Variable-Rate Coding with Feedback for Universal Communication Systems*

Shriram Sarvotham, Dror Baron, and Richard G. Baraniuk
Dept. of Electrical and Computer Engineering, Rice University, Houston, TX 77005

Abstract

Classical coding schemes that rely on joint typicality (such as Slepian-Wolf coding and channel coding) assume known statistics and rely on asymptotically long sequences. However, in practice the statistics are unknown, and the input sequences are of finite length. In this finite regime, we must allow a non-zero probability of codeword error ϵ and also pay a penalty in the rate. The penalty manifests itself as redundancy in source coding and a gap to capacity in channel coding. Our contributions in this paper are two-fold. First, we develop a universal scheme for the specific example of the binary symmetric channel. With n channel uses and a fixed ϵ, the gap to capacity of our scheme is $O(\sqrt{n})$. The prior art for channel coding with known statistics shows that the penalty is $\Omega(\sqrt{n})$. Therefore, we infer for communication systems that rely on joint typicality that the penalty needed to accommodate universality is $\Theta(\sqrt{n})$. Second, we derive the penalty incurred in variable-rate universal channel coding when we restrict the number of rounds of feedback. Surprisingly, it turns out that using only two or three rounds of feedback provides near-optimal performance for systems of practical interest. This rule of thumb is valuable for practical communication systems.

1 Introduction

Classical coding theorems that rely on joint typicality such as the Slepian-Wolf theorem [1–3] and the channel coding theorem [2, 3] assume known statistics and asymptotically long sequences. In practice, however, the input sequences are of a finite length. Unfortunately, no matter what fixed rate we choose to encode the sequences, the sequences may be jointly atypical, and hence we may be incapable of reconstructing them at the decoder. Therefore, in this *finite-length regime* we must allow a non-zero probability of codeword error ϵ. Another problem that arises in practice is that the source or channel statistics are unknown, and we must use coding methods whose performance is *universal* with respect to a broad range of possible statistics. Of course these practical concerns impose a penalty in the coding rate. These penalties exist in both source coding (we must expend redundant bits to describe the source) and channel coding (we operate below capacity, i.e., the penalty is the gap to capacity).

In our previous work [4], we proposed a variable-rate scheme for the specific case of universal Slepian-Wolf coding. Our scheme used feedback from the decoder (Fig. 1 (a)) to allow the encoder to estimate the source statistics. The penalty needed in our scheme was shown to be $O(\sqrt{n})$,[1] where n was the number of source bits.[2] Because the penalty

*This work was supported by NSF, AFOSR, ONR, and the Texas Instruments Leadership University Program. Email: {shri, drorb, richb}@rice.edu; Web: dsp.rice.edu.

[1]For two functions $f(n)$ and $g(n)$, $f(n) = O(g(n))$ if $\exists c, n_0 \in \mathbb{R}^+,\ 0 \le f(n) \le cg(n)$ for all $n > n_0$. Similarly, $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$, and $f(n) = \Theta(g(n))$ if $\exists c_1, c_2, n_0 \in \mathbb{R}^+,\ 0 \le c_1 g(n) \le f(n) \le c_2 g(n)$ for all $n > n_0$. Finally, $f(n) = o(g(n))$ if for all positive $c > 0$, $\exists n_0 \in \mathbb{R}^+,\ 0 \le f(n) < cg(n)$ for all $n > n_0$.

[2]In Slepian-Wolf coding, we use n for the number of input bits and t for the number of codeword

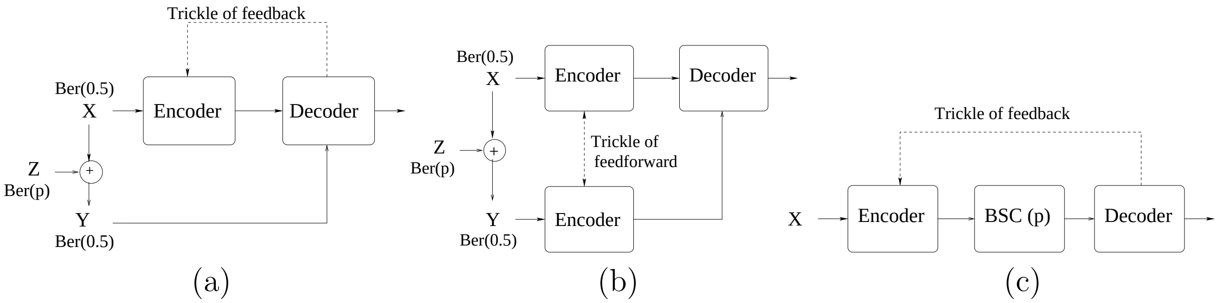

Figure 1: *(a) Universal Slepian-Wolf encoding with feedback from our previous work [4]. (b) Linked encoders for Slepian-Wolf encoding as considered by Kimura and Uyematsu [7]. (c) Universal channel coding with feedback.*

