left(\frac{1}{n}\right), \quad (41)$$

where  $k' \leq k$  is the effective dimension after the possible removal of redundant constraints.

For example, if  $k = 2$ ,  $\rho_1(z) = |z|$ , and  $\rho_2(z) = z^2$ , then obviously,  $\left(\frac{1}{n} \sum_{t=1}^n |z_t|\right)^2$  cannot exceed  $\frac{1}{n} \sum_{t=1}^n z_t^2$ , and so, if  $D_1 > \sqrt{D_2}$ , the constraint  $\sum_{t=1}^n |z_t| \leq nD_1$  is inactive (and hence removable) in the presence of the constraint  $\sum_{t=1}^n z_t^2 \leq nD_2$ . In this case, although  $k = 2$ , the effective dimension is  $k' = 1$ . Another obvious example of a superfluous distortion constraint occurs when there is a linear dependence. For instance, let  $k = 3$  and  $\rho_3(z) = a\rho_1(z) + b\rho_2(z)$ , where  $a$  and  $b$  are fixed positive reals. Then whenever  $D_3 \geq aD_1 + bD_2$ , the third distortion constraint is redundant and then  $k' = 2$  (or even less, if there are additional superfluous constraints). More generally, inactive constraints can be identified as those whose corresponding components of  $s_*$  vanish.

Another type of situations where the effective dimension  $k'$  is smaller than the formal dimension,  $k$ , occurs when the dependence of  $f(s)$  on some of the components of  $s$  disappears in the first place. Consider, for example, the case where  $k = 2$ ,  $\rho_1(z) = z^2$ , and the other distortion constraint is  $\max_{1 \leq t \leq n} |z_t| \leq A$ , in other words, we impose both an average quadratic constraint and a peak-limited distortion constraint with the motivation of avoiding large spikes in the reconstruction error signal. The peak-limited distortion constraint can be represented as an additive distortion constraint if we select  $\rho_2(z) = W(|z| - A)$ , where  $W(\cdot)$  is the IWF defined in (13) and the value

of  $D_2$  can then be selected to be an arbitrary finite, non-negative, real, say  $D_2 = 0$ . Here again  $k' = 1$ , but as said, this time, it is because the function  $f(s)$  depends only on one component of the vector  $s = (s_1, s_2)$  to begin with. To see this, observe that

$$\begin{aligned}
f(s) &= sD + \ln \left[ \int_{\mathbb{R}} e^{-s\rho(z)} dz \right] \\
&= s_1 D_1 + s_2 D_2 + \ln \left[ \int_{\mathbb{R}} e^{-s_1 \rho_1(z) + s_2 \rho_2(z)} dz \right] \\
&= s_1 D_1 + s_2 \cdot 0 + \ln \left[ \int_{\mathbb{R}} e^{-s_1 z^2 + s_2 W(|z| - A)} dz \right] \\
&= s_1 D_1 + \ln \left[ \int_{-A}^A e^{-s_1 z^2} dz \right] \\
&= s_1 D_1 + \ln \left( \sqrt{\frac{\pi}{s_1}} \cdot [1 - 2Q(A\sqrt{2s_1})] \right) \\
&= s_1 D_1 + \frac{1}{2} \ln \left( \frac{\pi}{s_1} \right) + \ln[1 - 2Q(A\sqrt{2s_1})], \tag{42}
\end{aligned}$$

where  $Q(\cdot)$  is the well-known  $Q$ -function, defined as

$$Q(t) = \int_t^\infty \frac{e^{-x^2/2} dx}{\sqrt{2\pi}}. \tag{43}$$

Here, although formally there are  $k = 2$  distortion constraints, the saddle-point integration is over one complex variable only, and so, the redundancy is  $\frac{\log n}{2n} + O\left(\frac{1}{n}\right)$  on top of the ordinary SLB,  $\frac{h(U^n)}{n} + \Phi(D)$ .

