Variable-length coding with feedback in the non-asymptotic regime

Yury Polyanskiy, H. Vincent Poor, and Sergio Verdú

Abstract—**Without feedback, the backoff from capacity due to non-asymptotic blocklength can be quite substantial for blocklengths and error probabilities of interest in many practical applications. In this paper, novel achievability bounds are used to demonstrate that in the non-asymptotic regime, the maximal achievable rate improves dramatically thanks to variable-length coding with feedback. For example, for the binary symmetric channel with capacity $1/2$ the blocklength required to achieve 90% of the capacity is smaller than 200, compared to at least 3100 for the best fixed-blocklength, non-feedback code. Virtually all the advantages of noiseless feedback are shown to be achievable with decision-feedback only. It is demonstrated that the non-asymptotic behavior of the fundamental limit depends crucially on the particular model chosen for the "end-of-packet" control signal.**

Index Terms—**Shannon theory, channel capacity, feedback, decision feedback, non-asymptotic analysis, memoryless channels, achievability bounds.**

I. Introduction

For a given channel, the fundamental limit of traditional coding with fixed blocklength and no feedback is given by the function $M^*(n, \epsilon)$ which is equal to the maximal cardinality of the code with blocklength n and probability of error ϵ. For several channels, including discrete memoryless channels (DMCs), the additive white Gaussian noise (AWGN) channel and some channels with memory the behavior of this function at fixed ϵ and moderate n is tightly characterized by the expansion [1], [12]

$$\log M^*(n, \epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n), \quad (1)$$

where C and V are the channel capacity and dispersion, resp[1].

In the context of fixed blocklength communication, Shannon showed [2] that noiseless feedback does not increase the capacity of memoryless channels but can increase the zero-error capacity. For a class of symmetric DMCs, Dobrushin demonstrated [11] that the sphere-packing bound holds even in the presence of noiseless feedback. Similarly, it can be shown [15] that for such channels the expansion (1) still holds with feedback as long as blocklength is not allowed to depend on feedback.

Nevertheless, it is known that feedback can be very useful provided that we allow variable-length codes. In his ground-

[1]As usual, $Q(x) = \int_x^\infty \frac{e^{-y^2}}{\sqrt{2\pi}} dy$.

breaking contribution, Burnashev [3] demonstrated that the error exponent improves in this setting and admits a particularly simple expression:

$$E(R) = \frac{C_1}{C}(C - R), \quad (2)$$

for all rates $0 < R < C$, where C is the capacity of the channel and C_1 is the maximal relative entropy between output distributions. Moreover, zero-error capacity may improve from zero to the Shannon capacity (as in the case of the binary erasure channel (BEC)) if variable-length is allowed. Furthermore, since existing communication systems with feedback (such as ARQ) have variable-length, in the analysis of fundamental limits for channels with feedback, it is much more relevant and interesting to allow codes whose length is allowed to depend on the channel behavior.

We mention a few extensions of Burnashev's work [3], [4] relevant to this paper. Yamamoto and Itoh proposed a simple and conceptually important two-phase coding scheme, attaining the optimal error exponent [5]. Using the notion of Goppa's empirical mutual information (EMI) several authors have constructed universal coding schemes attaining rates arbitrarily close to capacity with small probability of error [6], [7], exponentially decaying probability of error [8] and even attaining the optimal Burnashev exponent [9], [10] simultaneously for a collection of channels.

The error exponent analysis focuses on fixed rate, rather than fixed probability as in (1). Another aspect that was not previously addressed in the literature is the following. In practice, control information (such as initiation and termination) is not under the purview of the physical layer. However, the information theory literature typically assumes that all the feed-forward control information is carried through the same noisy channel as the information payload. This is most notably illustrated by Burnashev's model in which the error exponent is, in fact, limited by the reliability with which the termination information is conveyed to the receiver through the DMC while at the same time assuming that the feedback link has infinite reliability to carry not just a termination symbol but the whole sequence of channel outputs. To separate physical-channel issues from upper-layer issues, and avoid mismodelling of control signaling, it is important to realize that Initiation/Termination symbols are in fact carried through layers and protocols whose reliabilities need not be similar to those experienced by the payload. To capture this, we propose a simple modification of the (forward) channel model through the introduction of a "use-once" termination symbol whose

transmission disables further communication.

