Variable-Length Coding with Feedback: Finite-Length Codewords and Periodic Decoding

Tsung-Yi Chen, Adam R. Williamson and Richard D. Wesel

Department of Electrical Engineering

UCLA, Los Angeles, California 90095

Email: tsungyi@seas.ucla.edu; adamroyce@ucla.edu; wesel@ee.ucla.edu

Abstract—Theoretical analysis has long indicated that feedback improves the error exponent but not the capacity of single-user memoryless channels. Recently Polyanskiy et al. studied the benefit of variable-length feedback with termination (VLFT) codes in the non-asymptotic regime. In that work, achievability is based on an infinite-length random code and decoding is attempted at every symbol. The coding rate backoff from capacity due to channel dispersion is greatly reduced with feedback, allowing capacity to be approached with surprisingly small expected latency.

This paper is concerned with VLFT codes based on *finite-length* codes and decoding attempts only at *certain specified decoding times*. Note that with an underlying finite-length code, the transmitter may have to repeat code symbols. The penalties of using a finite block-length N and a sequence of specified decoding times are studied. This paper shows that properly scaling N with the expected latency can achieve the same performance up to second order terms as with $N = \infty$. The penalty introduced by limiting the decoding is a constant term and hence the performance approaches capacity as expected latency increases as long as the interval between periodic decoding times grows sub-linearly with the expected latency.

I. INTRODUCTION

While feedback cannot increase the capacity of a memoryless channel, it can significantly reduce the complexity of encoding and decoding at rates below capacity. The error exponent results of [1]–[4] suggest that feedback can be used to reduce the average block-length (or expected latency) required to approach capacity. As a practical demonstration, [5] showed that using an incremental redundancy (IR) scheme with feedback allows short convolutional codes to deliver bit error rate performance comparable to a long-block-length turbo code, but with lower latency. The demonstration of [5] qualitatively agrees with the error exponent analysis in [1]–[4].

Because of its asymptotic perspective, the error exponent theory does not provide an accurate prediction of short-block-length performance. For example, Yamamoto and Itoh [6] showed that the optimal Burnashev error exponent [7] is achievable by a two-phase ARQ coding scheme. However, at short block-lengths (i.e. for a small average number of channel uses) a considerable performance gap exists between ARQ and a well-designed IR scheme. Polyanskiy et al. [8] analyzed the benefit of feedback in the non-asymptotic regime and provide quantitative characterizations for short expected latency. They show that capacity can be closely approached in hundreds of

symbols rather than thousands using variable-length feedback codes with termination (VLFT codes), a form of IR.

The analysis of VLFT in [8] assumes an underlying codebook with infinite-length codewords to support incremental redundancy transmissions. In practice, a codebook with finite-length codewords may force the transmitter to repeat code symbols to support later incremental transmissions.

In [8], decoding is attempted at every symbol so that the communication may end after any given channel use. In practice, it may also be possible only to attempt decoding (and thus conclude communication) after channel uses that come at the end of a group of symbols because of packetization, decoding delays, and round-trip propagation time.

With these practical issues in mind, this paper studies the penalties that occur when the codebook is limited to finite-length codewords and/or decoding (and therefore termination) is only possible at periodic intervals.

II. PREVIOUS WORK AND MAIN RESULTS

A. Previous Work

We will consider discrete memoryless channels (DMC) throughout the paper and use the following notation: $x^n = (x_1, x_2, \ldots, x_n)$ denotes an n-dimensional vector, x_j the jth element of x^n, and x_i^j the ith to jth elements of x^n. We denote random variables by capitalized letters unless otherwise stated. The input and output alphabets are $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let the input and output product spaces be $\mathsf{X} = \mathcal{X}^n, \mathsf{Y} = \mathcal{Y}^n$ respectively. A channel is characterized by a conditional distribution $P_{\mathsf{Y}|\mathsf{X}} = \prod_{i=1}^n P_{Y_i|X_i}$ where the equality holds because the channel is memoryless. For codes that make use of a noiseless feedback link, we consider causal channels $\{P_{Y_i|X^iY^{i-1}}\}_{i=1}^\infty$ and additionally focus on causal memoryless channels $\{P_{Y_i|X_i}\}_{i=1}^\infty$.

