Universal Fixed-Length Coding Redundancy

Cheng Chang
UC Berkeley EECS
cchang@eecs.berkeley.edu

Anant Sahai
UC Berkeley EECS
sahai@eecs.berkeley.edu

***Abstract*— We consider the problem of choosing a block-length to achieve a desired probability of error for *fixed-length* lossless source coding and channel coding of a *finite* amount of payload data. This is closely related to the issue of redundancy. While Baron, *et al* in [3], [20] studied this problem for rates in the vicinity of entropy for *known* source distributions and in [2] for rates in the vicinity of capacity for *known* channels using central-limit-theorem style asymptotics, we are interested in all rates for *unknown* source distributions and *unknown* channels. By using the universal lower bound on the source coding error exponents in [5] and the similar universal lower-bound to the channel coding error exponent in [13], we derive universal upper bounds to the redundancy rate that are *non-asymptotic* in that they are explicitly valid for all alphabet sizes, all block lengths, and all target error probabilities. Because of the simplicity of the bounds, they also shed light on the large alphabet asymptotics for both source and channel coding.**

I. Introduction

Classical information theory results are often based on the assumption that the code lengths (and thus implicitly the data payloads) are *sufficiently long*. In practice, block length are finite and furthermore, many applications may prefer to generate short rather than long data payloads.[1] Hence it is important to understand the non-asymptotic behavior of coding systems, or equivalently, the gaps between asymptotic results and non-asymptotic reality [3], [2], [20]. In this paper, this is done in the context of fixed-length block coding that is "lossless" in the sense that a specified probability of block-error must be met.

A. Background: variable-length redundancy

The concept of redundancy has been traditionally studied for variable-length codes. For point-to-point lossless source-coding, a code's redundancy rate is defined as

$$\frac{1}{n}E_p(l(x_1^n)) - H(x) \tag{1}$$

where $l(x_1^n)$ is the code-length.

The story is trivial for the non-universal case since the Shannon code has code length $l(x_1^n) = \lceil \log p(x_1^n)^{-1} \rceil$. For i.i.d. random variables from distribution p_x, it is well known [6] that the expected code length is at most $nH(x) + 1$, giving an achievable redundancy of $\frac{1}{n}$.

Variable-length code redundancy becomes more interesting when the coding system does not know the source distribution. Because there is an expectation within (1), the redundancy of a specific code could depend on the source's actual distribution and the goal is to make the redundancy rate small for almost all possible distributions. Rissanen' Minimum Description Length (MDL) theory [19], [1] and Kontoyiannis [14] showed that the minimum redundancy is asymptotically $\frac{1}{2}(|\mathcal{X}| - 1)\log_2(n)$ plus some lower order (in n) terms.

Since the redundancy depends on the alphabet-size, it is natural to consider what happens if this also scales with n. This was done by Shamir for MDL and variable-length coding in [22]. We investigate large alphabet redundancy rates for fixed-length block coding for fixed-block coding in Section III-D.

B. Redundancy and fixed-length coding

First, we examine lossless source coding in two forms: point-to-point as illustrated in Figure 1 and with decoder side-information as illustrated in Figure 2. Sources are memoryless from finite-alphabet $\mathcal{X}$, with $\mathcal{Y}$ being the finite side-information alphabet if it is available. In both cases, a block of n source symbols is mapped by the encoder $\mathcal{E}$ into m encoded bits. The decoder is a map $\mathcal{D}$ from $\{0,1\}^m \times \mathcal{Y}^n$ to $\mathcal{X}^n$ if there is side-information available, and from $\{0,1\}^m$ to $\mathcal{X}^n$ otherwise. The reconstruction at the decoder is denoted $\widehat{x}_1^n$. To make the block error probability $\Pr(x_1^n \neq \widehat{x}_1^n)$ smaller than some positive value $\epsilon > 0$, roughly $m \sim nH(x)$ bits are needed [23] if there is no side-information and around $m \sim nH(x|y)$ bits otherwise [24].