The organization of the paper is as follows. Section II presents a formal statement of the problem. Section III analyzes the maximal achievable rate with and without a termination symbol. Section IV focuses on zero-error communication. Complete proofs of all the results can be found in [15].

II. STATEMENT OF THE PROBLEM

In this paper we consider the following channel coding scenario. A non-anticipatory channel consists of a pair of input and output alphabets $\mathcal{A}$ and $\mathcal{B}$ together with a collection of conditional probability kernels $\{P_{Y_i|X_1^i Y_1^{i-1}}\}_{i=1}^{\infty}$. Such channel is called (stationary) memoryless if

$$P_{Y_i|X_1^i Y_1^{i-1}} = P_{Y_i|X_i} = P_{Y_1|X_1}, \quad \forall i \geq 1 \qquad (3)$$

and if $\mathcal{A}$ and $\mathcal{B}$ are finite, it is known as a DMC.

Definition 1: An (ℓ, M, ϵ) variable-length feedback (VLF) code, where ℓ is a positive real, M is a positive integer and $0 \leq \epsilon \leq 1$, is defined by:

1) A space $\mathcal{U}$ with[2] $|\mathcal{U}| \leq 3$ and a probability distribution P_U on it, defining a random variable U which is revealed to both transmitter and receiver before the start of transmission; i.e. U acts as common randomness used to initialize the encoder and the decoder before the start of transmission.

2) A sequence of encoders $f_n : \mathcal{U} \times \{1, \ldots, M\} \times \mathcal{B}^{n-1} \to \mathcal{A}$, $n \geq 1$, defining channel inputs

$$X_n = f_n(U, W, Y^{n-1}), \qquad (4)$$

where $W \in \{1, \ldots, M\}$ is the equiprobable message.

3) A sequence of decoders $g_n : \mathcal{U} \times \mathcal{B}^n \to \{1, \ldots, M\}$ providing the best estimate of W at time n.

4) A non-negative integer-valued random variable τ, a stopping time of the filtration $\mathcal{G}_n = \sigma\{U, Y_1, \ldots, Y_n\}$, which satisfies

$$\mathbb{E}[\tau] \leq \ell. \qquad (5)$$

The final decision $\hat{W}$ is computed at the time instant τ:

$$\hat{W} = g_\tau(U, Y^\tau), \qquad (6)$$

and must satisfy

$$\mathbb{P}[\hat{W} \neq W] \leq \epsilon. \qquad (7)$$

The fundamental limit of channel coding with feedback is given by the following quantity:

$$M_f^*(\ell, \epsilon) = \max\{M : \exists(\ell, M, \epsilon)\text{-VLF code}\}. \qquad (8)$$

Those codes that do not require the availability of U, i.e. the ones with $|\mathcal{U}| = 1$, are called *deterministic* codes. Although from a practical viewpoint there is hardly any motivation to allow for non-deterministic codes, they simplify the analysis and expressions just like randomized tests do in hypothesis testing. Also similar to the latter, the difference in performance between the deterministic and non-deterministic codes is negligible for any practically interesting M and ℓ.

In a VLF code the decision about stopping transmission is taken solely upon observation of channel outputs in a causal

manner. This is the setup investigated by Burnashev [3]. Note that since τ is computed at the decoder, it is not necessary to specify the values of $g_n(Y^n)$ for $n \neq \tau$. In this way the decoder is a map $g : \mathcal{B}^\infty \to \{1, \ldots, M\}$ measurable with respect to $\mathcal{G}_\tau$.

Definition 2: An (ℓ, M, ϵ) variable-length feedback code with termination (VLFT) is defined similarly to a VLF code with the exception that condition 4) in the Definition 1 is replaced by

4') A non-negative integer-valued random variable τ, a stopping time of the filtration $\mathcal{G}_n = \sigma\{W, U, Y_1, \ldots, Y_n\}$, which satisfies $\mathbb{E}[\tau] \leq \ell$.

The fundamental limit of channel coding with feedback and termination is given by the following quantity:

$$M_t^*(\ell, \epsilon) = \max\{M : \exists(\ell, M, \epsilon)\text{-VLFT code}\}. \qquad (9)$$

In a VLFT code, "termination" is used to indicate the fact that the practical realization of such a coding scheme requires a method of sending a reliable end-of-packet signal by means other than using the $\mathcal{A} \to \mathcal{B}$ channel (e.g., by cutting off a carrier). As we discussed in the introduction, timing (including termination) is usually handled by a different layer in the protocol. The following are examples of VLFT codes:

1) VLF codes are a special case in which the stopping time τ is determined autonomously by the decoder; due to availability of the feedback, τ is also known to the encoder so that transmission can be cut off at τ.