## 5 Sliding-Window Distortion Constraints

Consider a situation where one wishes to control not only the ‘intensity’ of the reconstruction error signal,  $\sum_{t=1}^n \rho(z_i)$ , but possibly to shape also its ‘continuity’ or its ‘smoothness’ by imposing additional limitations, say, on the empirical autocorrelations of the error signal,  $z^n = (z_1, z_2, \dots, z_n)$ , in order to suppress, for example, high frequencies, which might be disturbing for the human eye (in case of images and video streams) or the human ear (in case of audio signals). For instance, one might wish to constrain the error signal to obey the first lag autocorrelation constraint,  $\frac{1}{n} \sum_{t=2}^n z_t z_{t-1} \geq 0.95$  (and perhaps also further lags), in addition to the ordinary mean-square error constraint, say,  $\frac{1}{n} \sum_{t=1}^n z_t^2 \leq 0.01$ . More generally, consider the case where we impose  $k$  additive



constraints pertaining to sliding-window functions, of the form,

$$\sum_{t=m}^n \rho_j(z_{t-m+1}^t) \leq nD_j, \quad j = 1, 2, \dots, k, \quad (44)$$

where  $m$  is a positive integer that designates the size of the sliding window. For  $m = 1$ , we are back to ordinary additive distortion constraints, where the single-letter distortion function  $\rho_j$  operates on single symbols separately. If the sizes of the sliding window are different for the  $k$  various constraints, which take  $m$  to be the largest one. Accordingly, in the above example where  $k = 2$  and the constraints are  $\frac{1}{n} \sum_{t=1}^n z_t^2 \leq 0.01$  and  $\frac{1}{n} \sum_{t=2}^n z_t z_{t-1} \geq 0.95$ , we have  $\rho_1(z) = z^2$ ,  $D_1 = 0.01$ ,  $\rho_2(z_1, z_2) = -z_1 \cdot z_2$ , and  $D_2 = -0.95$ . In this case, the sliding-window size is  $m = 2$ , and the minus signs in the correlation constraint are due to the reversal of the direction of the inequality in  $\frac{1}{n} \sum_{t=2}^n z_t z_{t-1} \geq 0.95$ , as opposed to the direction of inequalities in our distortion constraints in general.

How does the SLB extend to the case of sliding-window constraints of this type? In this section, we assume that  $m$  is fixed while  $n$  tends to infinity, and we focus merely on the main terms of the resulting SLB, disregarding the redundancy terms.

A straightforward extension of Lemma 1 and the subsequent derivation in Section 4 so as to apply to a set of  $k$  sliding-window distortion constraints with window size  $m$ , yields the lower bound

$$\mathbf{E}\{L[\phi(U^n)]\} \geq h(U^n) - \inf_{\beta \in [0, \infty)^k} \left( \log \left[ \int_{\mathcal{U}^n} \exp_2 \left\{ -\beta \cdot \sum_{t=m}^n \rho(z_{t-m+1}^t) \right\} dz^n \right] + n\beta \cdot D \right), \quad (45)$$

where  $\beta = (\beta_1, \dots, \beta_k)$ ,  $D = (D_1, \dots, D_k)$ ,  $\rho(z_{t-m+1}^t) = (\rho_1(z_{t-m+1}^t), \dots, \rho_k(z_{t-m+1}^t))$ , and the dot operations are understood as inner products, as defined earlier.

The multi-dimensional integral

$$\int_{\mathcal{U}^n} \exp_2 \left\{ -\beta \cdot \sum_{t=m}^n \rho(z_{t-m+1}^t) \right\} dz^n = \int_{\mathcal{U}^n} \prod_{t=m}^n \exp_2 \{ -\beta \cdot \rho(z_{t-m+1}^t) \} dz^n \quad (46)$$

can be viewed as being obtained by iterated applications of a sliding-window integral operator whose kernel is given by  $K_\beta(z_{t-m+1}^t) = \exp_2 \{ -\beta \cdot \rho(z_{t-m+1}^t) \}$  and the integration is over one component of  $z^n$  at a time. Under certain regularity conditions, the value of this multidimensional integral grows exponentially with an exponential order of  $[\lambda(\beta)]^n$ , where  $\lambda(\beta)$  is the spectral radius of the operator kernel, namely, the dominant eigenvalue, and then the asymptotic form of the SLB

becomes

$$\liminf_{n \rightarrow \infty} \frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \bar{h}(U^\infty) - \inf_{\beta \in [0, \infty)^k} [\log \lambda(\beta) + \beta D]. \quad (47)$$