We are interested in zero-error communication with feedback in this paper and will therefore focus on the paradigm of VLFT coding. In order to be self-contained, we state the definition of VLFT codes in [8]:

Definition 1: An (ℓ, M, ϵ) variable-length feedback code with termination (VLFT code) is defined as:

1) A common random variable (r.v.) $U \in \mathcal{U}$ with a probability distribution P_U revealed to both transmitter and receiver before the start of transmission.

2) A sequence of encoders $f_n : \mathcal{U} \times \mathcal{W} \times \mathcal{Y}^{n-1} \to \mathcal{X}$ that defines the channel inputs $X_n = f_n(U, W, Y^{n-1})$. Here W is the message r.v. uniform in $\{1, \ldots, M\}$.

3) A sequence of decoders $g_n : \mathcal{U} \times \mathcal{Y}^n \to \mathcal{W}$ providing the estimate of W at time n.
4) A stopping time $\tau \in \mathbb{N}$ w.r.t. the filtration $\mathcal{F}_n = \sigma\{U, Y^n, W\}$ such that:

$$\mathbb{E}[\tau] \leq \ell. \tag{1}$$

5) The final decision $\hat{W} = g_\tau(U, Y^\tau)$ must satisfy:

$$\mathbb{P}[\hat{W} \neq W] \leq \epsilon. \tag{2}$$

As observed in [8], the setup of VLFT is equivalent to augmenting each channel with a special use-once input symbol, the termination symbol, that has infinite reliability. This assumption captures the fact that many practical systems communicate control signals in the upper protocol layers and the termination symbol effectively separates the control issue from the physical channel. The benefit of the infinitely reliable control signal can cause the VLFT achievable rate to be larger than that of the original feedback channel capacity because what would have been a decoding error without feedback becomes a codeword "erasure" under VLFT.

The class of fixed-to-variable codes [9], or FV codes, is a special class of VLFT codes that satisfies the following conditions:

$$f_n(U, W, Y^{n-1}) = f_n(U, W) \tag{3}$$
$$\tau = \inf\{n \geq 1 : g_n(U, Y^n) = W\}. \tag{4}$$

Such codes are zero-error VLFT codes and only use feedback to stop the transmission. Fountain codes and families of rate-compatible codes used with an IR scheme are examples of such codes. This class of codes is widely used in practical systems and will be the main focus of this paper.

Let the finite dimensional distribution of $(X^n, \bar{X}^n, Y^n)$ be:

$$P_{X^n Y^n \bar{X}^n}(x^n, y^n, \bar{x}^n) \tag{5}$$
$$= P_{X^n}(x^n) P_{X^n}(\bar{x}^n) \prod_{j=1}^{n} P_{Y_j | X^j Y^{j-1}}(y_j | x^j, y^{j-1}), \tag{6}$$

i.e. the distribution of $\bar{X}^n$ is identical to X^n but independent of Y^n. The information density $i(x^n; y^n)$ is defined as

$$i(x^n; y^n) = \log \frac{dP_{X^n Y^n}(x^n, y^n)}{d(P_{X^n}(x^n) \times P_{Y^n}(y^n))} \tag{7}$$
$$= \log \frac{dP_{Y^n | X^n}(y^n | x^n)}{dP_{Y^n}(y^n)}. \tag{8}$$

The following is the achievability result based on random coding union bound (RCU) in [8]:

Theorem 1 ([8], Thm. 10): Fixing $M > 0$, there exists an $(\ell, M, 0)$ VLFT code with

$$\ell \leq \sum_{n=0}^{\infty} \xi_n \tag{9}$$

where ξ_n is the following expectation:

$$\mathbb{E}\left[\min\left\{1, (M-1)\Pr[i(X^n; Y^n) \leq i(\bar{X}^n; Y^n) | X^n Y^n]\right\}\right]. \tag{10}$$

We take $i(X^0; Y^0) = 0$ and hence $\xi_0 = 1$. Additionally, from the proof of [8, Thm. 11], we have:

$$\xi_n \leq \mathbb{E}\left[\exp\left\{-[i(X^n; Y^n) - \log\gamma]^+\right\}\right], \tag{11}$$

where $\gamma > 0$ is chosen such that (9) is minimized.