2) *decision feedback codes* are a special case of VLF codes where

$$f_n(U, W, Y^{n-1}) = f_n(U, W). \qquad (10)$$

Such codes require very limited communcation over feedback: only a single signal to stop the transmission once the decoder is ready to decode.

3) variable-length codes (without feedback), or *VL codes*, defined in [14, Problem 2.1.25] and [13] are required to satisfy two additional requirements: τ is a function of (W, U) and (10) holds. The fundamental limit and the ϵ-capacity of variable-length codes are given by

$$M_v^*(\ell, \epsilon) = \max\{M : \exists(\ell, M, \epsilon)\text{-VL code}\}, \qquad (11)$$
$$[\![C_\epsilon]\!] = \lim_{\ell \to \infty} \frac{1}{\ell} \log M_v^*(\ell, \epsilon). \qquad (12)$$

4) fixed-to-variable codes, or *FV codes*, defined in [13] are also required to satisfy (10), while the stopping time is[3]

$$\tau = \inf\{n \geq 1 : g_n(U, Y^n) = W\}, \qquad (13)$$

and therefore, such codes are zero-error VLFT codes. Of course, not all zero-error VLFT codes are FV codes, since in general condition (10) does not necessarily hold.

5) automatic repeat request (*ARQ*) codes analyzed in [1, Section IV.E] are yet a more restricted class of deterministic FV codes, where a single fixed-blocklength,

---

[2]This bound on the cardinality does not incur any loss of optimality as shown in [15, Appendix A].

[3]As explained in [13], this model encompasses fountain codes in which the decoder can get a highly reliable estimate of τ autonomously without the need for a termination symbol.

non-feedback code is used repeatedly until the decoder produces a correct estimate.

The main goal of this paper is to analyze the behavior of $\log M_f^*(\ell, \epsilon)$ and $\log M_t^*(\ell, \epsilon)$ and compare them with the behavior of the fundamental limit without feedback, $\log M^*(n, \epsilon)$. Regarding the behavior of $\log M_f^*(\ell, \epsilon)$ Burnashev's result (2) can be restated as

$$\log M_f^*(\ell, \exp\{-E\ell\}) = \ell C \left(1 - \frac{E}{C_1}\right) + o(\ell), \quad (14)$$

for any $0 < E < C_1$. Although (14) does not imply any statement about the expansion of $\log M_f^*(\ell, \epsilon)$ for a fixed ϵ, it still demonstrates that in the regime of very small probability of error, the parameter C_1 emerges as an important quantity.

III. MAIN RESULTS

The first result shows that under variable-length coding allowing a non-vanishing error probability ϵ boosts the ϵ-capacity by a factor of $\frac{1}{1-\epsilon}$ even in the absence of feedback.

Theorem 1: For any non-anticipatory channel with capacity C that satisfies the strong converse for fixed-blocklength codes (without feedback), the ϵ-capacity under variable-length coding without feedback, cf. (12), is

$$[\![C_\epsilon]\!] = \frac{C}{1-\epsilon}, \epsilon \in (0,1). \quad (15)$$

In general, it is known [13, Theorem 16] that the VL capacity, $[\![C]\!] = \lim_{\epsilon \to 0} [\![C_\epsilon]\!]$, is equal to the conventional fixed-blocklength capacity without feedback, C, for any non-anticipatory channel (not necessarily satisfying the strong converse). On the other hand, the capacity of FV codes for state-dependent non-ergodic channels can be larger than C [13].

Our main result is the following:

Theorem 2: For an arbitrary DMC with capacity C we have for any $0 < \epsilon < 1$

$$\log M_f^*(\ell, \epsilon) = \frac{\ell C}{1-\epsilon} + O(\log \ell), \quad (16)$$

$$\log M_t^*(\ell, \epsilon) = \frac{\ell C}{1-\epsilon} + O(\log \ell). \quad (17)$$

More precisely, we have

$$\frac{\ell C}{1-\epsilon} - \log \ell + O(1) \leq \log M_f^*(\ell, \epsilon) \leq \frac{\ell C}{1-\epsilon} + O(1) \quad (18)$$

$$\log M_f^*(\ell, \epsilon) \leq \log M_t^*(\ell, \epsilon) \leq \frac{\ell C + \log \ell}{1-\epsilon} + O(1). \quad (19)$$

A consequence of Theorem 2 is that for DMCs, feedback (even in the setup of VLFT codes) does not increase the ϵ-capacity, namely,

$$\lim_{\ell \to \infty} \frac{1}{\ell} \log M_t^*(\ell, \epsilon) = [\![C_\epsilon]\!],$$

where $[\![C_\epsilon]\!]$ is defined in (12) and given by Theorem 1.