There are several equivalent formulas for calculating the spectral radius,  $\lambda(\beta)$ , of the of a sliding-window kernel  $K_\beta(\cdot)$  for a given  $\beta$ , for example, the Collatz-Wielandt formula [16], [17], and the Donsker-Varadhan formula [18] – for details, see, e.g., Subsection 3.4 of [19]. In certain special cases, such as those that involve a symmetric kernel for  $m = 2$ ,  $K_\beta(z, z')$ , the Rayleigh quotient formula

$$\lambda(\beta) = \sup_g \frac{\int_{U^2} g(z) K_\beta(z, z') g(z') dz dz'}{\int_{\mathcal{U}} g^2(z) dz}. \quad (48)$$

can also be used. Clearly, in the finite-alphabet case, the operator kernel reduces to a finite-dimensional matrix and then the spectral radius is simply the Perron-Frobenius eigenvalue (see, e.g., Theorem 8.2.11 of [20]).

As a simple example, consider the case  $k = m = 2$ ,  $\rho_1(z) = z^2$ ,  $D_1 = D$ ,  $\rho_2(z, z') = -z \cdot z'$ , and  $D_2 = -\theta$ , where  $D > 0$  and  $\theta \in [0, 1)$  are given parameters. Then, following the analogous derivation of Example 4 in [19], we find that the resulting SLB is given by

$$\liminf_{n \rightarrow \infty} \frac{\mathbf{E}\{L[\phi_n(U^n)]\}}{n} \geq \bar{h}(U^\infty) - \frac{1}{2} \log[2\pi e D(1 - \theta^2)], \quad (49)$$

or, in words, the rate penalty due to the additional autocorrelation constraint is  $\frac{1}{2} \log \frac{1}{1 - \theta^2}$  on top of the SLB pertaining to the ordinary quadratic distortion constraint,  $\frac{h(U^n)}{n} - \frac{1}{2} \log(2\pi e D)$ .

## 6 Individual Sequences and Finite-State Encoders

We conclude this article by deriving an individual-sequence counterpart of the SLB for finite-alphabet, deterministic sequences, which is based on a variation of Ziv and Lempel's generalized Kraft inequality (see Lemma 2 of [12]).

Generally speaking, our model for lossy compression of individual sequences is based on the following simple structure. Each source block of length  $m$ ,  $u^m \in \mathcal{U}^m$  ( $m$  – positive integer) is first mapped by an arbitrary reproduction encoder (or vector quantizer) into a reproduction vector,

$$v^m = q(u^m) \in \mathcal{V}_m \subseteq \mathcal{U}^m \quad (50)$$

and then, the concatenation of the resulting  $m$ -vectors,  $\{v^m\}$ , forming the sequence  $v_1, v_2, \dots$ , is compressed losslessly by a finite-state encoder following the model of [12].

To keep this article self-contained, we begin with some basic background on lossless compression of individual sequences by finite-state encoders and the 1978 Lempel-Ziv (LZ78) algorithm. Readers familiar with this background may safely skip Subsection 6.1 and move on directly to Subsection 6.2.

## 6.1 Background

Following the model of [12], consider a setting for lossless compression of  $v^n$  on the basis of finite-state (FS) encoders. An FS encoder is defined by the set  $E = (\mathcal{U}, \mathcal{Y}, \Sigma, f, g)$ , where:  $\mathcal{U}$  is the finite alphabet of each symbol,  $v_i$ , which is the same as the source alphabet, and whose size is  $r$ ;  $\mathcal{Y}$  is a finite collection of binary, variable-length strings, which is allowed to consist of empty string  $\lambda$  (whose length is zero);  $\Sigma$  is a set of  $s$  states of the encoder;  $f : \Sigma \times \mathcal{U} \rightarrow \mathcal{Y}$  is the output function, and  $g : \Sigma \times \mathcal{U} \rightarrow \Sigma$  is the next-state function. Given an infinite input reproduction vector (obtained by concatenating infinitely many output vectors from the reproduction encoder),  $\mathbf{v} = (v_1, v_2, \dots)$  with  $v_i \in \mathcal{U}$ ,  $i = 1, 2, \dots$ , the FS encoder  $E$  produces an infinite output sequence,  $\mathbf{y} = (y_1, y_2, \dots)$  with  $y_i \in \mathcal{Y}$ , henceforth referred to as the compressed bit-stream, while passing through a sequence of states  $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots)$  with  $\sigma_i \in \Sigma$ . The encoder is governed recursively by the equations:

