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
University of California, Los Angeles
Email: tychen@ee.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 mainly concerned with VLFT codes based on finite-length codes and decoding attempts only at certain specified decoding times. The penalties of using a finite block-length \( N \) and a sequence of periodic decoding times are studied. This paper shows that properly scaling \( N \) with the expected latency can achieve the same performance up to constant terms as with \( N = \infty \). The penalty introduced by periodic decoding times is a linear term of the interval between decoding times and hence the performance approaches capacity as the expected latency grows if the interval between 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 \(^1\)-\(^6\) suggest that feedback can be used to reduce the average block-length (or expected latency) required to approach capacity. As a practical demonstration, \(^7\) 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 \(^7\) qualitatively agrees with the error exponent analysis in \(^1\)-\(^6\).

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 \(^8\) showed that the optimal Burnashev error exponent \(^9\) 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. \(^10\) 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 \(^10\) assumes an underlying codebook with infinite-length codewords and decoding is attempted at every symbol so that the communication may be concluded after any given channel use. In practice, it may only be possible to use a codebook with finite-length codewords. 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 times. 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 rather than at every symbol.

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 \( j \)th element of \( x^n \), and \( x^n_j \) the \( j \)th to \( j \)th 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 \( \mathcal{X} = \mathcal{X}^n, \mathcal{Y} = \mathcal{Y}^n \) respectively. A channel is characterized by a conditional distribution \( P_{Y|X} = \prod_{i=1}^n P_{Y_i|X_i} \), where the equality holds because the channel is memoryless. For codes that make use of a noiseless feedback link, we consider causal channels \( \{P_{Y_i|X_{i-j}}\}_{j=1}^\infty \) 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 \(^10\):

**Definition 1:** An \((\ell,M,\epsilon)\) 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 \( f_n : U \times \mathcal{W} \times \mathcal{Y}^{n-1} \rightarrow \mathcal{X} \) 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:
   \[
   \mathbb{E}[\tau] \leq \ell.
   \] (1)
5) The final decision \( \hat{W} = g_\tau(U, Y^\tau) \) must satisfy:
   \[
   \mathbb{P}[\hat{W} \neq W] \leq \epsilon.
   \] (2)

As observed in [10], 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 [11], or FV codes, is a special class of VLFT codes that satisfies the following conditions:

\[
f_n(U, W, Y^{n-1}) = f_n(U, W)
\] (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:

\[
P_{X^n, Y^n, \bar{X}^n}(x^n, y^n, \bar{x}^n) = P_{X^n}(x^n)P_{\bar{X}^n}(\bar{x}^n) \prod_{j=1}^{n} P_{Y_i|X_j, Y_j-1}(y_j|x^j, y_j^{-1}),
\] (5)
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

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

The following is the achievability result in [10]:

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

\[
\ell \leq \sum_{n=0}^{\infty} \xi_n
\] (9)

where \( \xi_n \) is the following expectation:

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

The expression above is referred to as random coding union (RCU) bound. We take \( i(X^0; Y^0) = 0 \) and hence \( \xi_0 = 1 \).

Additionally, from the proof of [10] Thm. 11, we have:

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

**B. Problem Statement and Main Results**

Following the VLFT framework of [10], this paper studies the following problems:

(i) Finite-length codeword penalty for VLFT (FV) codes: The random coding approach in [10] 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.

(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. The case where the decoding times can be an arbitrary set of increments \( \{I_j\}_{j=1}^{\infty} \) is studied in [13].

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

**Definition 2:** Let \( M^*_\ell(\ell, N, I, \epsilon) \) be the maximum integer \( M \) such that there exist an \((\ell, M, N, I, \epsilon)\) VLFT code based on a code with block-length \( N \) and a decoder that only attempts decoding every \( I \) symbols. For zero-error codes where \( \epsilon = 0 \) we denote the maximum \( M \) as \( M^*_\ell(\ell, N, I) \) and for zero-error codes with \( I = 1 \) (i.e. decoding attempts after every received symbol) we denote the maximum \( M \) as \( M^*_\ell(\ell, N) \).