For both problems, the number of bits needed by a scheme *above* the asymptotic rate h is defined to be its redundancy rate $\mathcal{R}(n, \epsilon, h)$:

$$\mathcal{R}(n, \epsilon, h) = \frac{m(n, \epsilon)}{n} - h \tag{2}$$

where $h = H(x)$ for point-to-point coding and $h = H(x|y)$ for coding with side information. Given h, finding $\mathcal{R}(n, \epsilon, h)$ and finding $m(n, \epsilon)$ are equivalent.

[1] This often occurs when the natural scale of data production is slow relative to the latency requirements.

Care has to be taken in defining the redundancy of a fixed-length code that must work for a set of possible source distributions — a compound source. Whereas in variable-length coding it makes sense to let the benchmark h move along with the underlying distribution, that is not meaningful in fixed-length coding. Instead, h should be set to the supremum of possible h since the spirit of universal fixed-length coding requires that the code satisfy the $\leq \epsilon$ target probability of block-error for every possible source distribution in the set.

In [3], [20], the redundancy rate for *non-universal* lossless source coding with decoder side-information is studied for $\mathcal{X} = \mathcal{Y} = \{0, 1\}$) and the source x and side-information y connected by a binary symmetric channel.[2] By using the central limit theorem (CLT) on the number of flips of the binary symmetric channel, they showed that the redundancy rate is:

$$\mathcal{R}(n, \epsilon, q) = \frac{\Phi^{-1}(\epsilon)\sqrt{q(1-q)}\log(\frac{1-q}{q})}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad (3)$$

where Φ^{-1} is the inverse Gaussian error function. The CLT-based analysis is valid only for rates $\frac{m}{n}$ in the neighborhood of $H(x|y)$ and meaningful only as n gets large. When the joint distribution p_{xy} is unknown to the coding system, the authors in [3] studied the special case of linked encoders [18], where it is assumed that the encoder knows the empirical distribution of the side information y_1^n. The redundancy rate for general universal coding without linked encoders remained open.

The focus in this paper is to go beyond these asymptotic results[3]. In place of the CLT, error exponent results are employed to bound redundancy. Using our universal lower bound on the lossless source coding error exponents from [5], achievable redundancy rates are obtained for coding with/without decoder side-information. The resulting bounds are universal in that they only depend on the source's (conditional) entropy rate and alphabet, not on its detailed distribution. Because the bounds are valid for all n, ϵ, they even apply when the alphabet-size (Section III-D) or required error probability ϵ varies along with n. In such cases, as pointed out in [3], the CLT based analysis is not valid.

Following [2], a similar story is told for discrete memoryless channel coding where there are only a finite n channel uses available and the goal is to see how many bits m can be transmitted reliably using them. Denote by $\mathcal{R}(n, \epsilon)$ the number of payload bits per channel-use *below* the channel capacity that can be communicated while still guaranteeing an average decoding error probability $\leq \epsilon$ using only n channel uses.

$$\mathcal{R}(n, \epsilon) = C - \frac{m(n, \epsilon)}{n} \quad (4)$$

where C is the DMC channel capacity. In this case, the relevant universal lower-bound on the error exponent comes from exercise 5.23 of [13].

[2] So $x = y \oplus z$ where z is another independent Bernoulli random variable with $\Pr(z = 1) = q < 0.5$.

[3] No $o(\cdot)$ terms are allowed since these are only meaningful in an asymptotic sense.