$$y_i = f(\sigma_i, v_i), \quad (51)$$

$$\sigma_{i+1} = g(\sigma_i, v_i), \quad (52)$$

for  $i = 1, 2, \dots$ , with a fixed initial state  $\sigma_1 = \sigma_\star \in \Sigma$ . If at any step  $y_i = \lambda$ , this is referred to as idling as no output is generated, but only the state is kept updated in response to the input.

An encoder with  $s$  states, henceforth called an  $s$ -state encoder, is one for which  $|\Sigma| = s$ . For the sake of simplicity, we adopt a few notation conventions from [12]: Given a segment of input symbols  $v_i^j$  with  $i \leq j$  and an initial state  $\sigma_i$ , we use  $f(\sigma_i, v_i^j)$  to denote the corresponding output segment  $y_i^j$  produced by  $E$ . Similarly,  $g(\sigma_i, v_i^j)$  will denote the final state  $\sigma_{j+1}$  after processing the inputs  $v_i^j$ , beginning from state  $\sigma_i$ .

An FS encoder  $E$  is called information lossless (IL) if, given any initial state  $\sigma_i \in \Sigma$ , any positive integer  $n$ , and any input string,  $v_i^{i+n}$ , the set  $(\sigma_i, f(\sigma_i, v_i^{i+n}), g(\sigma_i, v_i^{i+n}))$  uniquely determines the corresponding input string  $v_i^{i+n}$ .

The incremental parsing process used by the LZ78 algorithm is a sequential procedure for processing a finite-alphabet input  $u^n$ . At each step of this process, one determines the shortest string that has not yet occurred as a complete phrase in the current parsed set with the possible exception of the last phrase, which might be incomplete. For example, applying this parsing method to the sequence

$$u^{15} = 011010011000100$$

yields

$$0,1,10,100,11,00,01,00.$$

Let us denote by  $c(u^n)$  the total number of distinct phrases generated by this procedure (here,  $c(u^{15}) = 8$ ). In addition, let  $LZ(u^n)$  represent the length in bits of the binary string produced by the LZ78 encoding of  $u^n$ . By Theorem 2 in [12], the following inequality holds:

$$LZ(u^n) \leq [c(u^n) + 1] \log\{2r[c(u^n) + 1]\} \quad (53)$$

which can easily be shown to be further upper bounded by

$$LZ(u^n) \leq c(u^n) \log c(u^n) + n \cdot \epsilon_1(n), \quad (54)$$

where  $\epsilon_1(n)$  tends to zero uniformly as  $n \rightarrow \infty$ . In words, the LZ78 code length for  $v^n$  is upper bounded by an expression whose leading term is  $c(u^n) \log c(u^n)$ . We shall refer to the quantity  $c(u^n) \log c(u^n)$  as the unnormalized LZ complexity of  $u^n$ , to distinguish from the normalized LZ complexity defined as  $\frac{c(u^n) \log c(u^n)}{n}$ , which means the per-symbol LZ complexity.

## 6.2 SLB for Individual Sequences

The following lemma provides a variation of the generalized Kraft inequality of [12].