However, while in the absence of feedback and within the paradigm of fixed-length coding, the backoff from ϵ-capacity (equal to capacity for DMCs) is governed by the $\frac{1}{\sqrt{n}}$ term (1),

variable-length coding with feedback completely eliminates that penalty. Thus, the capacity is attainable at a much smaller (average) blocklength. Furthermore, the achievability (lower) bound in (18) is obtained via decision feedback codes (10) that use feedback only to let the encoder know that the decoder has made its final decision. As (18) demonstrates, such a sparing use of feedback does not lead to any significant loss in rate even non-asymptotically. Naturally, such a strategy is eminently practical in many applications, unlike those strategies that require full, noiseless, instantaneous feedback. In the particular case of the BSC, a lower bound (18) with a weaker $\log n$ term has been claimed in [8].

Theorem 3: Fix a real number $\gamma > 0$, a channel $\{P_{Y_i|X_1^i Y_1^{i-1}}\}_{i=1}^\infty$ and an arbitrary process $X = (X_1, X_2, \ldots, X_n, \ldots)$ taking values in $\mathcal{A}$. Define a probability space with finite-dimensional distributions given by

$$P_{X^n Y^n \bar{X}^n}(a^n, b^n, c^n) = \quad (20)$$

$$P_{X^n}(a^n) P_{\bar{X}^n}(c^n) \prod_{j=1}^n P_{Y_j|X_1^j Y_1^{j-1}}(a_j|b^j, a^{j-1}), (21)$$

i.e. X and $\bar{X}$ are independent copies of the same process and Y is the output of the channel when X is its input. For the joint distribution (21) define a sequence of information density functions $\mathcal{A}^n \times \mathcal{B}^n \to \bar{\mathbb{R}}$

$$i(a^n; b^n) = \log \frac{dP_{Y^n|X^n}(b^n|a^n)}{dP_{Y^n}(b^n)}, \quad (22)$$

and a pair of hitting times:

$$\tau = \inf\{n \geq 0 : i(X^n; Y^n) \geq \gamma\}, \quad (23)$$
$$\bar{\tau} = \inf\{n \geq 0 : i(\bar{X}^n; Y^n) \geq \gamma\}. \quad (24)$$

Then for any M there exists an (ℓ, M, ϵ) VLF code with

$$\ell \leq \mathbb{E}[\tau] \quad (25)$$
$$\epsilon \leq (M-1)\mathbb{P}[\bar{\tau} \leq \tau]. \quad (26)$$

Furthermore, for any M there exists a deterministic (ℓ', M, ϵ) VLF code with ϵ satisfying (26) and

$$\ell' \leq \text{esssup} \, \mathbb{E}[\tau|X]. \quad (27)$$

Worsening the bound to (27) is advantageous, e.g., for symmetric channels, since we have $\mathbb{E}[\tau|X] = \mathbb{E}[\tau]$ and thus the second part of Theorem 3 guarantees the existence of a deterministic code without any sacrifice in performance. Theorem 3 is a natural extension of the DT bound [1, Theorem 17], since (26) corresponds to the second term in [1, (70)], whereas the first term in [1, (70)] is missing because the information density corresponding to the true message eventually crosses any level γ with probability one. Interestingly, pairing a fixed stopping rule with a random-coding argument has been already discovered from a different perspective: in the context of universal variable-length codes [6]–[10], stopping rules based on a sequentially computed EMI were shown to be optimal in several different asymptotic senses. Although invaluable for universal coding, EMI-based decoders are hard
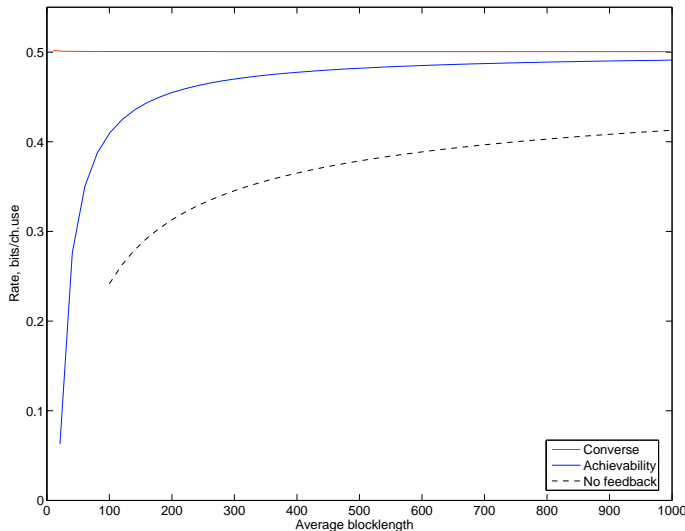
Fig. 1. Comparison of upper and lower bounds for the BSC(0.11) with variable-length and feedback; probability of error $\epsilon = 10^{-3}$.

