A Lower Bound on the Redundancy of Arithmetic-type Delay Constrained Coding

Eado Meron*, Ofer Shayevitz†, Meir Feder and Ram Zamir
Dept. of Electrical Engineering Systems
Tel Aviv University, Tel Aviv 69978, Israel
Email: {eado, ofersha, meir, zamir}@eng.tau.ac.il

Abstract

In a previous paper we derived an upper bound on the redundancy of an arithmetic-type encoder for a memoryless source, designed to meet a finite end-to-end strict delay constraint. It was shown that the redundancy decays exponentially with the delay constraint and that the redundancy-delay exponent is lower bounded by $\log(1/\alpha)$ where α is the probability of the most likely source symbol. In this work, we prove a corresponding upper bound for the redundancy-delay exponent, $C \cdot \log 1/\beta$ where β is the probability of the least likely source symbol. This bound is valid for almost all memoryless sources and for all arithmetic-type (possibly time-varying, memory dependent) lossless delay-constrained encoders. We also shed some light on the difference between our exponential bounds and the polynomial $O(d^{-5/3})$ upper bound on the redundancy with an average delay constraint d, derived in an elegant paper by Bugeaud, Drmota and Szpankowski for another class of variable-to-variable encoders, and show that the difference is due to the precision needed to memorize the encoder's state.

I Introduction

Historically, lossless source coding is divided into three classes: Block-to-Variable (BV), e.g. the Huffman code, Variable-to-Block (VB), e.g. Tunstall code, and the most general class of Variable-to-Variable (VV) codes, which include the former as special cases. It is well known that an optimal coding scheme of any class can attain a coding rate approaching the entropy of the source [1][2]. However, in the presence of resource constraints, a rate penalty, referred to as the *redundancy*, is unavoidable. In this work we focus on the redundancy in the encoding of a memoryless source that

*E. Meron is supported by The Yitzhak and Chaya Weinstein Research Institute for Signal Processing.

†O. Shayevitz is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

stems from a *strict end-to-end delay constraint* d, i.e., the n-th encoded symbol is always reproduced by time $n+d$. In BV and VB coding, this redundancy is known to be $O(d^{-1})$ [2][3]. In a previous work [4], we demonstrated (using an arithmetic-type encoder with memory and a fictitious symbol) that VV coding is significantly superior to BV and VB, as it can attain a redundancy of $O(\alpha^d)$, where α is the probability of the most likely source symbol. This should be contrasted with a result of Khodak [5], recently simplified and generalized [6], which proves that for a VV code that is a VB–BV concatenation, the redundancy is $O(d^{-\frac{5}{3}})$ for the even weaker *average* delay constraint d. This apparent contradiction is due to the fact that not all VV encoders can be represented in this way. An important example of that sort is an *ideal arithmetic encoder* [7], which admits a bounded expected delay yet has *zero* asymptotic redundancy [8].

In this paper, we complement the analysis of the tradeoff between delay and redundancy in arithmetic-type VV coding, i.e., considering encoders described by general nested interval mappings, including time varying and memory dependent mappings. We provide a lower bound of $\Omega(\beta^{8d})$ on the attainable redundancy in such an encoding of a memoryless source with a strict delay constraint d, where β is the smallest symbol probability. A tighter upper bound on the exponent, similar to a lower bound mentioned in [4] and replacing $\log 1/\beta$ with a function of the source's entropy, can be derived but due to space limitations will be given elsewhere.

II Definitions

Throughout the paper we consider a *discrete memoryless source* Q over a finite alphabet $\mathcal{X} = \{0, 1, \ldots, K-1\}$ with positive symbol probabilities $\{p_0, p_1, \ldots, p_{K-1}\}$. A finite source sequence is denoted by $x_m^n = \{x_m, x_{m+1}, \ldots, x_n\}$ with $x^n = x_1^n$, while an infinite one is denoted by x^∞. A random source sequence is denoted similarly using capital letters. The probability that the source emits a super-symbol $x^d \in \mathcal{X}^d$ is denoted $Q^d(x^d)$. The *entropy* of the source is denoted $H(Q)$. The *kullback-Leibler distance*, or *divergence*, between two sources P, Q over the same alphabet is denoted $D(P\|Q)$. We denote by $|A|$ the measure of a set $A \subseteq \mathbb{R}$. The *fractional* part of a number a is denoted by $\langle a \rangle$. The *difference modulo-1* $\langle A - B \rangle$ between two sets $A, B \subseteq \mathbb{R}$ is the set of all numbers $\langle a - b \rangle$ where $a \in A\,, b \in B$. All logarithms are taken to the base of 2.