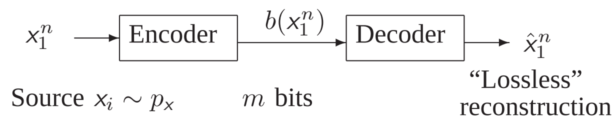

Fig. 1. Lossless fixed-length source coding

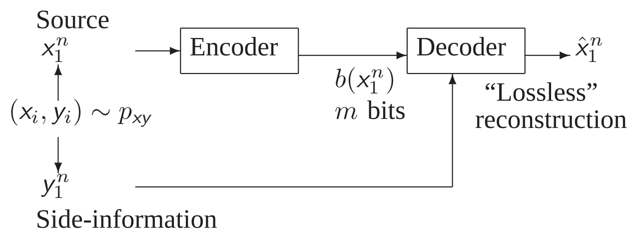

Fig. 2. Lossless source coding with decoder side-information

II. A Universal Quadratic Lower Bound on Fixed-length Block Coding Error Exponents

Channel coding error exponents are classical (see Chapter 5 of [13]) and the key feature is that Gallager's error exponents are non-asymptotic. We review lossless source-coding exponents and then quadratic lower bounds for both them and channel coding.

A. Lossless source coding error exponents

We summarize the relevant error exponent results from the literature [7], [11], [13]:

Lemma 1: [11] Assume the decoder has access to side information and the memoryless source and side information come from p_{xy}. If $R = \frac{m}{n}$, then a random binning encoder and jointly ML decoding system, illustrated in Figure 2, has an average error probability

$$\Pr(\hat{x}_1^n \neq x_1^n) \leq 2^{-nE_r(R)} \quad (5)$$

$$E_r(R) = \max_{0 \leq \rho \leq 1} \rho R - \bar{E}_0(\rho) \quad (6)$$

$$\text{where } \bar{E}_0(\rho) = \log_2\left(\sum_y (\sum_x p_{xy}(x, y)^{\frac{1}{1+\rho}})^{1+\rho}\right)$$

This bound holds $\forall n > 0$ — it is not asymptotic.

Without decoder side information, $\bar{E}_0$ simplifies to:

$$E_0(\rho) = (1+\rho)\log_2\left(\sum_{x\in\mathcal{X}} p_x(x)^{\frac{1}{1+\rho}}\right)$$

In Lemma 1, the random binning scheme at the encoder is uniform and thus universal in nature [11]. To do ML, however, the decoder needs to know the joint distribution. In [7], a decoder based on minimum entropy is shown to achieve the same error exponent asymptotically. For the universal decoder,

$$Pr(\hat{x}_1^n \neq x_1^n) \leq 2^{-n(E_r(R)-\phi(n))} \tag{7}$$

Where $\phi(n)$ is the vanishing term $\frac{2|\mathcal{X}|\log_2 n}{n}$ without side-information and $\phi(n) = \frac{2(|\mathcal{X}||\mathcal{Y}|-1)\log_2 n}{n}$ with decoder side-information.

B. Quadratic lower bounds to error exponents

1) Channel coding: Quadratic lower bounds to the error exponent function date back to [4] where they were first introduced to prove a coding theorem for compound channels. The bound there was valid only within $\frac{1}{2}$ nat of capacity and said that an error-exponent (here translated to base 2) of at least $\frac{(C-R)^2}{16|\mathcal{X}||\mathcal{Y}|\ln 2}$ was attainable[4] for a discrete memoryless channel $P(\cdot|\cdot)$, where $\mathcal{X}, \mathcal{Y}$ are the finite channel input and output alphabets. [15] gives a substantial improvement to this bound.

In exercise 5.23 of [13], an alternative quadratic lower bound is derived. Suppose that Q is the capacity-achieving distribution, then the random coding error exponent $E_r^c(R,Q)$, is lower bounded for all $R < C$.

$$E_r^c(R,Q) \geq \frac{\ln(2)(C-R)^2}{8/e^2 + 4[\ln|\mathcal{Y}|]^2} \tag{8}$$

We fixed a bug[5] in the original proof and further tightened the bound [5]. The bound (8) is universal in the formal sense that it only depends on the size of the output alphabet and the gap to capacity $(C-R)$ but not on the detailed channel statistics.

2) Source coding: These same techniques also yield bounds for lossless source-coding with and without decoder side-information.