to evaluate non-asymptotically as their analysis relies on inherently asymptotic methods, such as type-counting, cf. [10]. While the codes with encoders utilizing full noiseless feedback can achieve the Burnashev exponent (2), it was noted in [8] and [10] that the lower error exponent

$$E_1(R) = C - R \qquad (28)$$

is achievable at all rates $R < C$ with decision feedback codes (10). This property follows from Theorem 3 (see [15]).

Theorem 4: Consider an arbitrary DMC with capacity C. Then any (ℓ, M, ϵ) VLF code with $0 \leq \epsilon < 1$ satisfies[4]

$$\log M \leq \frac{C\ell + h(\epsilon)}{1 - \epsilon}, \qquad (29)$$

whereas each (ℓ, M, ϵ) VLFT code with $0 \leq \epsilon < 1$ satisfies

$$\log M \leq \frac{C\ell + \log(\ell + 1) + h(\epsilon) + \log e}{1 - \epsilon}, \qquad (30)$$

where $h(x) = -x \log x - (1 - x) \log(1 - x)$ is the binary entropy function.

Not only do Theorems 3 and 4 lead to a proof of Theorem 2, but also provide tight non-asymptotic bounds on the communication rate. A numerical comparison for the BSC with crossover probability $\delta = 0.11$ and $\epsilon = 10^{-3}$ is given in Fig. 1, where the upper bound is (29) and the lower bound is Theorem 3 (evaluated with various γ depending on the average blocklength). Note that for $BSC(\delta)$ the $i(X^n; Y^n)$ becomes a random walk taking steps $\log 2\delta$ and $\log(2 - 2\delta)$ with probabilities δ and $1 - \delta$, i.e.,

$$i(X^n; Y^n) = n \log(2 - 2\delta) + \log \frac{\delta}{1 - \delta} \sum_{k=1}^{n} Z_k, \qquad (31)$$

where Z_k are independent Bernoulli $\mathbb{P}[Z_k = 1] = 1 - \mathbb{P}[Z_k = 0] = \delta$. After simplifications (26) becomes:

$$\epsilon \leq (M-1)\mathbb{E}[f(\tau)], \qquad (32)$$

where

$$f(n) \triangleq \mathbb{E}[1\{\tau \leq n\} \exp\{-i(X^\tau; Y^\tau)\}]. \qquad (33)$$

[4]The inequality (29) is contained essentially in Lemmas 1 and 2 of [3].

The dashed line in Fig. 1 is the approximate fundamental limit for fixed blocklength codes without feedback given by the equation (1) with $O(\log n)$ substituted by $\frac{1}{2} \log n$; see [1, Theorem 53].

IV. ZERO-ERROR COMMUNICATION

The general achievability bound, Theorem 3, applies only to $\epsilon > 0$. What can be said about $\epsilon = 0$? Burnashev [3] showed that whenever $C_1 = \infty$, as $\ell \to \infty$ we have for some $a > 0$

$$\log M_f^*(\ell, 0) \geq C\ell - a\sqrt{\ell \log \ell} + O(\log \ell). \qquad (34)$$

For this reason, for such channels zero-error VLF capacity is equal to the conventional capacity. However, the penalty bound $\sqrt{\ell \log \ell}$ is rather loose, as the following result demonstrates.