Definition 1. A binary sequence $b^k = \{b_1, b_2, \ldots, b_k\}$ with $b_j \in \{0, 1\}$ is said to *represent* a *binary interval* $\big[0.b_1b_2,\ldots b_k0,\ 0.b_1b_2,\ldots b_k1\big) \subseteq [0,1)$.

Definition 2. An *interval-mapping encoder* (or an *arithmetic-type* encoder) $\mathcal{E}$, is a sequence of mappings $\mathcal{I}_n^{\mathcal{E}} : \mathcal{X}^n \mapsto \mathfrak{I}$, where $\mathfrak{I} = \{[\,a, b) \,|\, 0 \le a < b \le 1\}$, i.e., mappings of finite source sequences into intervals. The encoder must satisfy the following *disjoint nesting* property: $\mathcal{I}_{n+1}^{\mathcal{E}}(x^n y) \subset \mathcal{I}_n^{\mathcal{E}}(x^n)$ and $\mathcal{I}_n^{\mathcal{E}}(x^{n-1}y) \cap \mathcal{I}_n^{\mathcal{E}}(x^{n-1}z) = \phi$ for all $n \in \mathbb{N}$, $x^n \in \mathcal{X}^n\,, y, z \in \mathcal{X}\,, y \neq z$. After having observed x^n, we say that the interval-

mapping encoder has *emitted* the bit sequence $\mathcal{E}(x^n)$ representing the minimal binary interval containing $\mathcal{I}_n^{\mathcal{E}}(x^n)$.

In the sequel, any reference to an *encoder* should be understood as referring to an interval mapping encoder as defined above.

Definition 3. A (lossless) encoder is said to meet a strict *delay constraint* d, and is called *d-constrained*, if x^n is uniquely determined by $\mathcal{E}(x^n y^d)$ for all $n \in \mathbb{N}, x^n \in \mathcal{X}^n, y^d \in \mathcal{X}^d$. The family of d-constrained encoders is denoted $\mathfrak{I}_d$.

Remark. Using a d-constrained encoder means that at every time instant n, the decoder can uniquely determine the entire encoded sequence up to time $n-d$. Mind the difference between delay as defined above, and the standard block/parse length for BV,VB,VV codes. For instance, consider an encoder repetitively using of a BV code with block length K. Clearly, this is a K-constrained encoder. However, if the code is designed in a proper fashion, the encoder will be able to emit bits before the end of each block[1], and the decoder will be able to decode symbols before each codeword ends. Therefore, this is a d-constrained encoder with $d < K$.

III The Forbidden Points Concept

Let $\mathcal{I}_n^{\mathcal{E}}(x^n)$ be the current encoder's interval. In previous work [4][8] we have shown that in order for x^n to be decoded at a time $n+d$, the interval $\mathcal{I}_{n+d}^{\mathcal{E}}(x^{n+d})$ must not contain a certain countable set of points within $\mathcal{I}_n^{\mathcal{E}}(x^n)$, which are termed *forbidden points*. In particular, this means that the intervals of a d-constrained encoder do not cover the entire unit interval, i.e., $\bigcup_{x^n} \mathcal{I}_n^{\mathcal{E}}(x^n) \neq [0,1)$. The "origin" point of the forbidden points set is the midpoint of the minimal binary interval containing $\mathcal{I}_n^{\mathcal{E}}(x^n)$. It is easily verified that if $\mathcal{I}_{n+d}^{\mathcal{E}}(x^{n+d})$ contains this point, then the encoder has not emitted any bits since encoding x^n, hence if $\mathcal{I}_n^{\mathcal{E}}(x^n)$ is not precisely a binary interval then x^n cannot be decoded at time $n+d$. This observation can be generalized to produce the rest of the forbidden points, which are all the points reached by starting from the origin point and moving towards the edges of $\mathcal{I}_n^{\mathcal{E}}(x^n)$ with "maximal binary jumps", i.e., each jump is the maximal possible jump of the form $2^{-\ell}$ for some $\ell \in \mathbb{N}$, still ensuring we remain within $\mathcal{I}_n^{\mathcal{E}}(x^n)$. For a much required tutorial, see [4][8].

