Optimal Streaming Codes for Channels with Burst and Arbitrary Erasures

Silas L. Fong, Ashish Khisti and Baochun Li
University of Toronto
Toronto, ON M5S 3G4, Canada
E-mail: silas.fong@utoronto.ca

Wai-Tian Tan, Xiaoqing Zhu and John Apostolopoulos
Cisco Systems
San José, CA 95134, USA

***Abstract*—This paper considers transmitting a sequence of messages (a streaming source) over a packet erasure channel. In each time slot, the source constructs a packet based on the current and the previous messages and transmits the packet, which may be erased when the packet travels from the source to the destination. Every source message must be recovered perfectly at the destination subject to a fixed decoding delay. We assume that the channel loss model induces either a single burst erasure or multiple arbitrary erasures in all sliding windows of a fixed size. Under this channel loss model assumption, we fully characterize the maximum achievable rate by constructing streaming codes that achieve the optimal rate. In addition, our code construction implies the full characterization of the maximum achievable rate for convolutional codes with any given column distance, column span and decoding delay. Numerical results demonstrate that the optimal streaming codes outperform existing streaming codes of comparable complexity over some instances of the Gilbert-Elliott channel.**

I. Introduction

Low-latency video conferencing has been a cornerstone for communication and collaboration for individuals and enterprises. The advent of 5G promises to make high-throughput at low-latency ubiquitous. This enables new applications such as high-quality video conferencing, virtual reality (VR) and Internet-of-things (IoT) applications including vehicle-to-vehicle communication and mission-critical machine-type communication [1]. At the core of these important applications is the need to reliably deliver packets with low latency. Packet losses at the physical layer and the network layer are inevitable, which may be caused by unreliable wireless links or congestion at network bottlenecks. In order to alleviate the effect of packet losses on applications that are run over the Internet, two main error control schemes have been implemented at the transport and application layers: Automatic repeat request (ARQ) and forward error correction (FEC).

For long-distance low-latency communications, it is not suitable to use ARQ schemes for error control because each retransmission incurs an extra round-trip delay. More specifically, correcting an erasure using ARQ results in a 3-way delay (forward + backward + forward) and this aggregate (3-way) delay is required to be lower than 150 ms for interactive applications such as voice and video [2]. This aggregate delay makes ARQ impractical for communication between two distant points with aggregate delay larger than 150 ms (e.g., the aggregate delay between two diametrically opposite points on the earth's circumference is larger than 200 ms [3]).

In practice, packet losses experienced at the transport layer can be well approximated by statistical models [4], [5], including the well-known Gilbert-Elliott (GE) channel [6], [7] and its generalization the Fritchman channel [8]. In order to find good FEC codes for error correction at the application layer, it would be ideal if we could find the maximum achievable rate of a statistical model under a low decoding latency constraint and a given target error rate. However, characterizing such a rate over a statistical channel seems intractable. Therefore, researchers have been studying FEC codes over simplified channel models that provide useful approximations to practical low-latency communications over the Internet. One such model is the *packet-erasure channel* model that incurs burst erasures.

Correcting burst erasures using convolutional codes has a long history starting in the late 1950's, and the achievable rates for convolutional codes that correct burst erasures have been discussed in numerous works including [9]–[12], but the optimality of the convolutional codes under delay constraints was not discussed until the work by Martinian and Sundberg [13] in 2004. In [13], the maximum achievable rate for convolutional codes under delay constraints over a channel that induces burst erasures was fully characterized. Various generalizations of the burst erasure model and the low-latency convolutional codes in [13] have been proposed in [14]–[17].

This paper studies a packet-erasure channel model that induces *both burst and arbitrary erasures*. For this channel, optimal convolutional codes with rate 1/2 under delay constraints were discovered in [16] in 2013. Recently, lower and upper bounds on the maximum achievable rate of convolutional codes with delay constraints were provided by Badr et al. [18, Th. 1 and Th. 2]. The main result of this work is proposing low-latency convolutional codes that achieve the upper bound in [18, Th. 1], hence the maximum achievable rate of convolutional codes with delay constraints are fully characterized. Our simulation reveals that our proposed codes outperform all existing practical streaming codes over some instances of the GE channel. In addition, our main result implies the full characterization of the maximum achievable rate for convolutional codes with any given column distance, column span and decoding delay.

The rest of the paper is organized as follows. Section II describes the notation used in this paper. Section III presents the problem formulation and states the main result — the existence of optimal streaming codes over the packet erasure channel that induces burst and arbitrary erasures. Section IV presents the key preliminary results. Section V contains the proof of the main result. Section VI discusses optimal convolutional codes with any given column distance, column span and decoding delay. Section VII contains numerical results that compare the performance of the optimal convolutional codes with state-of-the-art schemes over the GE channel. Section VIII concludes this paper.