Lemma 2: [5] For a memoryless source x and decoder side information y, jointly generated iid from p_{xy} with conditional entropy $H(x|y) = h$ on finite alphabet $\mathcal{X} \times \mathcal{Y}$, the random coding error exponent $E_r(R)$, defined in (6), is lower bounded $\forall R \in [H(x|y), \log_2|\mathcal{X}|)$:

$$E_r(R) \geq \begin{cases} \frac{\log_2 e}{2(\log_2|\mathcal{X}|)^2}(R-h)^2 & \text{if } |\mathcal{X}| \geq 3 \\ \frac{1}{2}(R-h)^2 & \text{if } |\mathcal{X}| = 2 \end{cases} \tag{9}$$

To simplify notation, we use the following slightly looser bound:

$$E_r(R) \geq \frac{(R-h)^2}{2(\log_2|\mathcal{X}|)^2} \tag{10}$$

If there is no side-information, the source is from p_x, s.t. $H(x) = h$ and the same bound applies.

The proof details are in [5]. For both cases, the quadratic bounds are only determined by the gap to entropy and the size of the source alphabet $\mathcal{X}$. Curiously, the bound (10) has no dependence on the size of the side-information alphabet $\mathcal{Y}$.

III. Upper bounds on the fixed-length block coding redundancy rates

In this section we present the main result of the paper, upper bounds on the redundancy rate for fixed-length coding. In Sections III-A and III-B, upper bounds are derived for lossless source coding both with and without decoder side-information in both the universal and non-universal contexts based on Lemmas 1 and 2. Section III-C then derives similar bounds for channel coding.

A. Source coding with decoder side-information

The non-universal context (where the source model is known to the decoder) is a a straightforward corollary of Lemmas 1 and 2.

Corollary 1: For lossless source coding with decoder side-information with a memoryless source, for any fixed block length $n > 0$, and any target error probability $\epsilon \in (0,1]$, there exits a random code such that the redundancy rate $\mathcal{R}(n,\epsilon,|\mathcal{X}|)$ is at most $\frac{1}{n} + (\log_2|\mathcal{X}|)\sqrt{\frac{2\log_2(\epsilon^{-1})}{n}}$ as long as the decoder knows the joint source distribution.

Proof: From (5), we know that for rate at $\frac{m}{n}$,

$$Pr(\hat{x}_1^n \neq x_1^n) \leq 2^{-nE_r(\frac{m}{n})}$$

By the bound in (10), we know that

$$E_r(\frac{m}{n}) > \frac{(\frac{m}{n} - H(x|y))^2}{2(\log_2|\mathcal{X}|)^2}$$

Hence a sufficient condition for the error probability $Pr(\hat{x}_1^n \neq x_1^n)$ to be smaller than the target error probability ϵ is that

$$2^{-n\frac{(\frac{m}{n}-H(x|y))^2}{2(\log_2|\mathcal{X}|)^2}} \leq \epsilon$$

[4] This bound was then used in [4] to derive a bound on the maximum probability of error over a set of channels by implicitly quantizing the set of channels into a finite number of representative channels and then adapting the standard argument of expurgation of bad messages. This paper ignores this issue and just allows random codes to be used.

[5] This bug was also alluded to by [16], but they got around it by proving a weaker result that was sufficient for their purposes.

By the definition of redundancy rate $\mathcal{R}(n,\epsilon)$ in (2) and the fact that m is an integer introduces an additional redundancy of at most $\frac{1}{n}$, we conclude that:

$$\mathcal{R}(n,\epsilon,|\mathcal{X}|) = (\log_2|\mathcal{X}|)\sqrt{\frac{2\log_2(\epsilon^{-1})}{n}} + \frac{1}{n} \qquad (11)$$

is a sufficient condition for $Pr(\hat{x}_1^n \neq x_1^n) \leq \epsilon$. □

Taking an asymptotic perspective, for fixed target error probability ϵ, we achieve an $O((\log_2|\mathcal{X}|)\sqrt{\frac{\log_2\epsilon^{-1}}{n}})$ redundancy rate —- this is consistent in *order*[6] with [25] and later in [3]. From the expression of the upper bound on the redundancy rate in (11), it is clear that our bound only depends on the source alphabet size $|\mathcal{X}|$, the source block length n and the target error probability.