IV Main Result

Definition 4. For a memoryless source Q, the *redundancy* of the encoder $\mathcal{E}$ at time n, its *sup-redundancy* and its *inf-redundancy* are defined as[2]

$$R_n^{\mathcal{E}} \triangleq \frac{1}{n} D(Q^n \,\|\, |\mathcal{I}_n^{\mathcal{E}}|)\,, \qquad \overline{R}^{\mathcal{E}} \triangleq \limsup_{n\to\infty} R_n^{\mathcal{E}}\,, \qquad \underline{R}^{\mathcal{E}} \triangleq \liminf_{n\to\infty} R_n^{\mathcal{E}}$$

[1] These bits are the common prefix of all the codewords, that correspond to source sequences that begin with the source symbols seen thus far.

[2] Q^n and $|\mathcal{I}_n^{\mathcal{E}}|$ may be though of as probability distributions over $\mathcal{X}^n \cup \{\omega\}$, where ω is a fictitious symbol pertaining to regions of [0,1) uncovered by $\mathcal{E}$, with $Q^n(\omega) = 0$, $|\mathcal{I}_n^{\mathcal{E}}(\omega)| = 1 - \sum_{x^n} |\mathcal{I}_n^{\mathcal{E}}(x^n)|$.

The corresponding *sup-redundancy-delay function* and *inf-redundancy-delay function* and their exponents are defined as (recall that $\Im_d$ is the family of d-constrained interval-mapping encoders):

$$\overline{R}(d) = \inf_{\mathcal{E}\in\Im_d} \overline{R}^{\mathcal{E}}, \qquad \overline{E}_{rd} = \lim_{d\to\infty} -\frac{1}{d}\log \overline{R}(d)$$

$$\underline{R}(d) = \inf_{\mathcal{E}\in\Im_d} \underline{R}^{\mathcal{E}}, \qquad \underline{E}_{rd} = \lim_{d\to\infty} -\frac{1}{d}\log \underline{R}(d)$$

Theorem 1 (From [4]). *For any memoryless source Q, the sup-redundancy-delay exponent is lower bounded by*

$$\overline{E}_{rd} \geq \log(1/\alpha)$$

where $\alpha = \max\{p_0, \ldots, p_{K-1}\}$ is the maximal symbol probability.

We now state our main result, providing an upper bound for the inf-redundancy exponent for *almost any* source, which is meant w.r.t. (say) a uniform distribution over the probability simplex. Note that such a statement cannot hold for all sources, e.g. for 2-adic sources we can have zero redundancy with zero delay.

Theorem 2. *For almost any memoryless source Q, the inf-redundancy-delay exponent is upper bounded by*

$$\underline{E}_{rd} \leq 8\log(1/\beta)$$

where $\beta = \min\{p_0, \ldots, p_{K-1}\}$ is the minimal symbol probability.

Since the proof of Theorem 2 is both complex and tedious, we provide a brief outline of the main ideas behind the proof, which is given in section V. We first note that due to the strict delay constraint, at any time point the encoder must map the next d symbols into intervals that do not contain any forbidden points. Typically (for almost every interval), we will find an infinite number of forbidden points concentrated near the edges, with a typical "concentration region" whose size depends on the specific interval. Clearly, the distances between consecutive points diminishes exponentially to zero. Therefore, mapping symbols to the concentration region will result in a significant mismatch between the symbol probability and the interval length, and this phenomena incurs redundancy. This observation is made precise in Lemma 2.

Now, roughly speaking, there are two opposing strategies the encoder may use when mapping symbols to intervals. The first is to think short-range, namely to be as faithful to the source as possible by assigning interval lengths closely matching symbol probabilities (within the forbidden points constraint). This will likely cause the next source interval to have a relatively large concentration region, resulting in an inevitable redundancy at the subsequent mapping. The second strategy is to think long-range, by mapping to intervals with a small concentration region. This in general cannot be done while still being faithful to the source's distribution, hence this strategy also incurs in an inevitable redundancy. The latter observation is made precise in Lemma 4. Our lower bound resides in the balancing point of these two counterbalancing sources of redundancy.