in the finite-length known statistics case is $\Omega(\sqrt{n})$ bits, we concluded that the penalty for universality with feedback is $\Theta(\sqrt{n})$ bits. In this paper, we show how ideas from our previous work can be extended to additional coding schemes that rely on joint typicality. Another key contribution of this paper is to address the case where the feedback from the decoder to the encoder is limited. We evaluate the penalty imposed in this system and show that in practical systems the cost of restricting the feedback is small.

1.1 Related work

We survey several related results that consider universality or non-asymptotic sources in coding schemes that rely on joint typicality. Because Slepian-Wolf coding and channel coding are closely related [2] to each other, we survey the related work in both areas.

Fixed rate universal Slepian-Wolf setup: Csiszar and Korner [5] and Oohama and Han [6] consider a Slepian-Wolf setup where the coding rates are fixed, and the goal is to achieve universal error exponents similar to the known-statistics exponents. This approach is limited, because the fixed rates may greatly exceed the Slepian-Wolf limit.

Variable rates via linked encoders: Kimura and Uyematsu [7] proposed variable rate Slepian-Wolf coding with *linked encoders*, where a trickle of feed-forward between encoders is allowed (Fig. 1 (b)). In this setup, the penalty also includes communication between the encoders. Their result can be summarized in the following theorem.

Theorem 1 [7] *For unknown statistics, a zero-rate penalty (redundancy over Slepian-Wolf limit) suffices for Slepian-Wolf coding using the linked encoder setup.*

Lower bound via fixed rate coding in finite regime: Theorem 1 implies that for a fixed ϵ the penalty in the linked encoder setup [7] is $o(n)$ bits. The limitation here is that the $o(n)$ term may be quite large. Can we do better? We begin with a lower bound by Wolfowitz [2] (see also Baron et al. [8, 9]) on the penalty for using finite length sequences in communication systems in which the statistics are *known*.

Theorem 2 $[2, 8, 9]$ *For known statistics and a fixed ϵ, the penalty for using finite length sequences in coding schemes that rely on joint typicality is $\Theta(\sqrt{n})$ bits.*

bits. In channel coding, we use t for the number of input bits and n for the number of channel uses. We use this *reversed notation* because the Slepian-Wolf correlation and the channel noise sequences are distributed Bernoulli(n, p). Similarly, the Slepian-Wolf codeword and channel coding input sequences are both distributed similar to Bernoulli$(t, 0.5)$.

Wolfowitz proves Theorem 2 for channel coding, Slepian-Wolf coding and also discusses additional communication systems that rely on joint typicality [2]. The theorem lower bounds the penalty needed to accommodate universality. This is because we cannot perform better for universality than the known statistics case. *However, Theorem 2 does not upper bound the penalty for universality.*

Upper bound via linked encoders: In our previous work [8], we have shown that a penalty of $O(n^{2/3})$ bits suffices in the linked encoder Slepian-Wolf setup. Considering the lower bound of Theorem 2, the challenge is then to resolve the gap between $\Omega(\sqrt{n})$ and $O(n^{2/3})$ bits.

Slepian-Wolf via feedback from decoder: Also in previous work [4], we considered universality and non-asymptotic codeword length aspects of Slepian-Wolf coding. The setup comprised two binary sources X and Y that generate length-n Bernoulli sequences x and y with Bernoulli parameter 0.5.[3] The sequences are correlated such that y can be obtained from x by adding a correlation sequence z, which is Bernoulli with unknown parameter p (Fig. 1 (a)). The decoder must reconstruct x and y with codeword error probability upper bounded by ϵ. The decoder is allowed to provide a small amount of (assumed error-free) feedback to the encoder in order to help the encoder estimate the unknown parameter p and use variable coding rates that approach the asymptotic limit.