Since this was derived using error exponent techniques and there are no $o(\cdot)$ terms hiding anywhere, our analysis holds for *any* target error probability $\epsilon \in (0,1]$. In particular, ϵ can be a function of source block length n, $\epsilon(n)$. If the target block error probability $\epsilon(n) = \frac{1}{f(n)}$ where $f(n)$ is a polynomial function of n, then the redundancy goes to 0 as n goes to infinity. In contrast, if $\epsilon(n)$ is exponential $\epsilon(n) = 2^{-nE}$ in n, then by applying Corollary 1 and noticing that the error probability is zero if the rate is above $\log_2|\mathcal{X}|$, we know that the redundancy rate is no larger than

$$\mathcal{R}(n,2^{-nE},|\mathcal{X}|) = \min\{(\log_2|\mathcal{X}|)\sqrt{2E}, \log_2|\mathcal{X}| - H(x|y)\} \qquad (12)$$

This bound does not tend to zero as n gets large.

When the distribution is unknown, we can use minimum conditional empirical entropy decoding.

Corollary 2: For lossless source coding with decoder side-information with a memoryless source, for any fixed block length $n > 0$, and any target error probability $\epsilon \in (0,1]$, there exits a random code such that the redundancy rate $\mathcal{R}(n,\epsilon,|\mathcal{X}|,|\mathcal{Y}|,h)$ is at most $\frac{1}{n} + (\log_2|\mathcal{X}|)\sqrt{\frac{2\log_2(\epsilon^{-1})+4(|\mathcal{X}||\mathcal{Y}|-1)\log_2(n)}{n}}$ even if the encoder/decoder know[7] the joint distribution p_{xy} only up to its conditional entropy $H(x|y) \leq h$.

Proof: Follows that of Corollary 1 except that it starts with (7) and has to account for the $\phi(n)$ term in the error probability. By using the universal lower bound on the random coding error exponent $E_r(\frac{m}{n})$ in Lemma 2, we conclude that the redundancy rate is at most

$$\mathcal{R}(n,\epsilon,|\mathcal{X}|,|\mathcal{Y}|,h) = \frac{1}{n} + (\log_2|\mathcal{X}|)\sqrt{\frac{2\log_2(\epsilon^{-1}) + 4(|\mathcal{X}||\mathcal{Y}|-1)\log_2 n}{n}} \qquad (13)$$

This proves the desired result. □

Comparing the non-universal redundancy in (11) to the universal one in (13), the only difference is the vanishing term $\frac{4(|\mathcal{X}||\mathcal{Y}|-1)\log_2(n)}{n}$ that comes from the union bound analysis [7] in the universal decoding case. In particular, the achievable redundancy for universal coding depends on the side-information alphabet size $|\mathcal{Y}|$ which is not the case for non-universal coding. Furthermore, for fixed target error probability $\epsilon > 0$, this is the dominant term in the redundancy as n gets large. This term is comparable to the one coming from the error probability even when $\epsilon(n) = \frac{1}{f(n)}$ where $f(n)$ is a polynomial function of n.

From the above, we see that any improvement over the union bound analysis for universal decoding would also improve the redundancy rate result here. In the next section, a *type matching decoder* is used to tighten the redundancy rate for lossless source coding without side-information.

B. Lossless source coding without decoder side-information

Just as in the previous section, by using Lemmas 1 and 2, the redundancy rate for the case without side-information can easily be upper bounded by

$$\mathcal{R}(n,\epsilon,|\mathcal{X}|) = (\log_2|\mathcal{X}|)\sqrt{\frac{2\log_2(\epsilon^{-1})}{n}} + \frac{1}{n} \qquad (14)$$

when the source distribution is known. For universal coding when the source distribution is unknown, we have the upper bound

$$\mathcal{R}(n,\epsilon,|\mathcal{X}|) = \frac{1}{n} + (\log_2|\mathcal{X}|)\sqrt{\frac{2\log_2(\epsilon^{-1}) + 4(|\mathcal{X}|-1)\log_2(n)}{n}} \qquad (15)$$

which is obtained by replacing $|\mathcal{Y}|$ with 1 in (13).