V Proof of Theorem 2

Definition 5. Let $\mathcal{I}(x^n) = \mathcal{I}_n^{\mathcal{E}}(x^n)$ be the current encoder's interval. We define the d-instantaneous redundancy to be

$$r_d(x^n) = \sum_{x_{n+1}^{n+d} \in \mathcal{X}^d} Q(x_{n+1}^{n+d}) \log \frac{Q(x_{n+1}^{n+d})}{|\mathcal{I}(x^{n+d})| \,/ |\mathcal{I}(x^n)|}$$

Namely, the difference between the optimal and assigned codelength for the next d symbols.

Lemma 1. *The sup-redundancy and inf-redundancy are also given by*

$$\overline{R}^{\mathcal{E}} = \limsup_{n\to\infty} \frac{1}{nd} \sum_{k=1}^{n} \mathbb{E}(r_d(X^k)) \,, \qquad \underline{R}^{\mathcal{E}} = \liminf_{n\to\infty} \frac{1}{nd} \sum_{k=1}^{n} \mathbb{E}(r_d(X^k))$$

Proof. We prove the Lemma only for the inf-redundancy, the other part is the same. First, note that for any fixed $d \in \mathbb{N}$ the inf-redundancy can be written as

$$\underline{R}^{\mathcal{E}} = \liminf_{n\to\infty} \frac{1}{n} \mathbb{E}(-\log |I(X^n)|) - H(Q) = \liminf_{n\to\infty} \frac{1}{n} \mathbb{E}(-\log |I(X^{n+d})|) - H(Q)$$

Now let us expand the expected d-instantaneous redundancy as follows:

$$\mathbb{E}(r_d(X^k)) = \mathbb{E}\left(\mathbb{E}\left(\log \frac{|\mathcal{I}(X^k)|}{|\mathcal{I}(X^{k+d})|} \;\middle|\; X^k \right) - d \cdot H(Q) \right) = \mathbb{E} \log \frac{|\mathcal{I}(X^k)|}{|\mathcal{I}(X^{k+d})|} - d \cdot H(Q)$$

Therefore,

$$\liminf_{n\to\infty} \frac{1}{nd} \sum_{k=1}^{n} \mathbb{E}(r_d(X^k)) = \liminf_{n\to\infty} \frac{1}{n} \left(\frac{1}{d} \sum_{k=1}^{d} \mathbb{E} \log |\mathcal{I}(X^k)| + \frac{1}{d} \sum_{k=1}^{d} \mathbb{E}\left(-\log |\mathcal{I}(X^{n+k})| \right) \right)$$
$$- H(Q) \;\leq\; \liminf_{n\to\infty} \frac{1}{n} \mathbb{E}\left(-\log |\mathcal{I}(X^{n+d})| \right) - H(Q) = \underline{R}^{\mathcal{E}}$$

where in the inequality transition we used the nesting property. The inequality in the other direction is derived similarly, concluding the proof. □

Definition 6. The *normalized distance* between points in a given interval, is their Euclidian distance divided by the length of the interval.

Definition 7. Let $I \subseteq [0,1)$ be any interval. We define $\delta_I = \delta_I(\beta, d)$ as the maximal normalized distance between neighboring forbidden points of I, that is smaller than $\frac{\beta^d}{4}$.

Lemma 2. *Let $\mathcal{I}(x^n)$ be the current encoder's interval. Then $r_d(x^n) > \delta_{\mathcal{I}(x^n)}$*

Proof. Given the encoder's mapping of the next d symbols, let γ be the normalized size of the smallest interval between two consecutive forbidden points, into which some super-symbol x_{n+1}^{n+d} is assigned. If $\gamma < \delta_{\mathcal{I}}$, then some super-symbol has been assigned an interval whose normalized length is at least four times smaller than the super-symbol's probability. The remaining symbols are therefore mapped into an interval of normalized size at most[3] $1-3\gamma$. Thus, the d-instantaneous redundancy can be lower bounded as follows:

$$r_d(x^n) \geq \min_{p=Q^d(x_{n+1}^{n+d})} \left\{ p \log \frac{p}{\gamma} + (1-p) \log \frac{1-p}{1-3\gamma} \right\} \tag{1}$$