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:

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

\[
\log M^*_\ell(\ell, N, I) \geq \ell C - O(I).
\] (12)

Specifically, if we choose \( N > \ell + \frac{\log(\ell + 1) + \log \epsilon}{C} \) 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 \( I \). The proof is provided in Section III-C.

Of course, for practical applications that apply feedback to obtain reduced latency, the non-asymptotic behavior is critical. Numerical results on a binary symmetric channel demonstrate that properly selected values of \( N \) and \( I \) can 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 binary symmetric channel. Section V concludes the paper.
III. Finite Block-Lengths and Limited Decoding

In [10], 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 [10] is retained. We focus on achievability results under these constrained scenarios using proofs based on random FV codes. The general converse established in [10] 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 except that feedback indicates when it is the time to terminate transmission.

Letting \(\zeta_j\) be the marginal error event at the \(j\)th transmission, the expected latency \(E[\tau]\) is given as:

\[
E[\tau] = \sum_{n=1}^{\infty} n P[\tau = n] = \sum_{n \geq 1} P[\tau > n] = \sum_{n \geq 1} \prod_{j=1}^{n} \zeta_j .
\]

Consider a code \(C_N\) with finite block-length \(N\) where each element is a length-\(N\) \(X\)-valued string. Achievability results for an \((\ell, M, N, I, \epsilon)\) “truncated” VLFT code follow from a random coding argument. In particular we have the following:

**Theorem 3:** For any \(M > 0\) there exists an \((\ell, M, N, I, \epsilon)\) truncated VLFT code with

\[
\ell \leq \sum_{n=0}^{N-1} \xi_n \quad \epsilon \leq \xi_N .
\]

where \(\xi_n\) is the same as [10]. The proof is in the appendix.

Achievability results for \(\epsilon = 0\) can be obtained using an \((\ell, M, N, I, 1, 0)\) “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 through, for example, 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 4:** For every \(M > 0\) there exists an \((\ell, M, N, I, 0)\) repeated VLFT code such that

\[
\ell \leq \frac{1}{1 - \xi_N} \sum_{n=0}^{N-1} \xi_n
\]

where \(\xi_n\) 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 5:** For an \((\ell, M, N, 1, 0)\) repeated VLFT code with \(N = \Omega(\log M)\), we have the following expansion for a stationary DMC with capacity \(C\):

\[
\ell \leq \frac{\log M}{C} + O(1) .
\]

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

\[
O(1) \leq \frac{b_2 \log M}{C(M^b/c_\Delta - b_2)} + \frac{b_0 \log M}{C_\Delta M^a/c_\Delta} + a\]

where \(a\) depends on the mean and uniform bound of \(i(X; Y)\), and \(b_0, a\) are constants related to \(\Delta\) 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 define a pair of random walks to simplify the proofs:

\[
S_n \triangleq i(X^n; Y^n) \quad \bar{S}_n \triangleq i(X^n; Y^n) .
\]

For any measurable function \(f\) we have the property:

\[
E[f(\bar{X}^n; Y^n)] = E[f(X^n; Y^n)] \exp\{-S_n\} .
\]

Observe that \(S_n\) and \(\bar{S}_n\) are sums of i.i.d. r.v.s with positive and negative means \(E[i(X; Y)] = C\) and \(E[i(X; Y)] = -L\) (\(L\) is the lautum information [13] respectively. In particular \(\{S_n - nC\}\) is a bounded martingale and hence by Doob’s optional stopping theorem we have for a stopping time \(\tau\):

\[
E[S_{\tau}] = C E[\tau] .
\]

These properties are utilized in the following proofs.