Theorem 5: For a $BEC(\delta)$ with capacity C we have

$$\log_2 M_f^*(\ell, 0) = \ell C + O(1). \qquad (35)$$

Regarding any channel with $C_1 < \infty$ (e.g. the BSC), the following negative result holds:

Theorem 6: For any DMC with $C_1 < \infty$ we have

$$\log M_f^*(\ell, 0) = 0 \qquad (36)$$

for all $\ell \geq 0$.

The shortcoming of VLF coding found in Theorem 6 is overcome in the paradigm of VLFT coding. Our main tool is the following achievability bound.

Theorem 7: Fix an arbitrary channel $\{P_{Y_i|X_1^i Y_1^{i-1}}\}_{i=1}^\infty$ and a process $X = (X_1, X_2, \ldots, X_n, \ldots)$ with values in $\mathcal{A}$. Then for every positive integer M there exists an $(\ell, M, 0)$ VLFT code with

$$\ell \leq \sum_{n=0}^{\infty} \mathbb{E}\left[\min\left\{1, (M-1)\mathbb{P}[i(X^n; Y^n) \leq i(\bar{X}^n; Y^n)|X^n Y^n]\right\}\right], \qquad (37)$$

where X^n, $\bar{X}^n$, Y^n and $i(\cdot; \cdot)$ are defined in (21) and (22). Moreover, this is an FV code which is deterministic and uses feedback only to compute the stopping time, i.e. (10) holds.

Theorem 8: For an arbitrary DMC we have

$$\log M_t^*(\ell, 0) = \ell C + O(\log \ell). \qquad (38)$$

More specifically we have

$$\log M_t^*(\ell, 0) \leq \ell C + \log \ell + O(1), \qquad (39)$$

$$\log M_t^*(\ell, 0) \geq \ell C + O(1). \qquad (40)$$

Furthermore, the encoder achieving (40) uses feedback to calculate the stopping time only, i.e. it is an FV code.

Theorem 8 suggests that VLFT codes may achieve capacity even at very short blocklengths. To illustrate this numerically we first notice that Theorem 7 particularized to the BSC with i.i.d. input process X and an equiprobable marginal distribution yields the following result[5].

Corollary 9: For the BSC with crossover probability δ and for every positive integer M there exists an $(\ell, M, 0)$ VLFT code satisfying

$$\ell \leq \sum_{n=0}^{\infty} \sum_{t=0}^{n} \binom{n}{t} \delta^t (1-\delta)^{n-t} \min\left\{1, M \sum_{k=0}^{t} \binom{n}{k} 2^{-n}\right\}. \qquad (41)$$

[5]This expression is to be compared with the (almost) optimal non-feedback achievability bound for the BSC, [1, Theorem 34].
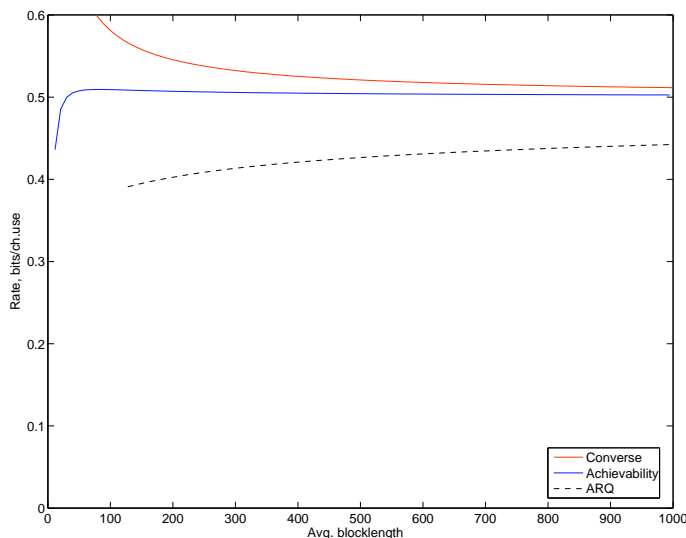
Fig. 2. Zero-error communication over the BSC(0.11) with a termination symbol. The lower bound is (41); the upper-bound is (30).

A comparison of (41) and the upper bound (30) is given in Fig. 2. We see that despite the requirement of zero probability of error, VLFT codes attain the capacity of the BSC at blocklengths as short as 30. Additionally, we have depicted the (approximate) performance of the best non-feedback code paired with the simple ARQ strategy, see [1, Section IV.E]. Note that the ARQ strategy indeed gives a valid zero-error VLFT code. The comparison on Fig. 2 suggests that even having access to the best possible block codes the ARQ is considerably suboptimal. It is interesting to note in this regard, that a Yamamoto-Itoh [5] strategy also pairs the best block code with a noisy version of ARQ (therefore, it is a VLF achievability bound). Consequently, we expect a similar gap in performance.