The redundancy result can be improved by using the universal source coding system without side-information depicted in Figure 3.

Corollary 3: For point-to-point lossless source coding, for any fixed block length n, and any positive target error probability ϵ, there exits a code such that the redundancy rate $\mathcal{R}(n,\epsilon,h)$ is at most

$$(\log_2|\mathcal{X}|)\sqrt{\frac{2\log_2(\epsilon^{-1})}{n}} + \frac{(|\mathcal{X}|-1)\log_2(n)}{n} + \frac{2}{n} \qquad (16)$$

[6] In (3), the term $\Phi^{-1}(\epsilon)$ can be approximated by $\sqrt{\log_2(\epsilon^{-1})}$ for small ϵ, hence we claim consistency in order.

[7] This is a departure from traditional notions of universality in which the joint distribution is only known up to some *qualitative* properties like being memoryless, Markov, etc. Here, it is also assumed to be known in a partially *quantitative* manner so that fixed-block coding makes sense. If the distribution was entirely unknown, then in worst case it would have independent side-information and a uniform source distribution. Thus only a rate of $\log|\mathcal{X}|$ would suffice. As pointed out in [5], knowing $H(x|y) \leq h$ is a non-convex constraint on the joint distribution.

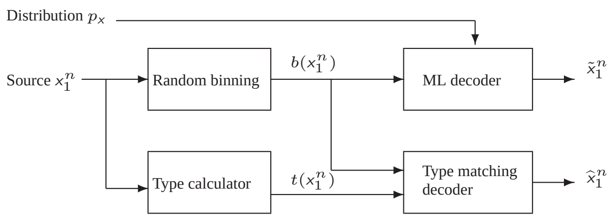

Fig. 3. Universal type matching decoding is as good as ML decoding

even if both the encoder and decoder are ignorant of the underlying source distribution except knowing the entropy rate h.

Proof: The encoder illustrated in Figure 3 has two parts. The first part is a type calculator. The output is the type index of the sequence $t(x_1^n)$. The number of types is upper bounded by $n^{|\mathcal{X}|-1}$, thus the number of bits to describe the type is at most $\lceil (|\mathcal{X}|-1)\log_2(n)\rceil$. The second part of the encoder is uniform random binning as in the previous section. The index of the bin is denoted by $b(x_1^n) \in \{0,1\}^m$

The decoder is a *type matching decoder*. First we define the set

$$S(b(x_1^n)) = \{z_1^n : b(z_1^n) = b(x_1^n), \text{ and } t(z_1^n) = t(x_1^n)\}$$

The decoder picks a sequence in $S(b(x_1^n))$ as the estimate $\widehat{x}_1^n$. Obviously $x_1^n \in S(b(x_1^n))$, so if $|S(b(x_1^n))| = 1$, then the decoding is going to be $\widehat{x}_1^n = x_1^n$. If $|S(b(x_1^n))| > 1$, notice that all the sequences in $S(b(x_1^n))$ have the same type and thus the same probability under any distribution, and thus we can use any tie-breaker from *a* maximum-likelihood(ML) decoding rule. All that remains is to show that the type-matching decoder performs at least as good as the ML decoder if they use the same tie-breaker.