Since the derivative of the function inside the minima w.r.t p is positive for $p > \frac{\gamma}{1-2\gamma}$, and since $p \geq \beta^d > 4\gamma$, then the minima in (1) is attained for[4] $p = \beta^d$, hence

$$\begin{aligned} r_d(x^n) &\geq \beta^d \log \frac{\beta^d}{\gamma} + (1-\beta^d) \log \frac{1-\beta^d}{1-3\gamma} \geq 2\beta^d + \log(1-\beta^d) \\ &\geq \beta^d \left(2 - \frac{\log e}{1-\beta^d} \right) \geq \delta_{\mathcal{I}(x^n)} \end{aligned}$$

where we have used $\log(1-p) \geq -\frac{p}{1-p}$ for $0 < p < 1$, and where the last transition is true[5] for $d > 2$. If on the other hand $\gamma \geq \delta_{\mathcal{I}}$, then all of the d-fold alphabet has been assigned to an interval of size at most $1-2\gamma$ which results in a d-instantaneous redundancy lower bounded by

$$r_d(x^n) \geq \log \frac{1}{1-2\gamma} \geq \log \frac{1}{1-2\delta_{\mathcal{I}}} \geq 2\delta_{\mathcal{I}(x^n)}$$

□

Definition 8. A number $a \in [0,1)$ is called (m,ℓ)–*constrained* if

$$a = 0.\underbrace{00\ldots0}_{m'(a)}\ \underbrace{1\phi\ldots\phi}_{m}\ \underbrace{00\ldots0}_{\ell}\ \phi\ldots$$

where $m'(a)$ is the length of the zeros prefix of a, and ϕ is the "don't care" symbol. The (m,ℓ)–*constrained region* $\mathcal{C}_{m,\ell}$ is the set of all such numbers.

Definition 9. A number $a \in [0,1)$ is called (m,ℓ)–*violating* if

$$a = 0.\underbrace{00\ldots0}_{m'(a)}\ \underbrace{1\phi\ldots\phi}_{m}\ \underbrace{\phi\ldots\ldots\ldots\ldots\phi}_{\ell \text{ not all '0' or all '1'}}\phi\ldots \tag{2}$$

The (m,ℓ)–*violating region* $\mathcal{V}_{m,\ell}$ is the set of all such numbers. The complement $\overline{\mathcal{V}}_{m,\ell} = [0,1) \setminus \mathcal{V}_{m,\ell}$ is called the (m,ℓ)–*permissible region.*

[3] Recall the properties of the forbidden points.

[4] Recall that β is the minimal symbol probability.

[5] d should be such that $\beta^d < 1 - \frac{4\log e}{7} \approx 0.1756$, and recall that $\beta \leq \frac{1}{2}$.

Corollary 1. *For $\mu > 0$, if $a \in \mathcal{V}_{(m,\lceil \mu m \rceil - 1)}$ is approximated by $b \in \mathcal{C}_{(m,\lceil \mu m \rceil)}$, then*

$$|a - b| \geq 2^{-m'(a)} \cdot 2^{-\lceil m(1+\mu) \rceil} \geq \frac{a}{2} \cdot 2^{-\lceil m(1+\mu) \rceil}$$

Definition 10. Let[6] $L\mathcal{C}_{m,\ell} = \langle -\log \mathcal{C}_{m,\ell} \rangle$ and $L\overline{\mathcal{V}}_{m,\ell} = \langle -\log \overline{\mathcal{V}}_{m,\ell} \rangle$, and define

$$\mathcal{D}^{(1)}_{m,\ell} \stackrel{\triangle}{=} \langle L\overline{\mathcal{V}}_{m,\ell} - L\mathcal{C}_{m,\ell} \rangle, \quad \mathcal{D}^{(2)}_{m,\ell} \stackrel{\triangle}{=} \langle \mathcal{D}^{(1)}_{m,\ell} - \mathcal{D}^{(1)}_{m,\ell} \rangle$$

Corollary 2. *If $a \notin \mathcal{D}^{(1)}_{m,\lceil \mu m \rceil}$ and $b \in L\mathcal{C}_{m,\lceil \mu m \rceil}$ then $a + b \notin L\overline{\mathcal{V}}_{m,\lceil \mu m \rceil}$.*