**Proof:** Following the definition of (21) and (22), we first weaken the RCU bound by [11] and choosing \(\gamma = M:\)

\[
E \left[ \min \left\{ (1, (M - 1) P[\bar{S}_n \geq S_n | X^n Y^n]) \right\} \right] \leq E \left[ \exp \{-[S_n - \log M]^+\} \right] .
\]

Then from Thm[4] we have:

\[
\ell \leq \frac{1}{1 - \xi_N} \sum_{n=0}^{N-1} E \left[ \exp \{-[S_n - \log M]^+\} \right] .
\]

Consider an auxiliary stopping time \(\tau\) w.r.t. the filtration

\[
\mathcal{F}_n = \sigma\{X^n, \bar{X}^n, Y^n\}:
\]

\[
\tau' = \inf \{n \geq 0 : S_n \geq \log M \} \wedge N
\]

Denoting \(E[X; E] = E[X 1_E]\) where \(1_E\) is the indicator...
Therefore from (24) we have:
\[
\sum_{n=0}^{N-1} \mathbb{E} \left[ \exp \{-|S_n - \log M|^+ \} \right] = \mathbb{E} \left[ r' + \sum_{k=0}^{N-1-r'} \exp \{-|S_{k+r'} - \log M|^+ \}; r' < N \right] + \mathbb{N} \mathbb{P}[r' \geq N].
\] (29)

On \( \{r' < N\} \) we have \( i(X^{r'}; Y^{r'}) \geq \log M \) and hence:
\[
[S_{k+r'} - \log M]^+ = [S_{k+r'} - S_r + S_r - \log M]^+ \geq \left[ i(X^{k+r'}; Y^{k+r'}) - i(X^r; Y^r) \right]^+ \]
\[
= \left[ i(X^k; Y^k) \right]^+ \quad (31)
\]

Observe that \( S_n \) and \( \tilde{S}_n \) are sums of i.i.d. r.v.s with positive and negative means respectively. Thus by Chernoff inequality we have that:
\[
\mathbb{E} \left[ e^{-i(X^n; Y^n)} \right] = \mathbb{P} \left[ S_n > 0 \right] + \mathbb{P} \left[ \tilde{S}_n \leq 0 \right] \leq a_1 e^{-ka_2},
\] (35)

wher first equality follows from (18). Thus there is a constant \( a_3 > 0 \) such that:
\[
\sum_{k=1}^{N-1-r'} \mathbb{E} \left[ \exp \{ -|S_k|^+ \} \right] \leq \sum_{k=1}^{N-1} \mathbb{E} \left[ \exp \{ -|S_k|^+ \} \right] \leq a_3.
\] (37)

We assume that \( i(X^n; Y^n) \) has bounded jumps, and hence on the set \( \{r' < N\} \) there is a constant \( a_4 \) such that
\[
i(X^{r'}; Y^{r'}) - \log M \leq a_4 C.
\] (39)

Therefore from (24) we have:
\[
\mathbb{E}[r'] \leq \frac{\log M}{C} + a_4.
\] (40)

Letting \( a_5 = a_4 + a_4 \) and combining (37) and (40) we have:
\[
\mathbb{E}[r'] \leq N - \log M + a_5 \geq N + \mathbb{N} \mathbb{P}[r' \geq N] + a_5.
\] (41)

Again by noting that \( S_n \) is a sum of i.i.d. r.v.’s with mean \( C \), for \( NC > \log M \) we have by Chernoff inequality that:
\[
\mathbb{P}[\tau' \geq N] = \mathbb{P}[S_N < \log M] \leq b_1 \exp \left\{ -b_1 N \left( C - \frac{\log M}{N} \right) \right\} \leq b_0 \exp \left\{ -b_1 C \right\}.
\] (43)

Combining (41), (42) and (45) we have the following for \( \ell \):
\[
\ell \leq \left( \frac{\log M}{C} + \frac{b_0 \log M}{C \Delta} + a_5 \right) (1 - \mathbb{P}[\xi_n])^{-1} \quad (46)
\]