B. Problem Statement and Main Results

Following the VLFT framework of [8], this paper studies the following problems:

(i) Finite-length codeword penalty for VLFT (FV) codes: The random coding approach in [8] generates random codebooks in an infinite product space (i.e. with codewords of infinite length). We study the performance penalty incurred by using random codebooks with a finite block-length, possibly repeating codeword symbols.

(ii) The penalty associated with limitations on decoding times: We study the performance penalty incurred when decoding is only allowed after every I symbols are received, i.e. periodic decoding times.

Note that [8, Sec. V] studied the case of an excess delay constraint where further transmission is not allowed after a given length N. In contrast, we have no such constraint and repetition is allowed to drive the error probability to zero.

For the rest of the paper we only consider channels with bounded information density $i(X; Y)$. Define a VLFT code with finite block-length N and uniform increment I as follows:

Definition 2: An $(\ell, M, N, I, \epsilon)$ VLFT code modifies 2) and 4) in definition 1:

2') A sequence of encoders

$$f_{n+kN} : \mathcal{U} \times \mathcal{W} \times \mathcal{Y}_{kN+1}^{n+kN-1} \mapsto \mathcal{X}$$

that satisfies $f_n = f_{n+N}$.

4') A stopping time $\tau \in \{kI : k \in \mathbb{N}\}$ w.r.t. the filtration $\mathcal{F}_n$ s.t. $\mathbb{E}[\tau] \leq \ell$.

Define the fundamental limit of an $(\ell, M, N, I, \epsilon)$ VLFT code as follows:

Definition 3: Let $M_t^*(\ell, N, I, \epsilon)$ be the maximum integer M such that there exist an $(\ell, M, N, I, \epsilon)$ VLFT code. For zero-error codes where $\epsilon = 0$ we denote the maximum M as $M_t^*(\ell, N, I)$ and for zero-error codes with $I = 1$ (i.e. decoding attempts after every received symbol) we denote the maximum M as $M_t^*(\ell, N)$.

All of the results that follow assume an arbitrary but fixed channel $\{P_{Y_j | X_j}\}_{j=1}^{N}$ and a process $\{X_j\}_{j=1}^{N}$ taking values in $\mathcal{X}$ where N could be set as infinity. Our main asymptotic result is the following expansion for a stationary DMC with capacity C:

$$\log M_t^*(\ell, N, I) \geq \ell C - O(I) \tag{12}$$

by choosing $N = \ell + \Omega(\log \ell)$. Specifically, if we choose $N > \ell + \frac{\log(\ell+1) + \log e}{C}$ and have decoding attempts separated by an increment I, then the expansion is the same as the case with $N = \infty$ and the constant term depends on the choice of the increment I. This is a corollary from the theorems in Section III-C and is stated at the end of Section III-C.

For practical applications that apply feedback to obtain reduced latency, the non-asymptotic behavior is critical. Numerical results on a binary symmetric channel (BSC) show that properly selected values of N and I yield excellent expected throughput with expected latency on the order of 200 symbols.

The rest of the paper is organized as follows: Section III-A investigates the penalty incurred by using VLFT codes based on finite block-lengths. Section III-B studies the penalty incurred by limiting decoding attempts, and Section III-C studies the penalty when both limitations are applied. Section IV gives numerical results for a BSC. Section V concludes the paper.

III. FINITE BLOCK-LENGTHS AND LIMITED DECODING

In [8], attention was focused on $(\ell, M, N, I, \epsilon)$ VLFT codes with $N = \infty$ and $I = 1$. This section studies the penalties associated with using finite N and $I \geq 1$. We focus on the $\epsilon = 0$ case. The random coding framework of [8] is retained. We focus on achievability results under these constrained scenarios using proofs based on random FV codes. The general converse established in [8] still applies since these additional constraints can only further limit performance.

A. The Finite-Block-Length Limitation

This subsection investigates $(\ell, M, N, I, \epsilon)$ VLFT codes with finite N but retains decoding at every symbol ($I = 1$). FV codes (as described in Section II-A) are employed so that encoding does not depend on the feedback. However, feedback does indicate when to terminate transmission.