II. Notation

The set of non-negative integers is denoted by $\mathbb{Z}_+$. All the elements of any matrix considered in this paper are taken from a common finite field $\mathbb{F}$, where 0 and 1 denote the additive identity and the multiplicative identity respectively. The set of k-dimensional row vectors over $\mathbb{F}$ is denoted by $\mathbb{F}^k$, and the set of $k \times n$ matrices over $\mathbb{F}$ is denoted by $\mathbb{F}^{k\times n}$. For any matrix $\mathbf{G}$, we let $\mathbf{G}^t$ denote the transpose of $\mathbf{G}$. A row vector in $\mathbb{F}^k$ is denoted by $\mathbf{a} \triangleq [a_0\ a_1\ \ldots\ a_{k-1}]$ where a_ℓ denotes the $(\ell+1)^{\text{th}}$ element of $\mathbf{a}$. The k-dimensional identity matrix is denoted by $\mathbf{I}_k$ and the $L \times B$ all-zero matrix is denoted by $\mathbf{0}^{L\times B}$. An $L \times B$ parity matrix of a systematic maximum-distance separable (MDS) $(L+B, L)$-code is denoted by $\mathbf{V}^{L\times B}$, which possesses the property that any L columns of $[\mathbf{I}_L\ \mathbf{V}^{L\times B}] \in \mathbb{F}^{L\times(L+B)}$ are independent. It is well known that a systematic maximum-distance separable (MDS) $(L+B, L)$-code always exists as long as $|\mathbb{F}| \ge L + B$ [19].

III. Streaming Codes for Channels with Burst and Arbitrary Erasures

A. Problem formulation

The source wants to send a sequence of messages $\{\mathbf{s}_i\}_{i=0}^\infty$ to the destination. Each $\mathbf{s}_i$ is an element in $\mathbb{F}^k$ where $\mathbb{F}$ is some finite field. In each time slot $i \in \mathbb{Z}_+$, the message $\mathbf{s}_i$ is encoded into a length-n packet $\mathbf{x}_i \in \mathbb{F}^n$ to be transmitted to the destination through an erasure channel, and the destination receives $\mathbf{y}_i \in \mathbb{F}^n \cup \{*\}$ where $\mathbf{y}_i$ equals either $\mathbf{x}_i$ or the erasure symbol '$*$'. The code is subject to a delay constraint of T time slots, meaning that the destination must produce an estimate of $\mathbf{s}_i$, denoted by $\hat{\mathbf{s}}_i$, upon receiving $\mathbf{y}_{i+T}$.

In any sliding window that consists of W consecutive time slots, we assume that there exists either one burst erasure with length no larger than B or multiple arbitrary erasures with total count no larger than N. Since any burst erasure of length N can be viewed as N arbitrary erasures, we assume without loss of generality (wlog) that $B \ge N$. In addition, we assume wlog that $T \ge B$, or otherwise a burst erasure of length B starting from a certain time slot would wipe out the message transmitted in the same time slot. Throughout this paper, we assume that $W \ge T+1$ unless specified otherwise. The choice of the window size $W \ge T+1$ can be explained intuitively as follows — A message generated in a time slot must be decoded by the destination in T time slots, implying that the "lifespan" of each message is $T+1$. Setting the window size no smaller than the lifespan of a message enables us to investigate how the erasure pattern within the lifespan of a message affects the recovery of the packet. Nevertheless, the case where $W < T+1$ will also be discussed in Section VIII.

The goal of this paper is to characterize the maximum coding rate k/n that can be communicated over the packet-erasure channel such that every message can be perfectly recovered at the destination with delay T.

B. Standard definitions

Definition 1: An $(n, k, T)_\mathbb{F}$-streaming code consists of the following:
1. A sequence of messages $\{\mathbf{s}_i\}_{i=0}^\infty$ where $\mathbf{s}_i \in \mathbb{F}^k$.
2. An encoder $f_i : \underbrace{\mathbb{F}^k \times \ldots \times \mathbb{F}^k}_{i+1 \text{ times}} \to \mathbb{F}^n$ for each $i \in \mathbb{Z}_+$, where $\mathbf{x}_i = f_i(\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_i)$.
3. A decoder $\varphi_{i+T} : \underbrace{\mathbb{F}^n \cup \{*\} \times \ldots \times \mathbb{F}^n \cup \{*\}}_{i+T+1 \text{ times}} \to \mathbb{F}^k$ for each $i \in \mathbb{Z}_+$, where $\hat{\mathbf{s}}_i = \varphi_{i+T}(\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{i+T})$.