Notice that we are only interested in the first two terms of the expansion \((1 - x)^{-1} = 1 + x + x^2 + \ldots \) on \([0,1]\). Thus
\[
\ell \leq \frac{\log M}{C} + \frac{b_0 \log M}{C \Delta} + \frac{b_2 \log M}{C \Delta} + a_6 \quad (47)
\]

for some \( a_6 > 0 \). Hence for \( M \) large enough we have (19). ■

An expansion of \( \log M^* (\ell, N) \) requires \( N \) growing with \( \ell \). The components of the correction term in Thm. 5 however, depend on both \( N \) (as \( M/C \Delta) \) and \( M \). Indeed for a fixed \( \ell \), all \( M \) satisfying (19) and (20) are achievable. The argument we make below is that for any fixed constant \( c_0 > 0 \), there is an \( \ell_0 \) that depends logarithmically on \( c_0 \) such that the expansion \( \log M^* \geq C \ell - c_0 \) is true for all \( \ell \geq \ell_0 \). We first invoke the converse for an \( (\ell, M, \infty, 1, 0) \) VFLT code:

**Theorem 6 (10, Thm. 11):** Given a stationary DMC with capacity \( C \) we have the following for an \( (\ell, M, \infty, 1, 0) \) VFLT code:
\[
\log M^* \leq \ell C + \log(\ell + 1) + \log e.
\] (48)

Combining Thms. 5 and 6 we have the following:

**Corollary 1:** For an \( (\ell, M, N, 1, 0) \) repeated VFLT code with \( N = (1 + \delta) \ell \) and a proper choice of \( \delta > 0 \), we have the following for a stationary DMC with capacity \( C \):
\[
\log M^* \geq (C + O(1)).
\] (49)

**Proof:** We first choose \( N \) to scale with \( \ell \) with a factor \( \delta \):
\[
N = (1 + \delta) \ell.
\] (50)

Then by the converse we have:
\[
\frac{\log M}{N} \leq C + \frac{\log(\ell + 1) + \log e - \delta C}{(1 + \delta) \ell} \leq C - \delta'.
\] (51)

The term \( \delta' \) on the right is positive by setting:
\[
\delta > \frac{\log(\ell + 1) + \log e}{\ell C}.
\] (53)

Again by Chernoff inequality we have:
\[
\mathbb{P}[\tau' \geq N] = \mathbb{P}[S_N < \log M] \leq \mathbb{P}[S_N - NC < -N \delta'] \leq b_0' \exp \left\{ -b_1' \ell \right\}.
\] (56)

Since \( \delta \) is chosen such that \( \log M/N \) is less than capacity, we

\footnote{As opposed to the expression in (10), we use a minus sign for \( O(1) \) term to make the penalty clear.}
also have \((42)\). By reordering \((41)\) we have for some \(b_2', b_3' > 0\) such that:
\[
\log \frac{M}{C} \geq \ell \left[ 1 - b_2 e^{-\ell (1 + \delta) b_3} - (1 + \delta)b_0' \ell - b_3' \right],
\]
which implies \(\log M \geq C - O(1)\) for large enough \(\ell\). To conclude the 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 \(\delta\) decreasing with \(\ell\), to obtain the infinite-block-length expansion of \(M^*_\ell(\ell, \infty)\) provided in \([10]\).

Thus, the restriction to a finite block-length \(N\) does not restrict the asymptotic performance if \(N\) is selected properly with respect to \(\ell\). 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.

\[\text{B. Limited, Regularly-Spaced, Decoding Attempts}\]

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 \(i(X^n; Y^n)\) 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 7:** For an \((\ell, M, N, I, 0)\) VLFT code with uniform increments \(I\) and \(N = \infty\) we have the following expansion for a stationary DMC with capacity \(C\):
\[
\log M^*_\ell(\ell, \infty, I) \geq \ell C - O(I). \tag{58}
\]

**Proof:** Consider the same random coding scheme as in Thm. 4 but now the auxiliary stopping time is given as \(\tau_0 = n_1 + \tau_0 - 1\) where \(\tau_0\) is also a stopping time given as:
\[
\tau_0 = \min \{j > 0 : S_{n_j} = i(X^n; Y^n) \geq \log M\}. \tag{59}
\]

The rest is similar to the proof of Thm. 4.