*Lemma 5.* For every IL FS encoder with  $s$  states, every  $\alpha > 1$ , every  $\beta \geq 0$ , and every positive integer  $\ell$  which is an integer multiple of  $m$ ,

$$\sum_{\sigma \in \Sigma} \sum_{w^\ell \in \mathcal{U}^\ell} \exp_2 \left\{ -\alpha L[f(\sigma, q(w^\ell))] - \beta \rho(w^\ell - q(w^\ell)) \right\} \leq \frac{s^2 \left[ \sum_{z \in \mathcal{U}} 2^{-\beta \rho(z)} \right]^\ell}{2^{\alpha-1} - 1}, \quad (55)$$

where  $q(w^\ell) \equiv q(w_1^m, w_{m+1}^{2m}, \dots, w_{\ell-m+1}^\ell) \triangleq (q(w_1^m), q(w_{m+1}^{2m}), \dots, q(w_{\ell-m+1}^\ell))$ , the latter being defined as in (50) for vectors in  $\mathcal{U}^\ell$ .

*Proof.* From the postulated IL property, it follows that given  $\sigma \in \Sigma$ , there cannot be more than  $s2^k$  distinct vectors,  $\{v^\ell\}$ , such that  $L[f(\sigma, v^\ell)] = k$ , for every positive integer  $k$ . Therefore,

$$\begin{aligned}
& \sum_{\sigma \in \Sigma} \sum_{w^\ell \in \mathcal{U}^\ell} \exp_2 \left\{ -\alpha L[f(\sigma, q(w^\ell))] - \beta \rho(w^\ell - q(w^\ell)) \right\} \\
&= \sum_{\sigma \in \Sigma} \sum_{k \geq 1} \sum_{\{v^\ell: L[f(\sigma, v^\ell)] = k\}} \sum_{\{w^\ell: q(w^\ell) = v^\ell\}} \exp_2 \left\{ -\alpha k - \beta \rho(w^\ell - q(w^\ell)) \right\} \\
&\leq \sum_{\sigma \in \Sigma} \sum_{k \geq 1} s \cdot 2^k \cdot 2^{-\alpha k} \sum_{z^\ell \in \mathcal{U}^\ell} 2^{-\beta \rho(z^\ell)} \\
&= s^2 \cdot \left[ \sum_{u \in \mathcal{U}} 2^{-\beta \rho(z)} \right]^\ell \cdot \sum_{k=1}^{\infty} 2^{-(\alpha-1)k} \\
&= \frac{s^2 [\sum_{z \in \mathcal{U}} 2^{-\beta \rho(z)}]^\ell}{2^{\alpha-1} - 1}. \tag{56}
\end{aligned}$$

This completes the proof of Lemma 5.

Now, let  $\ell$  divide  $n$ . For a given FS IL encoder  $E$ , a given  $u^n \in \mathcal{U}^n$ , and the associated state sequence  $\sigma^n \in \Sigma^n$  (generated from  $u^n$  by  $E$  using the next-state function  $g$  recursively), consider the joint empirical distribution

$$\hat{P}(\sigma, w^\ell) = \frac{\ell}{n} \sum_{i=0}^{n/\ell-1} \mathcal{I}\{\sigma_{i\ell+1} = \sigma, u_{i\ell+1}^{i\ell+\ell} = w^\ell\}, \quad \sigma \in \Sigma, w^\ell \in \mathcal{U}^\ell \tag{57}$$

Then, according to Lemma 5,

$$\begin{aligned}
& \frac{s^2 [\sum_{z \in \mathcal{U}} 2^{-\beta \rho(z)}]^\ell}{2^{\alpha-1} - 1} \\
&\geq \sum_{\sigma \in \Sigma} \sum_{w^\ell \in \mathcal{U}^\ell} \exp_2 \left\{ -\alpha L[f(\sigma, q(w^\ell))] - \beta \rho(w^\ell - q(w^\ell)) \right\} \\
&= \sum_{\sigma \in \Sigma} \sum_{w^\ell \in \mathcal{U}^\ell} \hat{P}(\sigma, w^\ell) \exp_2 \left\{ -\alpha L[f(\sigma, q(w^\ell))] - \beta \rho(w^\ell - q(w^\ell)) - \log \hat{P}(\sigma, w^\ell) \right\} \\
&\geq \exp_2 \left\{ -\alpha \sum_{\sigma \in \Sigma} \sum_{w^\ell \in \mathcal{U}^\ell} \hat{P}(\sigma, w^\ell) L[f(\sigma, q(w^\ell))] - \beta \sum_{w^\ell \in \mathcal{U}^\ell} \hat{P}(w^\ell) \rho(w^\ell - q(w^\ell)) + \hat{H}(S, U^\ell) \right\}
\end{aligned}$$