Corollary 3. *If $I, J \subseteq [0,1)$ are each a finite union of M intervals each of size no larger than r, then $|\langle I - J \rangle| \leq 2M^2 r$.*

The (m, ℓ)–permissible region within the interval $[1/2, 1)$ is comprised of $2^{m-1} + 1$ subintervals. By definition, the size of each is upper-bounded by $2^{-(m'+m+l)+1}$. Applying $\langle -\log(\cdot) \rangle$ to all such intervals in the $[1/2, 1)$ interval (corresponding to $m' = 0$) will stretch each of them by a factor of at most $\frac{2}{\log 2} < 3$. All other permissible intervals (those with $m' > 0$) coincide on the unit interval after applying the $\langle -\log(\cdot) \rangle$ operator. Thus, the measures of $L\overline{\mathcal{V}}_{m,\ell}$ and $L\mathcal{C}_{m,\ell} \subset L\overline{\mathcal{V}}_{m,\ell}$ are bounded by:

$$|L\mathcal{C}_{m,\ell}| < |L\overline{\mathcal{V}}_{m,\ell}| < (2^{m-1} + 1)2^{-(m+l)+1}3 < 4 \cdot 2^{-l}.$$

Using corollary 3, the measures of $\mathcal{D}^{(1)}_{m,\ell}$ and $\mathcal{D}^{(2)}_{m,\ell}$ are bounded by

$$|\mathcal{D}^{(1)}_{m,\ell}| < 2^m \cdot 2^{-(m+l-1)} \cdot 2^m \cdot 2 = 4 \cdot 2^{m-l}, \quad |\mathcal{D}^{(2)}_{m,\ell}| < (2^{2m})^2 \cdot 2^{-(m+l-2)} \cdot 2 = 8 \cdot 2^{3m-l} \quad (3)$$

Definition 11. A source is called *λ–regular* if there exists a pair of source symbols (i, j), such that and any $\mu > 3$ there exists some $m_0 \in \mathbb{N}$ so that

$$\lambda = \left\langle \log \frac{p_i}{p_j} \right\rangle \notin \bigcup_{m=m0}^{\infty} \mathcal{D}^{(2)}_{m,\lceil \mu m \rceil} \quad (4)$$

Note that $0 \in \mathcal{D}^{(2)}_{m,\lceil \mu m \rceil}$ for any m and μ, hence no source can be 0–regular.

Lemma 3. *Almost any source is λ-regular for some $\lambda > 0$.*

Proof. It follows immediately from (3) that for any $\mu > 3$, the set $\bigcup_{m=m_0}^{\infty} \mathcal{D}^{(2)}_{m,\lceil \mu m \rceil}$ has a measure approaching zero with m_0, hence the result. □

Lemma 4. *Suppose Q is a λ–regular source, and for some $\mu > 3$ let*

$$A^{\mu}_{d} = \left\{ x^d \in \mathcal{X}^d \;\middle|\; \langle -\log Q^d(x^d) \rangle \notin \mathcal{D}^{(1)}_{m,\lceil \mu m \rceil}, \, m = \lceil -d \log \beta \rceil \right\}$$

Then for any $\epsilon > 0$ we have $Q^d(A^{\mu}_{d}) > \frac{1}{2} - \epsilon$ for large enough d.

[6] The log and $\langle \cdot \rangle$ operations are taken pointwise on the set elements.

Proof. Let (i,j) be the symbols attaining λ, and suppose $d > \frac{m_0}{\log(1/\beta)}$ so that (4) is satisfied. Now consider some super-symbol $x^d \notin A_d^\mu$. By replacing a single occurrence of the symbol i in x^d by an occurrence of the symbol j, we get a new super-symbol $\widetilde{x}^d$, such that

$$\langle -\log Q^d(\widetilde{x}^d)\rangle = \langle -\log Q^d(x^d) + \lambda\rangle$$

Since $\lambda \notin \mathcal{D}^{(2)}_{m,\lceil\mu m\rceil}$ for any $m \geq m_0$, and since this region is composed of all the possible distances between points in $\mathcal{D}^{(1)}_{m,\lceil\mu m\rceil}$, then we must conclude that $\langle -\log Q^d(\widetilde{x}^d)\rangle \notin \mathcal{D}^{(1)}_{m,\lceil\mu m\rceil}$ for any $m \geq m_0$, hence $\langle -\log Q^d(\widetilde{x}^d)\rangle \in A_d^\mu$.