\[
\ell \leq n_1 + I \sum_{j=1}^{\infty} \mathbb{P}[\zeta_j]
\]
\[
\leq n_1 + I \sum_{j=1}^{\infty} \mathbb{E} \left[ \exp \left\{ - \left[ S_{n_j} - \log M \right]^+ \right\} \right]
\]
\[
\leq n_1 + I \mathbb{E}[\tau_0 - 1]
\]
\[
+ \sum_{k=0}^{\infty} \mathbb{E} \left[ \exp \left\{ - \left[ S_{n_{\tau_0+k}} - \log M \right]^+ \right\} \right]
\]
\[
\leq \mathbb{E}[\tau_0] + I \sum_{k=0}^{\infty} \mathbb{E} \left[ \exp \left\{ - i(X^n; Y^n) \right\} \right]
\]
\[
\leq \mathbb{E}[\tau_0] + I a_3
\]
\[
\leq \frac{\log M}{C} + I a_3 + I a_3, \tag{65}
\]

where \(65\) follows by applying Chernoff inequality and \(65\) is because the jumps of \(i(X^n; Y^n)\) is bounded by \(I \times a_4 C\) for some \(a_4 > 0\). Reordering the equations gives the result. 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 that \(I = \log_2 \log_2 M\) yields good results for short block-lengths.

\[\text{C. Finite Block-Length and Limited Decoding Attempts}\]

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 8:** For an \((\ell, M, N, I, 0)\) VLFT code with \(N = \Omega(\log M)\), we have the following for a stationary DMC with capacity \(C\):
\[
\ell \leq (1 + \mathbb{P}[\zeta_N])^{-1} \log \frac{M}{C} + \mathbb{P}[\tau_0 \geq m] + O(I) \tag{66}
\]
\[
\leq \frac{\log M}{C} + O(I), \tag{67}
\]

where \(\tau_0\) is the stopping time in terms of the number of decoding attempts up to and including the first success. The proof is similar to Thm. 4 and can be found in the appendix. The proof of Thm. 4 now follows:

**Proof of Thm. 4:** For an \((\ell, M, N, I, 0)\) VLFT code pick \(N\) as follows:
\[
N = (1 + \delta)\ell, \quad \text{where} \quad \delta > \frac{\log(\ell + 1) + \log e}{\ell C}. \tag{68}
\]

The result follows by a similar argument as for Cor. 4. The restriction on the initial block-length \(n_1\) only makes a constant difference.\n
\[\text{IV. Numerical Results}\]

We give a numerical example of our results for a binary symmetric channel (BSC). For a BSC with transition probability \(p\) we used the RCU bound in \([10]\), \([14]\), which gives the following expression:
\[
\xi_n \leq \sum_{l=0}^{n} \left( \frac{n}{l} \right) p^l (1-p)^{n-l} \min \left\{ 1, M \sum_{j=0}^{l} \left( \frac{n}{j} \right) 2^{-n} \right\}. \tag{63}
\]

Fig. 1 shows the performance of VLFT codes over a BSC with \(p = 0.0789\) with \(N = \infty \rightarrow N = \ell + \Omega(\log \ell)\), and \(N = \log M/C\). Since \(\ell\) scales linearly with \(\log M/C\), for the case of \(\ell + \Omega(\log \ell)\) we choose \(N\) to scale as:
\[
N = \frac{\log M}{C} + a \log \left( \frac{\log M}{C} \right) + b, \tag{69}
\]

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\%
expected latency increases. ARQ performance (in which latency has reached a constant gap from the longer, respectively, than the block-length that corresponds to capacity. In other words, $N = 1.43 \log M/C$ and $N = 1.67 \log M/C$ 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 (12), when $I$ grows linearly with $\log M$ (i.e. $[0.15 \log M]$) then there is a constant gap from the $I = 1$ case. However, if $I$ grows as $[\log_2 \log_2 M]$ 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 $N = \ell + \Omega(\log \ell)$ and $I = O(\log \ell)$ the optimal expansion of $M^*$ is achieved.