V. DISCUSSION

We have demonstrated that by allowing variable length, even a modicum of feedback is enough to considerably speed up convergence to capacity. For example, we constructed a feedback code that achieves 90% of the capacity of the BSC at blocklength 200; see Fig. 1. In contrast, to obtain the same performance without feedback requires a blocklength of at least 3100. Practically, this opens the possibility of utilizing the full capacity of the link without the complexity required to implement coding of very long packets. Indeed, a major ingredient of the achievability bounds in this paper, the idea of terminating early on favorable noise realizations, can be used to show that any point on the achievability curve of Fig. 1 is realized by pairing some linear block code with the stopping rule (23). In other words, known linear codes can be decoded with significantly less (average) delay if used in the variable-length setting.

Theoretically, the benefit of feedback is manifested by the absence of the $\sqrt{\ell}$ term in the expansions (16) and (17), whereas this term is crucial to determine the non-asymptotic performance without feedback. Intuitively, without feedback the main effect governing the $\sqrt{n}$ behavior was the stochastic

variation of information density around its mean, which is tightly characterized by the central limit theorem. In the variable-length setup with feedback the main idea is that of Wald-like stopping once the information density of some message is large enough. Therefore, there is virtually no stochastic variation and this explains the absence of any references to the central limit theorem and the fact that dispersion is zero.

We have also analyzed a modification of the coding problem by introducing a termination symbol (VLFT codes), which is practically motivated in many situations in which control signals are sent over a highly reliable upper layer. Not only this leads to the possibility of communicating with zero-error, but also dramatically improves the transient behavior, see Fig. 2, which is analytically expressed by the absence of not only the $\sqrt{\ell}$ term but also of the $\log \ell$ term in the bound (40). Furthermore, in Fig. 2 we see that fountain codes can achieve 90% of the capacity of the BSC at average blocklength < 20 and with zero probability of error. Practically, of course, "zero-error" should be understood as the reliability being essentially the probability with which the termination symbol is correctly detected.

REFERENCES

[1] Y. Polyanskiy, H. V. Poor and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307-2359, May 2010.

[2] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, Vol. 2, No. 3, pp. 8-19, Sept. 1956.

[3] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random transmission time," *Problems of Information Transmission,* vol.12, no.4, pp. 10-30, 1976.

[4] M. V. Burnashev, "Sequential discrimination of hypotheses with control of observations," *Math. USSR, Izvestia,* vol. 15, no. 3, pp. 419-440, 1980.

[5] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. 25, no. 6, pp. 729-733, Nov. 1979.

[6] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel-Aviv Univ., Tel-Aviv, Israel, 2003.

[7] S. C. Draper, B. J. Frey, and F. R. Kschischang, "Efficient variable length channel coding for unknown DMCs," *Proc. 2004 IEEE Int. Symp. Information Theory (ISIT)*, Chicago, IL, USA, June 2004.

[8] A. Tchamkerten and E. Telatar, "A feedback strategy for binary symmetric channels," *Proc. 2002 IEEE Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, July 2002.

[9] A. Tchamkerten and E. Telatar, "Optimal feedback schemes over unknown channels," *Proc. 2004 IEEE Int. Symp. Information Theory (ISIT)*, Chicago, IL, USA, June 2004.

[10] A. Tchamkerten and E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2126-2145, May 2006.

[11] R. L. Dobrushin, "Asymptotic bounds on error probability for transmission over DMC with symmetric transition probabilities," *Theory of Probab. Applicat.,* vol. 7, pp. 283-311, 1962.

[12] Y. Polyanskiy, H. V. Poor and S. Verdú, "Dispersion of the Gilbert-Elliot channel," *Proc. IEEE Int. Symp. Information Theory (ISIT),* Seoul, Korea, July 2009; *IEEE Trans. Inform. Theory, 2010.*

[13] S. Verdú and S. Shamai, "Variable-rate channel capacity," *IEEE Trans. Inform. Theory*, Jun. 2010

[14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems,* Academic, New York, 1981.

[15] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, Apr. 2010, submitted for publication.