Therefore, for almost any[7] *type* whose super-symbols are outside A_d^μ, there exists another type with almost the same probability, whose super-symbols are inside A_d^μ. Moreover, this transition between types is invertible, and hence unique. Therefore, by the Asymptotic Equipartition Property [1] the source must have, at worst, a probability mass arbitrarily close to a half inside A_d^μ as d increases. □

Proof of Theorem 2. Set $\mu > 3$, and let $m = \lceil -d\log 1/\beta\rceil$ throughout the proof. We shall lower bound $\mathbb{E}(r_d(X^k) + r_d(X^{k+d}))$. Define the event

$$E = \left(\delta_{\mathcal{I}(x^k)} > \beta^{\mu d} \quad \vee \quad \delta_{\mathcal{I}(x^{k+d})} > \beta^{\mu d}\right)$$

Given E, by Lemma 2 we have $r_d(x^k) + r_d(x^{k+d}) > \beta^{\mu d}$, which means redundancy. Assuming E^c, we have $\mathcal{I}(x^k), \mathcal{I}(x^{k+d}) \in \mathcal{C}_{m,\lceil\mu m\rceil}$. Recalling the set A_d^μ of Lemma 4 and using Pinsker's inequality, we get

$$r_d(x^k) \geq \left(\sum_{x_{k+1}^{k+d}\in A_d^\mu}\left|Q^d(x_{k+1}^{k+d}) - \frac{|\mathcal{I}(x^{k+d})|}{|\mathcal{I}(x^k)|}\right|\right)^2 = \left(\sum_{x_{k+1}^{k+d}\in A_d^\mu}\left|\frac{Q^d(x_{k+1}^{k+d})|\mathcal{I}(x^k)| - \mathcal{I}(x^{k+d})|}{|\mathcal{I}(x^k)|}\right|\right)^2$$

$$\overset{(a)}{\geq}\left(\frac{1}{|\mathcal{I}(x^k)|}\sum_{x_{k+1}^{k+d}\in A_d^\mu}\frac{1}{2}\,Q^d(x_{k+1}^{k+d})|\mathcal{I}(x^k)|\cdot\beta^{(\mu+1)d}\right)^2 = \left(\frac{Q^d(A_d^\mu)}{2}\right)^2\cdot\beta^{2(\mu+1)d} \overset{(b)}{\geq} c\beta^{2(\mu+1)d}$$

In transition (a) we used the fact that both $\mathcal{I}(x^k), \mathcal{I}(x^{k+d}) \in \mathcal{C}_{m,\lceil\mu m\rceil}$ and that we sum over $Q(x_{k+1}^{k+d}) \in A_d^\mu$ together with Corollaries 1 and 2. In transition (b) we used Lemma 4 to lower bound the probability of the set A_d^μ, with $c = \left(\frac{1}{4} - \epsilon\right)^2$. In summary,

$$\begin{aligned}\mathbb{E}(r_d(X^k) + r_d(X^{k+d})) &\geq \mathbb{P}(E)\beta^{\mu d} + (1-\mathbb{P}(E))\cdot c\beta^{2(\mu+1)d}\\ &\geq \inf_{q\in[0,1]}\left(q\beta^{\mu d} + (1-q)\cdot c\beta^{2(\mu+1)d}\right) \geq c\beta^{2(\mu+1)d}\end{aligned}$$

and plugging into Lemma 1 we get

$$\underline{R}^{\mathcal{E}} = \liminf_{n\to\infty}\frac{1}{2nd}\sum_{k=1}^{n}\mathbb{E}(r_d(X^k) + r_d(X^{k+d})) \geq \frac{c}{2d}\,\beta^{2(\mu+1)d}$$

[7] Except for the case where the ith symbol does not occur, but this type has a negligible probability for large enough d.