The goal is to reconstruct x and y at the decoder using minimal total communication. Our scheme partitions x and y into blocks of geometrically increasing lengths and allocates error probabilities among the blocks. Each block is encoded at a different rate and feedback is provided at the end of each block. We proved the following theorem.

Theorem 3 [4] *For unknown statistics and a fixed ϵ, the penalty (redundancy) for Slepian-Wolf coding using our variable rate scheme with feedback is $O(\sqrt{n})$ bits.*

Universal channel coding: Having provided an order bound on the penalty for universality in Slepian-Wolf coding, we now consider channel coding. In this setting, the channel encoder cannot vary its rate without feedback from the decoder. We outline two such approaches. Tchamkerten and Telatar [10, 11] consider universal variable-rate schemes, where the encoder receives the channel output as feedback. For some channels (including the BSC), their schemes achieve the known-statistics error exponents. However, because the channel output is communicated to the encoder, they require $\Theta(n)$ bits of feedback, which we view as excessive. At the other extreme, rateless codes (c.f. Luby [12]) require minimal feedback. In this approach, the channel encoder continues transmitting until the decoder has received enough information to recover the input. For LT codes [12] that achieve error probability ϵ, the penalty using input sequences of length t is $O(\sqrt{t}\log^2(t/\epsilon))$ bits. Our scheme improves the penalty to $O(\sqrt{t}\sqrt{\log(1/\epsilon)})$.

1.2 Contributions

Order bound for universality: In this paper, we characterize the cost of universality for communication systems that rely on joint typicality. We show how our previous work [4] can be extended to additional communication systems (beyond Slepian-Wolf coding). As an example, we derive the penalty incurred to accommodate universality in channel coding. In our proposed scheme, we use variable-rate block-sequential channel coding. The input sequence is partitioned into blocks, and each block is encoded separately. The decoder transmits a trickle of feedback at the end of each block, which enables the

[3]Recall from Footnote 2 that we use reversed notation.

encoder to estimate the statistics and vary the coding rate. Although the derivation for the penalty associated with universal channel coding is similar to the derivation for Slepian-Wolf coding, we present it here for completeness.

Cost of restricted feedback: Our proposed solution may require many rounds of feedback from the decoder to the encoder. Although the bit-rate required for feedback is negligible relative to other terms discussed in this paper, multiple rounds of feedback increase the complexity in the control mechanisms of the communication system and thus may be difficult to implement. To address this concern, the second key contribution of this paper is to analyze systems that use few rounds of feedback. We derive the overall penalty incurred by such systems and show that two or three rounds of feedback may suffice in many communication systems of practical interest.

2 Problem statement

Consider Fig. 1 (c). A binary symmetric source X generates a length-t sequence x. The sequence is encoded at some variable rate and then transmitted over a binary symmetric channel (BSC) with unknown cross-over probability p. The decoder must reconstruct x with codeword error probability upper bounded by ϵ. The decoder may send the encoder a trickle of (assumed error-free) feedback. The goal is to reconstruct x at the decoder using minimal total communication.

3 Proposed solution

Main idea: We use variable-rate block-sequential channel coding. The encoder partitions x into blocks and encodes each block at a rate that depends on its current estimate of p. The decoder receives a block, decodes it, and then reveals the number of channel errors detected in the block to the encoder. Using this information, which is communicated via feedback, the encoder refines its estimate of p and moves to the next block.

Notation: Let the number of rounds of feedback be k. This implies that the input x is divided into $k+1$ blocks for block-transmission. Let b_i be the encoded length of the i^{th} block. Denote the total number of bits sent over the channel (not including feedback) by n, i.e., $n = \sum_{i=1}^{k+1} b_i$. Let $B_i = \sum_{j=1}^{i} b_j$ be the sum of the lengths of the first i blocks. We assume that the capacity C of the BSC satisfies $C > \eta$ for some $\eta > 0$ (this assumption is weak). Let the probability of error in block i be ϵ_i. We allocate the probability ϵ of codeword error among the different blocks, i.e., $\sum_{i=1}^{k} \epsilon_i = \epsilon$.[4] Let s be the total number of feedback bits. In particular, the i^{th} feedback round describes the number of bit errors in block i, and thus requires $\log_2(b_i)$ bits. Finally, we define the penalty r as

$$r \triangleq n - \frac{t}{C} + s.$$

Details: We propose that the block lengths increase as powers of 2 except for the last block; that is, $b_i = 2^{i-1}$ for $1 \leq i \leq k$. The last block contains the remaining $n - B_k = n - 2^k - 1$ bits. We assign the block error probabilities such that $\epsilon_i = \epsilon b_i/(2^{k+1} - 1)$, where we compute k (and therefore the ϵ_i's) based on η. (The assumption $C > \eta$ enables conservative design of the ϵ_i's.) If $C > \eta$, then the penalty incurred by our scheme increases by a *constant multiplicative factor*, which depends on C and η. This constant does not affect the order terms in the analysis. It is possible to reduce this constant using more complex schemes; we leave that option for future work.