Definition 2: An $(n, k, m, T)_\mathbb{F}$-convolutional code is an $(n, k, T)_\mathbb{F}$-streaming code constructed as follows: Let $\mathbf{G}_0^{\text{conv}}, \mathbf{G}_1^{\text{conv}}, \ldots, \mathbf{G}_m^{\text{conv}}$ be $m+1$ generator matrices in $\mathbb{F}^{k\times n}$. Then

$$\mathbf{x}_i = \sum_{\ell=0}^{m} \mathbf{s}_{i-\ell}\, \mathbf{G}_\ell^{\text{conv}} \tag{1}$$

for each $i \in \mathbb{Z}_+$ where $\mathbf{s}_{-1} = \mathbf{s}_{-2} = \ldots = \mathbf{s}_{-m} = \mathbf{0}^{1\times k}$ by convention.

Remark 1: For an $(n, k, m, T)_\mathbb{F}$-convolutional code, m is commonly referred to as the *encoder memory* (see, e.g., [20, Sec. 1.4]), and the role of T specifies the decoding delay associated with the convolutional code.

Definition 3: An erasure sequence is a binary sequence denoted by $e^\infty \triangleq \{e_i\}_{i=0}^\infty$ where

$$e_i = \mathbf{1}\{\text{erasure occurs at time } i\}.$$

A (W, B, N)-erasure sequence is an erasure sequence e^∞ that satisfies the following: For each $i \in \mathbb{Z}_+$ and any window

$$\mathcal{W}_i \triangleq \{i, i+1, \ldots, i+W-1\},$$

either $N < \sum_{\ell\in\mathcal{W}_i} e_\ell \le B$ holds with all the 1's in $(e_i, e_{i+1}, \ldots, e_{i+W-1})$ occupying consecutive positions or $\sum_{\ell\in\mathcal{W}_i} e_\ell \le N$ holds with no restriction on the positions of 1's. The set of (W, B, N)-erasure sequences is denoted by $\Omega_{(W,B,N)}^\infty$.

Definition 4: Let $e \in \{0, 1\}$. The input-output relation of the erasure channel $g_n : \mathbb{F}^n \times \{0, 1\} \to \mathbb{F}^n \cup \{*\}$ subject to $e \in \{0, 1\}$ is defined as

$$g_n(\mathbf{x}, e) = \begin{cases} \mathbf{x} & \text{if } e = 0, \\ * & \text{if } e = 1. \end{cases} \tag{2}$$

For any erasure sequence e^∞ and any $(n, k, T)_\mathbb{F}$-streaming code, $\mathbf{y}_i = g_n(\mathbf{x}_i, e_i)$ holds for each $i \in \mathbb{Z}_+$.

Definition 5: An $(n, k, T)_{\mathbb{F}}$-streaming code is said to be (W, B, N)-achievable if the following holds for any (W, B, N)-erasure sequence $e^\infty \in \Omega^\infty_{(W,B,N)}$: For all $i \in \mathbb{Z}_+$ and all $\mathbf{s}_i \in \mathbb{F}^k$, we have $\hat{\mathbf{s}}_i = \mathbf{s}_i$.

Definition 6: The (W, T, B, N)-capacity, denoted by $C_{(W,T,B,N)}$, is the maximum rate achievable by all $(n, k, T)_{\mathbb{F}}$-streaming codes that are (W, B, N)-achievable.

C. Main Result

Recall that $W > T \geq B \geq N$ holds due to the explanations given in Section III-A.

Theorem 1: Fix any $W > T \geq B \geq N$ and suppose

$$|\mathbb{F}| > 2\left(\binom{T+1}{N} + T - B + 2\right). \tag{3}$$

Then, there exists an $(n, k, T, T)_{\mathbb{F}}$-convolutional code that is (W, B, N)-achievable where $k = T - N + 1$ and $n = k + B = T + B - N + 1$.

Combining Theorem 1, Definition 6 and the existing upper bound [18, Th. 1] that

$$C_{(W,T,B,N)} \leq \frac{T - N + 1}{T + B - N + 1},$$

we fully characterize the (W, T, B, N)-capacity to be

$$C_{(W,T,B,N)} = \frac{T - N + 1}{T + B - N + 1} \tag{4}$$

for all (W, T, B, N).