Since this lower bound holds for any $\mu > 3$ and any d-constrained encoder $\mathcal{E} \in \Im_d$, we immediately have

$$\underline{R}(d) \geq \frac{c}{2d}\,\beta^{2(\mu+1)d}\,, \qquad \underline{E}_{rd} \leq 8\log(1/\beta)$$

□

VI Discussion

Given the results above it is natural to wonder what are the reasons for the difference in performance (polynomial vs. exponential redundancy as a function of the delay) between the VV codes considered in [5][6], that are a concatenation of VB and BV codes, and the class of interval-mapping VV codes considered herein. First, note that it is usually impossible to represent a VB-BV code by an interval-mapping encoder. For instance, a Huffman-like code might map two source sequences differing only in the last symbol into bit sequences that do not have a common prefix, that thus represent far away intervals. Moreover, VB-BV codes are not *designed* to be sequential, in the sense that they do not emit any bits until the block/parse is over.

An interval-mapping encoder that tries to emulate a VB-BV encoder (although this is not always precisely possible), would reach a perfectly binary interval roughly every poly(d) steps. This has two consequences: First, since a certain amount of redundancy is incurred at such points, the overall redundancy diminishes polynomially with the delay constraint. Second, since a binary interval is a re-scaled version of the unit interval (without changing the locations of future forbidden points), the encoder is practically reset at these instances, and the prefix has no future effect on its behavior. This provides the intuition as to why the VB-BV concatenation is a harsh restriction, which is not called for in the d-constrained coding scenario.

However, although the reset points of VB-BV schemes incur redundancy, they allow the encoder to start-over and therefore VB-BV encoders are more effective in a *precision limited* setting. The VV encoders discussed in this paper can hence attain their superior performance by allowing a finer precision for keeping the encoder's state/interval. However, it should be noted that only a finite precision is necessary when maintaining a delay constraint, and that precision can be derived from Lemma 2. Therefore, the redundancy of our interval-mapping encoder when operating in a resource limited setting, is dominated by the larger of two sources: The above delay-precision constraint, and the external complexity-precision constraint.

It is generally possible to have an encoder which is not interval-mapping, but is yet sequential in the sense of emitting all the bits it can at any time-point. Even the VB-BV encoders of [5][6] can be generally modified to emit bits before the end of the block/parse is reached, making the encoder sequential and the delay smaller then the block/parse length, since symbols may be decoded earlier. For example, if we terminate the d-constrained arithmetic-type encoder described in [4] using a large enough block length n, then we get a BV code with delay d, but a block-length $n \gg d$. Nevertheless, we can actually show that the results of this paper are essentially valid

for all lossless d-constrained encoders, therefore indicating that the interval-mapping encoders achieve the best delay-redundancy tradeoff possible.

Finally, we believe that the zero-measure set of sources for which the bound of Theorem 2 may not hold, can be reduced from non-λ-regular sources to the set of 2-adic sources only, which is the smallest possible since for these sources zero delay and zero redundancy are simultaneously attainable. Moreover, we believe that the divergence between a source's distribution and the closest 2-adic distribution determines the minimal delay from which our bound is tight. It also seems that the theory of uniformly-distributed sequences modulo-1 can be a useful tool to that end [9].

References

[1] T.M. Cover and J.A Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[2] M. Drmota, Y. Reznik, S.A. Savari, and W. Szpankowski, "Precise asymptotic analysis of the tunstall code," in *Proc. of the International Symposium on Information Theory*, 2006.

[3] W. Szpankowski, "Asymptotic average redundancy of huffman (and other) block codes," *IEEE Transactions on Information Theory*, vol. 46, no. 7, Nov 2000.

[4] O. Shayevitz, E. Meron, M. Feder, and R. Zamir, "Bounds on redundancy in constrained delay arithmetic coding," in *Proc. of the Data Compression Conference*, 2007, pp. 133–142.

[5] G.L. Khodak, "Delay-redundancy relation of vb-encoding (in russian)," *All-union Conference on Theoretical Cybernetics, Novobirsk*, 1969.

[6] Y. Bugeaud, M. Drmota, and W. Szpankowski, "On the construction of (explicit) Khodak's code and its analysis," *IEEE Transactions on Information Theory, submitted.*

[7] F. Jelinek, *Probabilistic Information Theory*, McGraw-Hill, New York, 1968.

[8] O. Shayevitz, R. Zamir, and M. Feder, "Bounded expected delay in arithmetic coding," in *Proc. of the International Symposium on Information Theory*, 2006.

[9] M. Drmota and R.F. Tichy, *Sequences, Discrepancies and Applications*, Springer-Verlag, 1997.