[4] The union bound ensures that the probability of error in any of the blocks is upper bounded by ϵ.

4 Analysis of variable-rate scheme with feedback

Theorem 4 *For a BSC with unknown statistics, capacity $C > \eta$, and a fixed ϵ, the penalty of our variable-rate channel coding scheme with feedback is $r = O(\sqrt{n})$ bits.*

Parts of the proof that are tedious and less insightful are not included. See Sarvotham et al. [13] for the complete proof.

Proof of Theorem 4: We compute the penalty for block i as

$$r_i = c_1 \frac{b_i}{\sqrt{B_{i-1}}} \Phi^{-1}(\epsilon_i) + c_2 \sqrt{b_i} \Phi^{-1}(\epsilon_i) + o(\sqrt{b_i}),$$

where $\Phi^{-1}(.)$ is the inverse error function. The first term captures the penalty due to universality. The remaining terms account for the penalty in non-asymptotic channel coding [2, 9]. *The $\Phi^{-1}(\epsilon_i)$ comes from the Central Limit Theorem [2, 9]; backing off δ standard deviations from the channel capacity buys us $\Phi(\delta)$ probability of error.*

For blocks 1 through k we have $b_i = 2^{i-1}$. Therefore $B_i = 2^i - 1$ and so

$$r_i = c_1 \frac{2^{i-1}}{\sqrt{2^{i-1} - 1}} \Phi^{-1}(\epsilon_i) + c_2 \sqrt{2^{i-1}} \Phi^{-1}(\epsilon_i) + o(2^{\frac{i}{2}}).$$

The first and second terms (corresponding to penalties associated with universality and finite length coding) are of the same order and can be combined. Therefore,

$$r_i = c_3\, 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) + o(2^{\frac{i}{2}})$$

for $1 \leq i \leq k$. In block $k+1$, we can transmit a maximum of 2^k bits and so

$$r_{k+1} \leq c_3\, 2^{\frac{k+1}{2}} \Phi^{-1}(\epsilon_{k+1}) + o(2^{\frac{k+1}{2}}).$$

The total penalty r is the sum of the penalties r_i incurred over all the blocks along with feedback penalty s, yielding

$$r \leq \left[\sum_{i=1}^{k+1} c_3\, 2^{\frac{i}{2}}\, \Phi^{-1}(\epsilon_i) \right] + \left[\sum_{i=1}^{k+1} o(2^{\frac{i}{2}}) \right] + s,$$

where the inequality appears because of block $k+1$. It can be shown that $\sum_{i=1}^{k+1} o(2^{\frac{i}{2}}) = o(2^{\frac{k}{2}})$ [13], and so

$$r \leq \left[\sum_{i=1}^{k+1} c_3\, 2^{\frac{i}{2}}\, \Phi^{-1}(\epsilon_i) \right] + o(2^{\frac{k}{2}}) + s. \tag{1}$$

It can easily be shown that the feedback cost satisfies $s = O(\log^2(n))$. Therefore, s can be absorbed by the $o(2^{\frac{k}{2}})$ term. To simplify the summation term in (1), recall that $\epsilon_i = \epsilon b_i / (2^{k+1} - 1)$. Therefore $\epsilon_{i+1} = 2\epsilon_i$, and the following upper-bound on $\Phi^{-1}(\frac{\xi}{2}) / \Phi^{-1}(\xi)$ will be useful. Its proof appears in [13].