Let ζ_j be the marginal error event at the jth transmission. Consider a code $\mathcal{C}_N$ with finite block-length N where each element is a length-N $\mathcal{X}$-valued string. Achievability results for an $(\ell, M, \epsilon, N, 1)$ "truncated" VLFT code follow from a random coding argument. In particular we have the following:

Theorem 2: For any $M > 0$ there exists an $(\ell, M, N, 1, \epsilon)$ truncated VLFT code with

$$\ell \leq \sum_{n=0}^{N-1} \xi_n \tag{13}$$

$$\epsilon \leq \xi_N. \tag{14}$$

where ξ_n is the same as (10). The proof is in the appendix.

Achievability results for $\epsilon = 0$ can be obtained using an $(\ell, M, N, 1, \epsilon)$ "repeated" VLFT code, which modifies the encoder and decoder pairs with an ARQ-type repetition. When the block-length-N codeword is exhausted without successful decoding, the transmission process starts from scratch discarding the previous received symbols. Using the original N symbols, for example through Chase code combining, would be beneficial, but this is not necessary for our achievability result. Specifically, we have the following result for a zero-error repeated VLFT code with a finite block-length N:

Theorem 3: For every $M > 0$ there exists an $(\ell, M, N, 1, 0)$ VLFT code such that

$$\ell \leq \frac{1}{(1 - \xi_N)} \sum_{n=0}^{N-1} \xi_n \tag{15}$$

where ξ_n is the same as (10). The proof is in the appendix.

Note that this is an FV code based on a finite-length codebook rather than an infinite one. The penalty of using a codebook with finite length is made clear in the following theorem and its corollary:

Theorem 4: For an $(\ell, M, N, 1, 0)$ repeated VLFT code with $N = O(\log M)$, we have the following expansion:

$$\ell \leq \frac{\log M}{C} + O(1). \tag{16}$$

Let $C_\Delta = C - \Delta, \Delta > 0$ and $N = \log M / C_\Delta$. The correction term is upper bounded as follows:

$$O(1) \leq \frac{b_2 \log M}{C(M^{b_3/C_\Delta} - b_2)} + \frac{b_0 \log M}{C_\Delta M^{b_1 \Delta / C_\Delta}} + a \tag{17}$$

where a depends on the mean and uniform bound of $i(X; Y)$, and b_j's are constants related to Δ and M. This choice of N has residual terms decaying with M very slowly. However, our numerical results indicate that this decay is fast enough for excellent performance in the short-block-length regime. We provide the proof of Thm. 5 in [10] but omit it from this paper due to limited space.

An expansion of $\log M_t^*(\ell, N)$ requires N growing with ℓ. The components of the correction term in Thm. 4, however, depend on both N (as $\log M / C_\Delta$) and M. Indeed for a fixed ℓ, all M satisfying (16) and (17) are achievable. The argument we make below is that for any fixed constant $c_0 > 0$, there is an ℓ_0 that depends logarithmically on c_0^{-1} such that the expansion $\log M_t^* \geq C\ell - c_0$ is true for all $\ell \geq \ell_0$. We first invoke the converse for an $(\ell, M, \infty, 1, 0)$ VFLT code:

Theorem 5 (Thm. 11, [8]): We have the following converse for an $(\ell, M, \infty, 1, 0)$ VLFT code:

$$\log M_t^* \leq \ell C + \log(\ell + 1) + \log e. \tag{18}$$

Combining Thms. 4 and 5 we have the following:

Corollary 1: For an $(\ell, M, N, 1, 0)$ repeated VLFT code with $N = (1 + \delta)\ell$ with a proper choice of $\delta > 0$ we have:

$$\log M_t^* \geq \ell C - O(1). \tag{19}$$

We provide the proof in [10] but omit it from this paper due to limited space.

To conclude this discussion of the penalty associated with finite block-length, we comment that N only needs to be scaled properly, i.e. $(1 + \delta)\ell$ for δ decreasing with ℓ, to obtain the infinite-block-length expansion of $M_t^*(\ell, \infty)$ provided in [8]. Thus, the restriction to a finite block-length N does not restrict the asymptotic performance if N is selected properly with respect to ℓ. However, the constant penalty is indeed different for infinite and finite N, which might not be negligible in the short-block-length regime. Still, our numerical results in Section IV indicate that relatively small values of N can yield good results for short block-lengths.