$$\begin{aligned}
&= \exp_2 \left\{ -\alpha \cdot \frac{\ell}{n} \sum_{i=0}^{n/\ell-1} L[f(\sigma_{i\ell+1}, v_{i\ell+1}^{i\ell+\ell})] - \beta \cdot \frac{\ell}{n} \sum_{i=0}^{n/\ell-1} \rho(u_{i\ell+1}^{i\ell+\ell} - q(u_{i\ell+1}^{i\ell+\ell})) + \hat{H}(S, U^\ell) \right\} \\
&= \exp_2 \left\{ -\alpha \cdot \frac{\ell}{n} \sum_{t=1}^n L[f(\sigma_t, v_t)] - \beta \cdot \frac{\ell}{n} \sum_{t=1}^n \rho(u_t - v_t) + \hat{H}(S, U^\ell) \right\}, \tag{58}
\end{aligned}$$

where  $\hat{P}(w^\ell) = \sum_{\sigma \in \Sigma} \hat{P}(\sigma, w^\ell)$  and  $\hat{H}(S, U^\ell)$  is the joint entropy of the auxiliary random variables  $S \in \Sigma$  and  $U^\ell \in \mathcal{U}^\ell$ , induced by the empirical joint distribution  $\hat{P}$ . It follows then that

$$\begin{aligned}
&\alpha \cdot \frac{1}{n} \sum_{t=1}^n L[f(\sigma_t, v_t)] + \beta \cdot \frac{1}{n} \sum_{t=1}^n \rho(u_t - v_t) \\
&\geq \frac{\hat{H}(S, U^\ell)}{\ell} - \frac{\log(s^2)}{\ell} - \log \left[ \sum_{z \in \mathcal{U}} 2^{-\beta \rho(z)} \right] + \frac{\log(2^{\alpha-1} - 1)}{\ell} \\
&\geq \frac{\hat{H}(U^\ell)}{\ell} - \log \left[ \sum_{z \in \mathcal{U}} 2^{-\beta \rho(z)} \right] - \frac{2 \log s}{\ell} + \frac{\log(2^{\alpha-1} - 1)}{\ell} \\
&\geq \frac{c(u^n) \log c(u^n)}{n} - \log \left[ \sum_{z \in \mathcal{U}} 2^{-\beta \rho(z)} \right] - \Delta_n(\ell) - \frac{2 \log s}{\ell} + \frac{\log(2^{\alpha-1} - 1)}{\ell}, \tag{59}
\end{aligned}$$

where  $\lim_{n \rightarrow \infty} \Delta_n(\ell) = \frac{1}{\ell}$  and the last step can be found in eq. (12) of [21] as well as references therein. By taking the limit of  $\ell \rightarrow \infty$  (followed by the limit  $n \rightarrow \infty$ ), the last three terms can be made arbitrarily small, whereas the first two terms serve as the individual-sequence counterpart of the right-hand side of eq. (20). Moving the distortion term to the right-hand side and optimizing over  $\beta$ , we obtain

$$\begin{aligned}
\alpha \cdot \frac{1}{n} \sum_{t=1}^n L[f(\sigma_t, v_t)] &\geq \frac{c(u^n) \log c(u^n)}{n} - \inf_{\beta \geq 0} \left\{ \beta \cdot \frac{1}{n} \sum_{t=1}^n \rho(u_t - v_t) + \log \left[ \sum_{z \in \mathcal{U}} 2^{-\beta \rho(z)} \right] \right\} - \\
&\quad \Delta_n(\ell) - \frac{2 \log s}{\ell} + \frac{\log(2^{\alpha-1} - 1)}{\ell} \\
&= \frac{c(u^n) \log c(u^n)}{n} - \Phi \left( \frac{1}{n} \sum_{t=1}^n \rho(u_t - v_t) \right) - \\
&\quad \Delta_n(\ell) - \frac{2 \log s}{\ell} + \frac{\log(2^{\alpha-1} - 1)}{\ell}. \tag{60}
\end{aligned}$$