IV. Preliminaries for the Proof of Theorem 1

Definition 7: An $(n, k, T)_{\mathbb{F}}$-block code consists of a sequence of k source symbols $\{s_i\}_{i=0}^{k-1}$ where $s_i \in \mathbb{F}$, a generator matrix $\mathbf{G} \in \mathbb{F}^{k\times n}$ defined as

$$\mathbf{G} \triangleq \begin{bmatrix} \mathbf{I}_k & \mathbf{P} \end{bmatrix}$$

where $\mathbf{P} \in \mathbb{F}^{k\times(n-k)}$ is some parity-check matrix, and a decoding function φ_{i+T} for each $i \in \{0, 1, \ldots, k-1\}$ where φ_{i+T} is used by the destination at time $i + T$ to estimate s_i according to

$$\hat{s}_i = \begin{cases} \varphi_{i+T}(y_0, y_1, \ldots, y_{i+T}) & \text{if } i + T \leq n - 1, \\ \varphi_{i+T}(\underbrace{y_0, y_1, \ldots, y_{n-1}, *, \ldots, *}_{i+T+1 \text{ symbols}}) & \text{if } i + T > n - 1. \end{cases}$$

Definition 8: An $(n, k, T)_{\mathbb{F}}$-block code is said to be (W, B, N)-achievable if the following holds for any (W, B, N)-erasure sequence $e^\infty \in \Omega^\infty_{(W,B,N)}$: Suppose $y_i = g_1(x_i, e_i)$ holds for each $i \in \mathbb{Z}_+$ with g_1 being defined in (2). Then for all $i \in \{0, 1, \ldots, k-1\}$ and all $s_i \in \mathbb{F}$, we have $\hat{s}_i = s_i$.

The following lemma follows from the standard argument of interleaving a block code into a convolutional code [12] (see also [13, Sec. IV-A]) by means of periodic interleaving.

Lemma 1 ([21, Lemma 1]): Given an $(n, k, T)_{\mathbb{F}}$-block code which is (W, B, N)-achievable, we can construct an $(n, k, n-1, T)_{\mathbb{F}}$-convolutional code which is (W, B, N)-achievable.

For the case $T - N + 1 \geq B$, the following lemma shows the existence of a generator matrix that leads to a (W, B, N)-achievable $(n, k, T)_{\mathbb{F}}$-block code with rate $\frac{T-N+1}{T+B-N+1} > \frac{1}{2}$. One component of the generator matrix is an $m \times (N + m)$ *N-diagonal matrix* defined as

$$\mathbf{D}_N^{m\times(N+m)} \triangleq \begin{bmatrix} d_0^{(0)} & \cdots & d_{N-1}^{(0)} & 0 & \cdots & \cdots & 0 \\ 0 & d_0^{(1)} & \cdots & d_{N-1}^{(1)} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & d_0^{(m-1)} & \ldots & d_{N-1}^{(m-1)} & 0 \end{bmatrix} \tag{5}$$

with arbitrary values for $\{d_\ell^{(i)}\}_{\substack{0\leq i\leq m-1 \\ 0\leq \ell\leq N-1}}$.

Lemma 2 ([21, Lemmas 2 and 3]): Fix any $T \geq B \geq N$ and let $k \triangleq T - N + 1$ and $n \triangleq k + B$. Suppose $k \geq B$, which is equivalent to $k/n \geq 1/2$ (high-rate regime). If $\mathbb{F}$ satisfies (3), there exists a $\mathbf{P}$ having the form

$$\mathbf{P} = \left[\begin{array}{c|c} \multicolumn{2}{c}{\mathbf{D}_N^{(B-N)\times B}} \\ \hline \mathbf{0}^{N\times(B-N)} & \mathbf{P}_{\text{right}} \\ \hline \multicolumn{2}{c}{\mathbf{V}^{(k-B)\times B}} \end{array}\right] \tag{6}$$

such that $\mathbf{G} = [\mathbf{I}_k \ \mathbf{P}]$ is the generator matrix of a (W, B, N)-achievable code, where $\mathbf{D}_N^{m\times(N+m)}$ is an N-diagonal matrix as defined in (5), $\mathbf{P}_{\text{right}}$ is a $N \times N$ matrix, and $\mathbf{V}^{(k-B)\times B}$ denotes a $(k - B) \times B$ parity matrix of a systematic MDS code.