Suppose that we make a decoding error at the type matching decoder for some source sequence x_1^n, i.e. $|S(b(x_1^n))| > 1$ and the tie breaker favors another sequence $\widehat{x}_1^n \neq x_1^n$. Then the ML decoder using the same tie breaker rule is going to favor the same sequence $\widehat{x}_1^n$ if sequences with type $t(x_1^n)$ are the most likely ones within the same bin $b(x_1^n)$. If this type is not the most likely one in the bin, then the ML decoder is guaranteed to make an error. Either way, the ML decoder will make a decoding error as well. So the type matching coding system achieves an error probability of at most $2^{-nE_r(\frac{m}{n})}$ by using at most $m+(|\mathcal{X}|-1)\log_2(n)+1$ bits. Therefore, in order to achieve target error probability ϵ, by following the same argument in Corollary 1, we conclude that the redundancy rate for universal lossless source coding is at most (16). □

C. Channel coding

The derivation of the universal upper bound on channel coding redundancy rate defined in (4) is exactly the same as that for source coding with decoder side-information. For a random code with input distribution Q and universal minimum mutual information decoding, it is shown that [12] the decoding error

$$P_e \leq 2^{-n(E_r^c(R,Q) - \frac{2|\mathcal{X}||\mathcal{Y}|\log_2(n)}{n})} \tag{17}$$

A straightforward consequence of (17) and (8) is that:

$$\mathcal{R}(n,\epsilon,|\mathcal{X}|,|\mathcal{Y}|) = \frac{1}{n} + \sqrt{\frac{\frac{8}{e^2}+4(\ln|\mathcal{Y}|)^2}{\ln 2} \times \frac{\log_2(\epsilon^{-1})+2(|\mathcal{X}||\mathcal{Y}|-1)\log_2(n)}{n}} \tag{18}$$

where the redundancy rate is the number of bits per channel use *below* the channel capacity that can be reliably delivered.

D. Redundancy rates for coding with large alphabets

To show the utility of our universal bounds on redundancy rate for fixed-length block coding, we investigate redundancy rates for coding with large alphabets as inspired by [21]. For fixed target error probability $\epsilon > 0$ and fixed finite alphabet size $|\mathcal{X}| < \infty$, the redundancy rate $\mathcal{R}(n,\epsilon)$ clearly converges to 0 as n goes to infinity. However, if the alphabet size also grows with block length n, the convergence is not guaranteed.

Since our upper bounds (11) (13) (16) and (18) on redundancy are valid for any target error probability $\epsilon \in (0,1]$, any block length $n > 0$ and alphabets $\mathcal{X},\mathcal{Y}$ of any sizes, we can let *everything* vary, be it n, ϵ or $|\mathcal{X}|$. One natural question is how fast can the alphabet size grow with block length n, while still not *not* requiring any more rate asymptotically. We list some sufficient conditions for $\mathcal{R}(n,\epsilon,|\mathcal{X}|,|\mathcal{Y}|)$ to converge to 0 as n goes to infinity:

- $|\mathcal{X}||\mathcal{Y}| = O(n^{1-\delta})$ for some $\delta > 0$ for universal source coding with decoder side-information
- $|\mathcal{X}| = o(\frac{n}{\log_2 n})$ for universal lossless source coding

In contrast $|\mathcal{X}|$ only needs to be $o(2^{\sqrt{n}})$ for source coding with known statistics.

Similarly, the redundancy penalty for channel coding converges to 0 if $|\mathcal{X}||\mathcal{Y}| = O(n^{1-\delta})$ for some $\delta > 0$. Previously, universal channel coding with large alphabets was studied in the arbitrarily varying channel context [8]. The work here suggests an alternate perspective on channel coding with continuous alphabets. In particular, it reveals that universal channel coding with continuous alphabets is possible if the channel uncertainty model is structured so that a quantized channel's capacity converges to the continuous channel capacity at some uniform (over channels) rate with finer-and-finer quantization of both inputs and outputs. Then, as long as the number of quantization bins is increased so that the product of input and output quantization bins grows sublinearly in the block-length, the results here immediately give a capacity-achieving coding theorem.[8]

IV. Conclusions and Open Problems

For both non-universal and universal coding, upper bounds on the redundancy rate for fixed-length block coding have been derived for lossless point-to-point source coding, lossless source coding with decoder side-information and channel coding. Our bounds are simple and explicit functions of target error probability, block length and alphabet size. Because these results are non-asymptotic, the large alphabet asymptotics for both source coding and channel coding can also be explored and sufficient conditions given for the redundancy to tend to zero.