B. Limited, Regularly-Spaced, Decoding Attempts

This subsection investigates $(\ell, M, N, I, \epsilon)$ VLFT codes with $N = \infty$ but decoding attempted only at specified,

regularly-spaced, symbols $(I > 1)$. The first decoding time occurs after n_1 symbols (which could be larger than I) so that the decoding attempts are made at the times $n_j = n_1 + (j-1)I$. The relevant information density process $i(X^{n_j}; Y^{n_j})$ is on the subsequence $n_j = n_1 + (j-1)I$. The main result here is that the constant penalty now scales linearly with I:

Theorem 6: For an $(\ell, M, N, I, 0)$ VLFT code with uniform increments I and $N = \infty$ we have the following expansion:

$$\log M_t^*(\ell, \infty, I) \geq \ell C - O(I). \tag{20}$$

Considering an I-fold product channel, invoking Cor. 1 yields a simple proof. We provide a full proof that makes clear the relation on the penalty terms in [10] but omit it here due to limited space. In view of the theorem, the increment I can grow slowly, e.g. $I = O(\log \ell)$ and can still permit an expected rate that approaches C without the dispersion penalty. In the non-asymptotic regime, however, the penalty might not be negligible. Our numerical results in Section IV indicate $I = \lceil \log_2 \log_2 M \rceil$ yields good results for short block-lengths.

C. Finite Block-Length and Limited Decoding Attempts

This subsection investigates $(\ell, M, N, I, 0)$ (repeated) VLFT codes with *both* finite N and $I > 1$. When these two limitations are combined, a key parameter is m, the number of decoding attempts before the transmission process must start from scratch if successful decoding has not yet been achieved. The main result follows from combining the results of Sections III-A and III-B. Once n_1, N and I are specified, the value of m is implied. Specifically, we have the following theorem:

Theorem 7: For an $(\ell, M, N, I, 0)$ VLFT code with $N = \Omega(\log M)$, we have the following for a stationary DMC with capacity C:

$$\ell \leq (1 + \mathbb{P}[\zeta_N])^{-1} \frac{\log M}{C} + \mathbb{P}[\tau_0 \geq m] + O(I) \tag{21}$$

$$\leq \frac{\log M}{C} + O(I), \tag{22}$$

where τ_0 is the stopping time in terms of the number of decoding attempts up to and including the first success.
The proof is omitted due to space limit. Summing up all the theorems above, we immediately have the following corollary:

Corollary 2: Choosing $N = \ell + \Omega(\log \ell)$ for a stationary DMC with capacity C, we have:

$$\log M_t^*(\ell, N, I) \geq \ell C - O(I). \tag{23}$$

Proof: For an $(\ell, M, N, I, 0)$ VLFT code pick N as

$$N = (1 + \delta)\ell, \text{ where } \delta > \frac{\log(\ell + 1) + \log e}{\ell C}. \tag{24}$$

The result follows by a similar argument as for Cor. 1. The restriction on the initial block-length n_1 only makes a constant difference. ∎

IV. NUMERICAL RESULTS

This section gives a numerical example of our results for a binary symmetric channel (BSC). For a BSC with transition
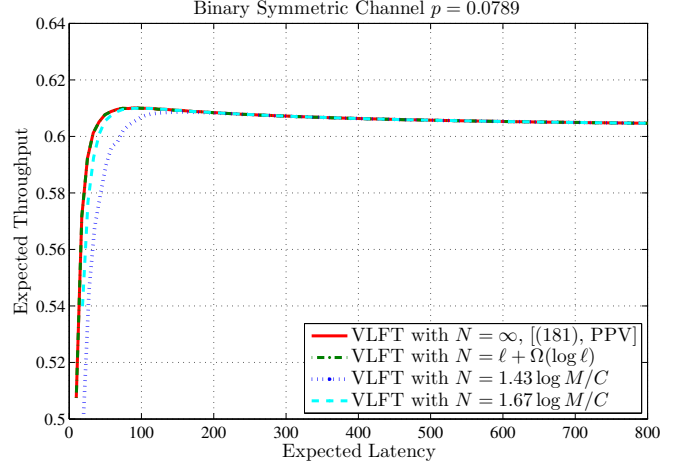