Lemma 1 *For any $\delta > 0$, we can find $\xi' > 0$ such that for all $\xi < \xi'$,*

$$\frac{\Phi^{-1}(\frac{\xi}{2})}{\Phi^{-1}(\xi)} < 1 + \delta.$$

We pick δ such that $1+\delta \in (1, \sqrt{2})$ and determine the corresponding ξ' for which $\frac{\Phi^{-1}(\xi/2)}{\Phi^{-1}(\xi)} < 1+\delta$ for all ξ smaller than ξ'.[5] Because Lemma 1 is applicable only when $\epsilon_i < \xi'$, we split the summation in (1) into two sums, such that Lemma 1 can be applied to the $\Phi^{-1}(\epsilon_i)$ terms in the first sum:

$$r \le \left[\sum_{i=1}^{m} c_3\, 2^{\frac{i}{2}}\, \Phi^{-1}(\epsilon_i)\right] + \left[\sum_{i=m+1}^{k+1} c_3\, 2^{\frac{i}{2}}\, \Phi^{-1}(\epsilon_i)\right] + o(2^{\frac{k}{2}}). \tag{2}$$

If $\epsilon_{k+1} \ge \xi'$, then the index m is chosen such that $\epsilon_m < \xi'$ and $\epsilon_{m+1} > \xi'$. Else $\epsilon_{k+1} < \xi'$, and we choose $m = k+1$. Consequently, the second summation is empty. Because $\epsilon_{i-1} = \epsilon_i/2$, we have $\epsilon_i < \xi'$ for $i \le m$, and hence the result from Lemma 1 can be applied to all the terms in the first summation. Denote the first summation in (2) by S_1. By change of variables $j \leftarrow m-i$, we have

$$S_1 = \sum_{i=1}^{m} c_3\, 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) = \sum_{j=0}^{m-1} c_3\, 2^{\frac{m-j}{2}} \Phi^{-1}(\epsilon_{m-j}).$$

From Lemma 1, $\Phi^{-1}(\epsilon_{m-j}) = \Phi^{-1}(\frac{\epsilon_m}{2^j}) < \Phi^{-1}(\epsilon_m)(1+\delta)^j$, hence S_1 can be bounded by

$$\begin{aligned} S_1 &< c_3\, 2^{\frac{m}{2}}\, \Phi^{-1}(\epsilon_m) \sum_{j=0}^{m-1} \left(\frac{1+\delta}{\sqrt{2}}\right)^j \\ &< c_3\, 2^{\frac{m}{2}}\, \Phi^{-1}(\epsilon_m)\, O(1) \quad \text{(summing the geometric series)} \\ &= O(2^{\frac{k}{2}} \Phi^{-1}(\epsilon_m)) \quad \text{(because } m \le k+1 \text{).} \end{aligned} \tag{3}$$

Note the dependence on $\Phi^{-1}(\epsilon_m)$. If ϵ is large enough, then $\epsilon_m \in (\xi'/2, \xi')$ and $\Phi^{-1}(\epsilon_m)$ is $O(1)$. However, if ϵ is small, then $m = k+1$, in which case $\Phi^{-1}(\epsilon_m) = \Phi^{-1}(\epsilon_{k+1}) = O(\Phi^{-1}(\epsilon))$. The second summation in (2), which we denote by S_2, can be simplified as

$$\begin{aligned} S_2 &= \sum_{i=m+1}^{k+1} c_3\, 2^{\frac{i}{2}}\, \Phi^{-1}(\epsilon_i) \\ &< c_3\, \Phi^{-1}(\xi') \sum_{i=m+1}^{k+1} 2^{\frac{i}{2}} \quad \text{(because } \Phi^{-1}(\epsilon_i) < \Phi^{-1}(\xi')\text{)} \\ &= O(2^{\frac{k}{2}}) \quad \text{(because } \sum_{i=m+1}^{k+1} 2^{\frac{i}{2}} = O(2^{\frac{k}{2}}) \text{ and } \Phi^{-1}(\xi') \text{ is constant).} \end{aligned} \tag{4}$$

Note that S_2 is often empty in practice because ξ' is large (Footnote 5). Combining (2)-(4), we infer that $r = O(2^{\frac{k}{2}})$. Because $n \in [2^k, 2^{k+1})$, $r = O(\sqrt{n})$. □

The $\Phi^{-1}(\epsilon)$ term in S_1 yields $r = O(\sqrt{n}\Phi^{-1}(\epsilon)) = O(\sqrt{n}\sqrt{\log(1/\epsilon)})$. Because $\Phi^{-1}(\epsilon)$ also appears in refined versions of Theorem 2 [2, 8, 9], we have a tight order bound.

The arguments detailed in this proof can be applied to additional communication systems that rely on joint typicality. In these systems, we account for the penalties associated with universality and non-asymptotics. Our analysis determines these penalties using the central limit theorem; the main idea is that backing off δ standard deviations from the asymptotic performance limits buys us $\Phi(\delta)$ probability of error.