This work is a corollary of our previous work on universal lower bounds to error exponents in [5]. As the lower bound in [5] is loose, the resulting upper bound on the redundancy rate is also loose. However, we observe that it is correct in order since the order is consistent with known results for special cases.

Or results are restricted to almost-lossless scenarios. For lossy source coding, the redundancy rate problem is studied in the variable-length setup [9], [26]. It would be interesting to derive the fixed-length redundancy rate which is based on a universal bound for the error exponent for lossy source coding investigated in [17].

Acknowledgments

We thank Gil Shamir for thought provoking comments and questions at CISS 2007 when [5] was presented and Robert Gallager for helpful discussions. We also thank the anonymous reviewers for their very careful reading and insightful comments that improved the presentation of the paper tremendously.

References

[1] D. Baron, M. A. Khojastepour, and R. G. Baraniuk. How quickly can we approach channel capacity? *38th Asilomar*, 2004.

[2] D. Baron, M. A. Khojastepour, and R. G. Baraniuk. Redundancy rates of Slepian-Wolf coding. *42nd Allerton Conference*, 2004.

[3] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Information Theory*, 44:2743–2760, 1998.

[4] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Annals of Mathematical Statistics*, 30(4):1229–1241, December 1959.

[5] C. Chang and A. Sahai. Universal quadratic lower bounds on source coding error exponents. *41st CISS*, 2007.

[6] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons Inc., New York, 1991.

[7] Imre Csiszár and János Körner. *Information Theory*. Akadémiai Kiadó, Budapest, 1986.

[8] I. Csiszar. Arbitrarily varying channels with general alphabets and states. *IEEE Trans. Inform. Theory*, 38:1725–1742, Novermber 1992.

[9] A. Dembo and I. Kontoyiannis. Critical behavior in data compression. *Technical Report no. TR-99-26, Department of Statistics, Purdue University*, 1999.

[10] S. Draper and A. Sahai. Universal anytime coding. *Control over Communication Channels Workshop*, Limassol, Cyprus, 2007.

[11] Robert Gallager. Source coding with side information and universal coding. *LIDS-P-937*, 1979.

[12] Robert Gallager. A random coding bound on fixed-composition codes. *http://web.mit.edu/gallager/www/notes/notes6.pdf*, 1992.

[13] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley, New York, NY, 1971.

[14] I. Kontoyiannis. Second-order noiseless source coding theorems. *IEEE Trans. Inform. Theory*, 43:1339–1341, July 1997.

[15] Samuel Kotz. Exponential bounds on the probability of error for a discrete memoryless channel. *Annals of Mathematical Statistics*, 32(2):577–582, June 1961.

[16] Amos Lapidoth and Emre Telatar. The compound channel capacity of a class of finite-state channels. *IEEE Trans. Inform. Theory*, 44:973–983, May 1998.

[17] Katalin Marton. Error exponent for source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 20:197–199, 1974.

[18] Y. Oohama. Universal coding for correlated sources with linked encoders. *IEEE Trans. Information Theory*, 42:837–847, 1996.

[19] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[20] S. Sarvotham, D. Baron, and R. G. Baraniuk. Non-asymptotic performance of symmetric Slepian-Wolf coding. *39th CISS*, 2005.

[21] Gil Shamir. *Personal communication*, March 15th, 2007.

[22] Gil Shamir. On the MDL principle for i.i.d. sources with large alphabets. *IEEE Trans. Inform. Theory*, 52:1939–1955, May 2006.

[23] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[24] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19, 1973.

[25] J. Wolfowitz. *Coding theorems of information theory*. Springer-Verlag, New York, 1978.

[26] Zhen Zhang, Enhui Yang, and Victor K. Wei. The redundancy of source coding with a fidelity criterion. *IEEE Trans. Inform. Theory*, 43:71–91, January 1997.

[8]Thus, the results of [10] can be extended to give a sufficient condition for continuous feedback channels to be used to stabilize an unstable random process.