Fig. 1: Performance comparison of VLFT code achievability based on RCU bound with different codebook block-lengths.
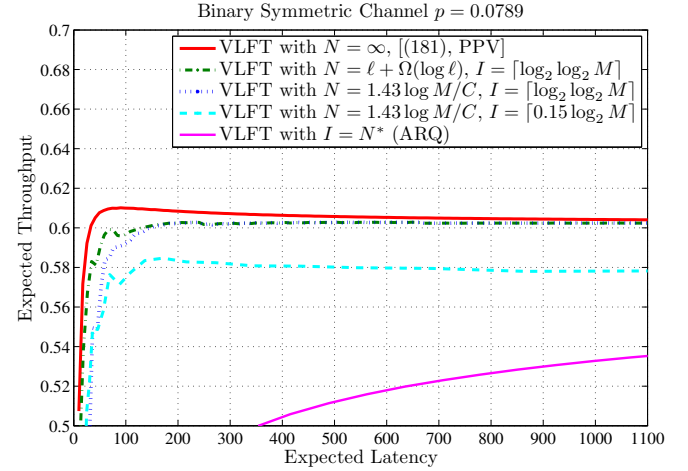

Fig. 2: Performance comparison of VLFT code achievability based on RCU bound with uniform increment and finite-length limitations.

probability p we used the RCU bound in [8], [11][1], which gives the following expression:

$$\xi_n \leq \sum_{t=0}^{n} \binom{n}{t} p^t (1-p)^{n-t} \min \left\{ 1, M \sum_{j=0}^{t} \binom{n}{j} 2^{-n} \right\}.$$

Fig. 1 shows the performance of VLFT codes over a BSC with $p = 0.0789$ with $N = \infty$ $N = \ell + \Omega(\log \ell)$, and $N = \log M/C_\Delta$. Since ℓ scales linearly with $\frac{\log M}{C}$, for the case of $\ell + \Omega(\log \ell)$ we choose N to scale as:

$$N = \frac{\log M}{C} + a \log \left(\frac{\log M}{C} \right) + b, \tag{25}$$

where $a, b > 0$ are constants to be chosen numerically. The numerical examples presented here use $a = 10$, $b = 30$. We choose $\Delta = 0.3C$ and $0.4C$, which are about 43% and 67% longer, respectively, than the block-length that corresponds to capacity. In other words, $N = 1.43 \log M/C$ and $N =$

---

[1]We replace $(M-1)$ by M for simplicity.

$1.67 \log M/C$ respectively. As expected latency increases in Fig. 1, expected throughput for the finite-N (repeated) VLFT codes converges to that of VLFT with $N = \infty$ before expected latency has reached 200 symbols. The penalty of $\Delta = 0.3C$ compared to $\Delta = 0.4C$ is only visible when M is small.

VLFT codes can have expected throughput higher than the original BSC capacity because of the beneficial effect of the error-free termination symbol. This effect becomes smaller as expected latency increases.

Fig. 2 shows the performance of the repeated VLFT code with various decoding-time increments I. As in (23), when I grows linearly with $\log M$ (i.e. $\lceil 0.15 \log_2 M \rceil$) then there is a constant gap from the $I = 1$ case. However, if I grows as $\lceil \log_2 \log_2 M \rceil$ then the gap from the $I = 1$ case decreases as expected latency increases. ARQ performance (in which $I = N*$, an optimized block-length) is also shown in the figure, which reveals a considerable performance gap from even the most constrained VLFT implementation we implemented.

V. Conclusion

This paper shows that the achievable performance of a VLFT code is mostly preserved when the block-length of the underlying code is restricted to be finite and decoding attempts are limited to regularly spaced decoding times. Specifically, if block-length $N = \Omega(\ell \log \ell)$ and $I = O(\log \ell)$ the optimal expansion of M_t^* is achieved.

The finite-block-length results for VLFT codes suggest that it is not necessary to use an infinitely long codebook or even a very large one. Numerical results show that a base code rate that is 60% of the capacity can closely approach performance of a VLFT code with $N = \infty$. Numerical results also show that decoding after every $\lceil \log_2 \log_2 M \rceil$ symbols is almost as good as decoding at every symbol in a practical application.