---

[5] For $\delta = 0.4$, the value of ξ' is about 0.3093. This indicates the typical values of ξ' that we can expect in practical settings.

5 Coding with finitely many rounds of feedback

In the variable-rate scheme of Section 3, we require $k = O(\log(t))$ rounds of feedback, where $t = O(n)$ is the length of the input sequence. For large t, k may become prohibitively large. As noted in Section 1.2, this may impose unreasonable demands on a practical communication system in terms of control and implementation. To address these concerns, we now describe a modified scheme in which we fix k.

In our scheme, we transmit $k+1$ blocks and use k rounds of feedback. Given that the input consists of t bits, in block i we encode $e_i(t,k)$ input bits,

$$e_i(t,k) = \Theta\left(t^{2^{k-i+1}\frac{2^i-1}{2^{k+1}-1}}\right),$$

where we assume that $e_i(t,k)$ are integers.[6] This choice of block lengths yields minimal order terms for the penalty, as shown below. We also choose $e_{k+1}(t,k)$, the length of the last block, such that the remaining bits are encoded.

The rate at which block i is encoded is chosen according to the current parameter estimate, in order to ensure that the probability of error in each of the $k+1$ blocks will be less than $\frac{\epsilon}{k+1}$. Again, the union bound ensures that the total probability of error is less than ϵ, as required. Note also that we assume $C > \eta$ to provide an initial parameter estimate in the first block. The resulting penalty is given by the following theorem.

Theorem 5 *For a BSC with unknown statistics, capacity $C > \eta$, a fixed ϵ, and any $\delta > 0$, the penalty in our modified scheme using k rounds of feedback is $r(t,k)$ bits, where*

$$r(t,k) = O\left(t^{\frac{2^k}{2^{k+1}-1}+\delta}\right). \tag{5}$$

Proof: We use induction to establish the result. The *inductive hypothesis* on k is that $r(t,k)$ satisfies the theorem statement of the theorem for up to k rounds of feedback.

Verifying the Hypothesis for $k=1$: For $k=1$, we use 2 blocks. Therefore, we encode $e_1(t,1) = \Theta(t^{\frac{2}{3}})$ and $e_2(t,1) = \Theta(t)$ input bits in the 2 blocks. The penalty is

$$r(t,1) = c_1 e_1(t,1) + c_2 \frac{e_2(t,1)}{\sqrt{e_1(t,1)}} + c_3\sqrt{e_2(t,1)},$$

where the first term accounts for the penalty in the first block (note that the assumption $C > \eta$ ensures that c_1 is finite) and the remaining terms account for the penalty in the second block.[7] We can now write

$$r(t,1) = c_1\Theta(t^{\frac{2}{3}}) + c_2\frac{\Theta(t)}{\sqrt{\Theta(t^{\frac{2}{3}})}} + c_3\sqrt{\Theta(t)} = \Theta(t^{\frac{2}{3}}).$$

Clearly, for any $\delta > 0$ (even $\delta = 0$) the penalty r satisfies the inductive hypothesis (5).

Extending the inductive hypothesis from k **to** $k+1$: We assume that the hypothesis is correct up to k rounds of feedback and then add another round of feedback. We must prove that the redundancy satisfies (5).

[6] Because $e_i(t,k)$ are *order terms*, rounding to the nearest integer does not change our results.

[7] The constants c_1, c_2,... that appear in the proof of Theorem 5 are different from the constants used in the proof of Theorem 4.

First, let us consider the first $k+1$ blocks of our scheme out of $k+2$ blocks in total. We define the total number of bits encoded during these rounds as ψ, and have

$$\begin{aligned} \psi &\triangleq \sum_{i=1}^{k+1} e_i(t, k+1) \\ &= \Theta(e_{k+1}(t, k+1)) \\ &= \Theta\left(t^{2\frac{2^{k+1}-1}{2^{k+2}-1}}\right) = t^{2\frac{2^{k+1}-1}{2^{k+2}-1}+\Theta(1/\log(t))}, \end{aligned} \tag{6}$$

where the last term of the summation determines its order (6). Now consider encoding $e_i(\psi, k)$ bits in block i. In words, we use the policy for k rounds during the first $k+1$ blocks of the policy for $k+1$ rounds, which uses $k+2$ blocks. In this case,