Let  $L_{\max} = \max_{\sigma, v} L[f(\sigma, v)]$ , Selecting  $\alpha = 1 + \zeta_\ell$ , where  $\zeta_\ell$  tends to zero at sub-exponential rate (say,  $\zeta_\ell = 1/\ell$ ), we have on the one hand, the previous inequality, and on the other hand,

$$\alpha \cdot \frac{1}{n} \sum_{t=1}^n L[f(\sigma_t, v_t)] \leq \frac{1}{n} \sum_{t=1}^n L[f(\sigma_t, v_t)] + \zeta_\ell \cdot L_{\max}, \tag{61}$$

and so, our main result in this section is the following:

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^n L[f(\sigma_t, v_t)] &\geq \frac{c(u^n) \log c(u^n)}{n} - \Phi \left( \frac{1}{n} \sum_{t=1}^n \rho(u_t - v_t) \right) - \\
&\quad \Delta_n(\ell) - \frac{2 \log s}{\ell} + \frac{\log(2^{\zeta_\ell} - 1)}{\ell} - \zeta_\ell \cdot L_{\max} \\
&\geq \frac{c(u^n) \log c(u^n)}{n} - \Phi \left( \frac{1}{n} \sum_{t=1}^n \rho(u_t - v_t) \right) - \\
&\quad \Delta_n(\ell) - \frac{2 \log s}{\ell} + \frac{\log(\zeta_\ell \ln 2)}{\ell} - \zeta_\ell \cdot L_{\max}, \tag{62}
\end{aligned}$$

with the last four terms tending to zero in the limit of  $\ell \rightarrow \infty$  followed by the limit  $n \rightarrow \infty$ .

To conclude, the individual-sequence counterpart of the SLB for finite-alphabet source sequences is essentially of the same form as the classical SLB of the probabilistic setting, except that the normalized entropy term is replaced by the normalized LZ complexity of the sequence and the function  $\Phi(\cdot)$  is calculated at the point of the actual distortion,  $\frac{1}{n} \sum_{t=1}^n \rho(u_t - v_t)$ .

## Appendix

*Proof of the second equality in eq. (4).* Let us define the density function,

$$g(z) = \frac{2^{-\beta \rho(z)}}{\int_{\mathcal{U}} 2^{-\beta \rho(z')} dz'}. \tag{A.1}$$

Then,

$$\begin{aligned}
\inf_{\beta \geq 0} \left\{ \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right] + \beta D \right\} &= \inf_{\beta \geq 0} \left\{ \sup_f [-D(f \| g)] + \log \left[ \int_{\mathcal{U}} 2^{-\beta \rho(z)} dz \right] + \beta D \right\} \\
&= \inf_{\beta \geq 0} \left\{ \sup_f \left[ \int_{\mathcal{U}} dz f(z) \log \left( \frac{2^{-\beta \rho(z)}}{f(z)} \right) \right] + \beta D \right\} \\
&= \inf_{\beta \geq 0} \sup_f \left\{ - \int_{\mathcal{U}} dz f(z) \log f(z) + \beta \left[ D - \int_{\mathcal{U}} f(z) \rho(z) dz \right] \right\} \\
&\stackrel{(a)}{=} \sup_f \inf_{\beta \geq 0} \left\{ - \int_{\mathcal{U}} dz f(z) \log f(z) + \beta \left[ D - \int_{\mathcal{U}} f(z) \rho(z) dz \right] \right\} \\
&= \sup_f \inf_{\beta \geq 0} \{ h(Z) + \beta [D - \mathbf{E}\{\rho(Z)\}] \} \\
&= \sup_f \begin{cases} -\infty & \mathbf{E}\{\rho(Z)\} > D \\ h(Z) & \mathbf{E}\{\rho(Z)\} \leq D \end{cases} \\
&= \sup_{\{Z: \mathbf{E}\{\rho(Z)\} \leq D\}} h(Z), \tag{A.2}
\end{aligned}$$

where (a) is due to the fact that the objective function,  $-\int_{\mathcal{U}} dz f(z) \log f(z) + \beta [D - \int_{\mathcal{U}} f(z) \rho(z) dz]$ , is concave in  $f$  and affine (and hence convex) in  $\beta$ .