VI. Appendix

Proof of Thm. 2: Consider a random codebook $\mathcal{C}_N = \{\mathsf{C}_1, \ldots, \mathsf{C}_M\}$ with M codewords of length-N and codeword symbols independent and identically distributed according to P_X. To construct a VLFT code consider the following (U, f_n, g_n, τ): The common random variable $U \in \mathcal{U} = \mathcal{X}^N \times \cdots \times \mathcal{X}^N$. ($M$-fold product space) is distributed as: $U \sim \prod_{j=1}^{M} P_{X^N}$. A realization of U corresponds to a deterministic codebook $\{\mathsf{c}_1, \ldots, \mathsf{c}_M\}$. Let $x(n)$ denote the nth coordinate of a vector x. The sequence (f_n, g_n) and stopping time τ are defined as

$$f_n(U, W) = \mathsf{C}_W(n). \tag{26}$$

$$g_n(U, W, Y^n) = \arg \max_{j=1,\ldots,M} i(\mathsf{C}_W(n); Y^n). \tag{27}$$

$$\tau = \inf\{n : g_n(U, Y^n) = W\} \wedge N. \tag{28}$$

The nth marginal error event ζ_n is given as:

$$\zeta_n = \left\{ \bigcup_{j \neq W} i(\mathsf{C}_j^n; Y^n) > i(\mathsf{C}_w^n; Y^n) \right\}. \tag{29}$$

The following upper bounds are immediate from properties of probability:

$$\mathbb{E}[\tau] = \sum_{n=0}^{N-1} \mathbb{P}[\tau > n] \leq \sum_{n=0}^{N-1} \mathbb{P}[\zeta_n] \leq \sum_{n=0}^{N-1} \xi_n. \tag{30}$$

As in [8, (151)-(153)], the last inequality follows from union bound and the fact that a probability measure is upper bounded by 1. With a similar bounding technique, the error probability can be upper bounded as:

$$\mathbb{P}[g_\tau(U, Y^\tau) \neq W] = \mathbb{P}[g_N(U, Y^N) \neq W, \tau = N]$$
$$\leq \mathbb{P}[\zeta_N] \leq \xi_N.$$

In other words, the error probability is upper bounded by the error probability of the base code $\mathcal{C}_N$. ∎

Sketch Proof of Thm. 3: The proof follows from random coding and the following modification of the triplet (f_n, g_n, τ) of Thm. 2: For $k = 1, 2, \ldots$ let (f_n', g_n', τ') be:

$$f_n'(U, W) = \begin{cases} f_n(U, W) & \text{if } n \leq N \\ f_{n-kN}(U, W) & \text{if } kN < n \leq (k+1)N \end{cases}$$

$$g_n'(U, Y^n) = \begin{cases} g_n(U, Y^n) & \text{if } n \leq N \\ g_{n-kN}(U, Y_{kN+1}^n) & \text{if } kN < n \leq (k+1)N \end{cases}$$

$$\tau' = \inf\{n : g_n'(U, Y^n) = W\}$$

The zero-error part is obvious from the definition of the stopping time τ'. As mentioned above, the new encoder/decoder sequence (f_n', g_n') is simply an extension of the VLFT code in Thm. 2 by performing an ARQ-like repetition. Applying RCU bound for each n finishes the proof. ∎

References

[1] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channel with feedback–I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no.2, pp. 172–182, Apr. 1966.

[2] J. Schalkwijk, "A coding scheme for additive noise channel with feedback–II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, no.2, pp. 183–189, Apr. 1966.

[3] A. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Trans. Inf. Theory*, vol. IT-15, no.1, pp. 52–60, Jan. 1969.

[4] K. S. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Probl. Pered. Inform.*, vol. 6, no.1, pp. 87–92, 1970.

[5] T.-Y. Chen, N. Seshadri, and B.-Z. Shen, "Is feedback a performance equalizer of classic and modern codes?" in *Proc. 2010 Inf. Theory and Applications Workshop (ITA)*, San Diego, CA, USA, Feb. 2010.

[6] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 25, pp. 729–733, Nov. 1979.

[7] M. V. Burnashev, "Data transmission over a discrete channel with feedback. random transmission time," *Probl. Inf. Transm*, no. 4, pp. 10–30, 1976.

[8] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57(8), pp. 4903–4925, Aug. 2011.

[9] S. Verdú and S. Shamai, "Variable-rate channel capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2651 –2667, Jun. 2010.

[10] T.-Y. Chen, A. R. Williamson, and R. D. Wesel, "Variable-length coding with feedback: Finite-length codewords and periodic decoding," *arXiv:1301.7464*, Jan. 2013.

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