$$\begin{aligned} e_i(\psi, k) &= \Theta\left(\psi^{2^{k-i+1}\frac{2^i-1}{2^{k+1}-1}}\right) = \Theta\left(t^{\left[2\frac{2^{k+1}-1}{2^{k+2}-1}+\Theta(1/\log(t))\right]\left[2^{k-i+1}\frac{2^i-1}{2^{k+1}-1}\right]}\right) \\ &= \Theta\left(t^{\left[2^{k-i+2}\frac{2^i-1}{2^{k+2}-1}\right]+o(1)}\right), \end{aligned}$$

where the $o(1)$ decays as t increases. Because the order term for $e_i(\psi, k)$ is identical to that for $e_i(t, k+1)$ except for the $o(1)$ term, for any fixed $\theta > 0$ we have

$$e_i(t, k+1) = o\left(e_i(\psi^{1+\theta}, k)\right).$$

Invoking the inductive hypothesis (5), the penalty during the first $k+1$ blocks is $o(r(\psi^{1+\theta}, k))$. Therefore, for any $\delta > 0$ there corresponds $\theta > 0$ for which $r(\psi^{1+\theta}, k)$ upper bounds the penalty during the first $k+1$ blocks. Increasing δ further will account for the more stringent $\frac{\epsilon}{k+2}$ that we require in each round.

We now consider the penalty r_{k+2} in the last block. This penalty is of the form

$$r_{k+2} = c_4 \frac{e_{k+2}(t, k+1)}{\sqrt{e_{k+1}(t, k+1)}} + c_5\sqrt{e_{k+2}(t, k+1)}, \tag{7}$$

where the terms are similar to before, and $e_{k+1}(t, k+1)$ accounts for the parameter estimate derived from the first $k+1$ blocks. Note, however, that c_4 and c_5 must satisfy the more stringent probability of error $\frac{\epsilon}{k+2}$, which is resolved by introducing the δ term. Therefore, the first term of (7) is of order $O(t^{\frac{2^{k+1}}{2^{k+2}-1}+\delta})$ and the second term is of order $O(t^{1/2+\delta})$. Combining these terms with the penalty for the first $k+1$ blocks, we have verified the hypothesis for $k+1$ rounds of feedback. □

Note that we do not have a converse bound for the penalty using a fixed number of rounds of feedback. Rather, we have shown that it is possible to achieve universality while paying a little extra penalty even if we restrict the number of rounds of feedback.

To investigate the potential practical impact of Theorem 5, we present numerical results. We compute the minimum penalty for a given t and k using dynamic programming. Fig. 2 plots the penalty r versus t for several fixed values of k. For small k, we benefit significantly from each additional round of feedback (for large t). However, for $k > 3$, the gains from additional feedback rounds are marginal. Therefore, in practice we obtain near-optimal performance using only a few rounds of feedback. Finally, we note that for small values of t, increasing k degrades the performance. This can be attributed to the constants in the order terms.

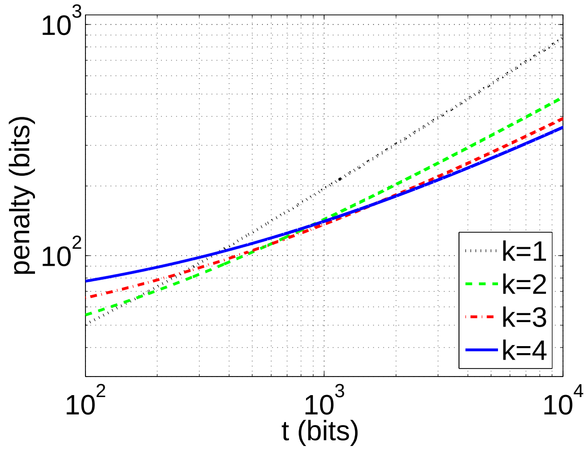

Figure 2: *Numerical results depicting the performance with finite rounds of feedback.*

6 Discussion and conclusions

For communication systems that rely on joint typicality, we have shown that the penalty due to universality is $O(\sqrt{n})$ bits. The prior art for known statistics lower bounds the penalty by $\Omega(\sqrt{n})$; hence it is $\Theta(\sqrt{n})$. We have shown how to achieve this limit by proposing a variable-rate coding scheme with feedback. Our scheme was demonstrated for universal channel coding over the binary symmetric channel — the total number of channel uses (including feedback) must exceed the ideal information-theoretic-limit by only $O(\sqrt{n})$. The scheme can be applied to any communication system in which the decoder relies on joint typicality. Example applications include compression of encrypted data [14, 15] and distributed compression of sensor network data [16].