The finite-block-length results for VLFT codes suggest that it is not necessary to use an infinitely long codebook even when very large one. Numerical results show that a base code rate that is $67\%$ of the capacity can closely approach performance of a VLFT code with an infinite block-length. Numerical results also show that decoding after every $\log_2 \log_2 M$ symbols is almost as good as decoding at every symbol.

VI. Appendix

Proof of Thm. 3 Consider a random codebook $C_N = \{c_1, \ldots, c_M\}$ 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 $(U, f_n, g_n, \tau)$: The common random variable
\[ U \in U = X^N \times \cdots \times X^N. \] is distributed as:
\[ U \sim \prod_{j=1}^M P_{X^N}. \]

A realization of $U$ corresponds to a deterministic codebook $\{c_1, \ldots, c_M\}$. Let $x(n)$ denote the $n$th coordinate of a vector $x$. The sequence $(f_n, g_n)$ is defined as
\[ f_n(U, W) = C_W(n) \] (72)
\[ g_n(U, W, Y^n) = \arg \max_{j=1,\ldots,M} i(C_W(n); Y^n) \] (73)
and the stopping time $\tau$ is defined as:
\[ \tau = \inf \{ n : g_n(U, Y^n) = W \} \land N. \] (74)
The $n$th marginal error event $\zeta_n$ is given as:
\[ \zeta_n = \left\{ \bigcup_{j \notin W} i(C^n_j; Y^n) > i(C^n_{w}; Y^n) \right\}. \] (75)
Following (13)-(15) we have
\[ \mathbb{E}[\tau] = \sum_{n=0}^{N-1} \mathbb{P}[\tau > n] \leq \sum_{n=0}^{N-1} \mathbb{P}[\zeta_n] \leq \sum_{n=0}^{N-1} \xi_n. \] (76)
As in (10) (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:
\[
\mathbb{P}[\mathcal{g}'(U, Y^T) \neq W] = \mathbb{P}[g^N(U, Y^N) \neq W, \tau = N]
\]
\[
= \mathbb{P} \left[ \bigcap_{j=1}^{N} \zeta_j \right]
\]
\[
\leq \mathbb{P}[\zeta_N] \leq \xi_N.
\]

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

**Proof of Thm.** Consider the FV code as in Thm. but with an initial-block-length \( n_1 \), an uniform increment \( I \) and a finite \( m \). The finite block-length is given by \( N = n_m \). Define the auxiliary stopping time as:
\[
\tau_0 = \inf \{ j > 0 : (X^n_j ; Y^n_j) \geq \log M \} \wedge m.
\]

Similar to Thm. we have \((86)\) to \((90)\), shown at the top of the page. Now we are left to choose the scaling of \( m \). Using a similar choice as in the proof of Thm. \( \xi \) which yields
\[
N = n_1 + (m - 1)I \geq \frac{\log M}{C}.
\]

Rest of the proof follows as in the proof of Thm. \( \xi \).

**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] B. Nakiboglu and R. Gallager, “Error exponents for variable-length block codes with feedback and cost constraints,” IEEE Trans. Inf. Theory, vol. 54, no. 3, pp. 945 –963, Mar. 2008.

[6] R. G. Gallager, “Variations on a theme by Schalkwijk and Kailath,” IEEE Trans. Inf. Theory, vol. 56, no. 1, pp. 6–17, Jan. 2010.

[7] 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.

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

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

[10] 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.

[11] S. Verdú and S. Shamai, “Variable-rate channel capacity,” IEEE Trans. Inf. Theory, vol. 56, no. 6, pp. 2651 –2667, Jun. 2010.
[12] T.-Y. Chen, A. R. Williamson, and R. D. Wesel, “Rate-compatible sphere-packing analysis,” draft, 2012.
[13] D. Palomar and S. Verdu, “Lautum information,” IEEE Trans. Inf. Theory, vol. 54, no. 3, pp. 964–975, Mar. 2008.
[14] 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.