## References

- [1] Berger, T. *Rate Distortion Theory - a Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, U.S.A., 1971.
- [2] Gray, R. M. *Source Coding Theory*, Kluwer Academic Publishers, Norwell, MA, U.S.A., 1990.
- [3] Cover, T. M.; Thomas, J. A. *Elements of Information Theory*, John Wiley & Sons, Hoboken, NJ, U.S.A., 2006.
- [4] Linder, T.; Zamir, R. “On the asymptotic tightness of the Shannon lower bound,” *IEEE Trans. Inform. Theory* **1994**, vol. 40, no. 6, pp. 2026–2031.
- [5] Koch, T. “The Shannon lower bound is asymptotically tight,” *IEEE Trans. Inform. Theory* **2016**, vol. 62, no. 11, pp. 6155–6161.
- [6] Kostina, V. “Data compression with low distortion and finite blocklength,” *Fifty-third Annual Allerton Conference* **2015**, Allerton House, UIUC, Illinois, U.S.A., pp. 1127–1134.
- [7] Kostina, V. “When is Shannon’s lower bound tight at finite blocklength?” *Fifty-fourth Annual Allerton Conference* **2016**, Allerton House, UIUC, Illinois, U.S.A., pp. 982–989.
- [8] Gastpar, M.; Sula, E. “Shannon bounds for quadratic rate-distortion problems,” *IEEE Journal on Selected Areas in Inform. Theory* **2024**, vol. 5, pp. 597–608.
- [9] Campbell, L. L. “Kraft inequality for decoding with respect to a fidelity criterion,” *IEEE Trans. Inform. Theory* **1973**, vol. IT-19, no. 1, pp. 68–73.
- [10] Merhav, N. “A comment on “A rate of convergence result for a universal  $D$ -semifaithful code”,” *IEEE Trans. Inform. Theory* **1995**, vol. 41, no. 4, pp. 1200–1202.
- [11] Rissanen, J. “Tight lower bounds for optimum code length,” *IEEE Trans. Inform. Theory* **1982**, vol. IT-28, no. 2, pp. 348–349.



- [12] Ziv, J.; Lempel, A. “Compression of individual sequences via variable-rate coding,” *IEEE Trans. Inform. Theory* **1978**, vol. IT-24, no. 5, pp. 530–536.
- [13] de Bruijn, N. G. *Asymptotic Methods in Analysis*, Dover Publications, New York, 2nd Edition, 1981.
- [14] Merhav, N.; Weinberger, N. “A toolbox for refined information-theoretic analyses with applications,” *Foundations and Trends in Communications and Information Theory* **2025**, vol. 22, no. 1, pp. 1–184.
- [15] Neuschel, T. “Apéry polynomials and the multivariate saddle point method,” *Contr. Approx.* **2014**, vol. 40, pp. 487–507.
- [16] Collatz, L. “Einschließungssatz für die charakteristischen Zahlen von Matrizen”, *Mathematische Zeitschrift* **1942**, vol. 48, no. 1, pp. 221–226.
- [17] Wielandt, H. “Unzerlegbare, nicht negative Matrizen,” *Mathematische Zeitschrift* **1950**, vol. 52, no. 1, 642–648.
- [18] M. D. Donsker and S. R. S. Varadhan, “Asymptotic evaluation of certain Markov process expectations for large time, IV” *Communications on Pure and Applied Mathematics*, vol. 36, no. 2, pp. 183–212, 1983.
- [19] Merhav, N.; Shamai (Shitz), S. “Volume-based lower bounds to the capacity of the Gaussian channel under pointwise additive input constraints,” submitted for publication **2025**, and available on-line at: <https://arxiv.org/pdf/2510.04095>
- [20] Horn, R. A.; Johnson, C. R. *Matrix Analysis*, Cambridge University Press, New York, NY, U.S.A., 1985.
- [21] N. Merhav, “Universal Slepian-Wolf coding for individual sequences,” *IEEE Trans. Inform. Theory*, vol. 71, no. 1, pp. 783–796, January 2025.