Our proposed solution may require many rounds of feedback from the decoder to the encoder. This may be difficult to implement because of increased complexity. We addressed this issue by deriving the overall penalty incurred for k rounds of feedback in Theorem 5. Even for very small values of k, the penalty comes very close to $O(\sqrt{n})$. This suggests that a small number of rounds of feedback may be sufficient in practice. For example, with two rounds of feedback, the penalty is $O(n^{4/7})$, which is only slightly greater than the limiting performance. As a rule of thumb, we recommend to use two or three rounds of feedback in practical communication systems.

Our vision is to further develop the science of non-asymptotic information theory. Towards this goal, the current work characterizes the order terms of the penalty for universality. As future work, we can refine these results to determine the constants involved in the order terms. Consider channel coding as an example. It has been shown that powerful channel coding techniques such as LDPC codes [17] can operate near capacity and have the same $\Theta(\sqrt{n})$ penalty term as the optimal coding scheme; only the constants in the order terms differ. In a similar vein, the constants involved in the order terms for universality will allow us to benchmark the effectiveness of practical universal schemes. Further work can also be done to extend these results for more complex communication systems such as sources and channels with memory, and systems in which the feedback channel is noisy.

Acknowledgments. We thank K. Ramchandran and T. Weissman, who inspired our study of and approach to this problem. We also thank M. A. Khojastepour for informative and inspiring conversations.

References

[1] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, Jul. 1973.

[2] J. Wolfowitz, *Coding theorems of information theory*, 3rd ed. Springer-Verlag, Berlin; New York, 1978.

[3] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley and Sons, New York, 1991.

[4] S. Sarvotham, D. Baron, and R. G. Baraniuk, “Variable-rate universal Slepian-Wolf coding with feedback,” to appear in *38th Asilomar Conf. on Signals, Systems, and Computers,* Pacific Grove, CA, Nov. 2005.

[5] I. Csiszar and J. Korner, “Towards a general theory of source networks,” *IEEE Trans. Inform. Theory*, vol. IT-26, no. 2, pp. 155–165, Mar. 1980.

[6] Y. Oohama and T. S. Han, “Universal coding for the Slepian-Wolf data compression system and the strong converse theorem,” *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1908–1919, Nov. 1994.

[7] A. Kimura and T. Uyematsu, “Weak variable-length Slepian-Wolf coding with linked encoders for mixed source,” *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 183–193, Jan. 2004.

[8] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, “Redundancy rates of Slepian-Wolf coding,” *Proc. 42nd Annual Allerton Conf. on Comm., Control, and Computing*, Oct. 2004.

[9] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, “How quickly can we approach channel capacity?,” *Proc. 38th Asilomar Conf. on Signals, Systems, and Computers,* Pacific Grove, CA, Nov. 2004.

[10] A. Tchamkerten and I. E. Telatar, “On the Universality of Burnashev’s Error Exponent,” *IEEE Trans. Inform. Theory,* pp. 2940-2944, Sep. 2005.

[11] A. Tchamkerten, “Feedback Communication over Unknown Channels,” *Ph.D Thesis,* EPFL, Mar. 2005.

[12] M. Luby, “LT codes,” *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science,* Vancouver, Canada, pp. 271–282, November 2002.

[13] S. Sarvotham, D. Baron, and R. G. Baraniuk, “Penalty for universality in communication systems that rely on joint typicality”, *Technical report*, in preparation, 2005.

[14] M. Johnson, P. Ishwar, V. M. Prabhakaran, D. Schonberg, and K. Ramchandran “On compressing encrypted data,” *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2992–3006, Oct. 2004.

[15] D. Schonberg, S. Draper, and K. Ramchandran “On blind compression of encrypted data approaching the source entropy rate,” *Proc. European Signal Processing Conf.*, Antalya, Turkey, Sept. 2005.

[16] Z. Xiong, A. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *IEEE Signal Processing Mag.*, vol. 21, pp. 80–94, Sept. 2004.

[17] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, “Finite-Length Scaling for Iteratively Decoded LDPC Ensembles,” *Proc. 41st Annual Allerton Conf. on Comm., Control, and Computing*, Oct. 2003.