Remark 2: For the special case $N = 1$ with delay $T = k + N - 1 = k$, the parity-check matrix $\mathbf{P}$ in Lemma 2 reduces to the parity-check matrix of the Martinian-Sundberg scheme [13, Th. 2] in which $\mathbf{P}$ was simply chosen to be $\begin{bmatrix} \mathbf{I}_B \\ \mathbf{V}^{(k-B)\times B} \end{bmatrix}$. For the case $N > 1$ with delay $T = k + N - 1$, the Martinian-Sundberg scheme is no longer (W, B, N)-achievable because the row weight (number of non-zero elements) in each of the first B columns in the generator matrix of the base block code equals 2, implying that the contribution of some source symbol can be completely erased by some choice of 2 arbitrary erasures. In contrast, our choice of $\mathbf{P}$ in Lemma 2 having the form (6) ensures that the minimum row weight in the generator matrix is $N+1$, implying that the contribution of every source symbol is not completely erased by any choice of N arbitrary erasures.

Remark 3: For the special case $N = B$ with delay $T = k + N - 1 = n - 1$, we can simply choose $\mathbf{P}$ in Lemma 2 to be $\mathbf{V}^{k\times B}$ such that the resultant code is a maximum distance separable (MDS) code.

Example 1: Suppose $(W, T, B, N) = (6, 5, 3, 2)$ where $k = 4 \geq B$. Fix $\mathbb{F} = \text{GF}(41)$ so that (3) is satisfied. By Lemma 2, there exists a $(7, 4, 5)$-block code which is $(6, 3, 2)$-achievable where $\mathbf{G} = [\mathbf{I}_k \ \mathbf{P}]$ with $\mathbf{P}$ having the form (6). An example for such a $\mathbf{G}$ is

$$\mathbf{G} = \left[\begin{array}{cccc:ccc} 1 & 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right],$$

where the minimum row weight of the resultant $\mathbf{G}$ equals 3.

The following lemma is the counterpart of Lemma 2 for the case $T - N + 1 < B$ where the coding rate is less than $1/2$.

Lemma 3 ([21, Lemmas 2 and 4]): Fix any $T \ge B \ge N$ and let $k \triangleq T - N + 1$ and $n \triangleq k + B$. Suppose $k < B$, which is equivalent to $k/n < 1/2$ (low-rate regime). If $\mathbb{F}$ satisfies (3), there exists a $\mathbf{P}$ having the form

$$\left[\begin{array}{c:c:c} \mathbf{P}_{\text{left}} & \multicolumn{2}{c}{\mathbf{D}^{(B-N)\times k}_{k-B+N}} \\ \hdashline \mathbf{V}^{(k-B+N)\times(B-k)} & \mathbf{0} & \mathbf{V}^{(k-B+N)\times(k-B+N)} \end{array}\right] \tag{7}$$

such that $\mathbf{G} = [\mathbf{I}_k \ \mathbf{P}]$ is the generator matrix of a (W, B, N)-achievable code, where $\mathbf{P}_{\text{left}}$ is a $(B - N) \times (B - k)$ matrix, $\mathbf{D}^{(B-N)\times k}_{k-B+N}$ is a $(k - B + N)$-diagonal matrix as defined in (5), and $\mathbf{V}^{(k-B+N)\times N} \triangleq \left[\mathbf{V}^{(k-B+N)\times(B-k)} \ \mathbf{V}^{(k-B+N)\times(k-B+N)}\right]$ constitutes a $(k - B + N) \times N$ parity matrix of a systematic MDS code.

Remark 4: Suppose $k < B$. Then $N > 1$ must hold, and our choice of $\mathbf{P}$ in Lemma 3 having the form (7) ensures that the minimum row weight in the generator matrix is $N+1$. As in the case $k \ge B$ discussed in Remark 2, we see from (7) that the contribution of every source symbol is not completely erased by any choice of N arbitrary erasures.

Example 2: Suppose $(W, T, B, N) = (6, 5, 4, 3)$ where $k = 3 < B$. Fix $\mathbb{F} = \text{GF}(47)$ so that (3) is satisfied. By Lemma 3, there exists a $(7, 3, 5)$-block code which is $(6, 4, 3)$-achievable where $\mathbf{G} = [\mathbf{I}_k \ \mathbf{P}]$ with $\mathbf{P}$ having the form (7). A candidate for such a $\mathbf{G}$ is

$$\mathbf{G} = \left[\begin{array}{ccc:cccc} 1 & 0 & 0 & 1 & 1 & 3 & 0 \\ 0 & 1 & 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{array}\right],$$

where the minimum row weight of the resultant $\mathbf{G}$ equals 4.

V. Proof of Theorem 1

Choose a sufficiently large $\mathbb{F}$ which satisfies (3). Let $k \triangleq T - N + 1$ and $n \triangleq k + B$. Consider the following two cases:

Case $k \ge B$:

By Lemma 2, there exists an $(n, k, T)_{\mathbb{F}}$-block code with generator matrix $\mathbf{G} = [\mathbf{I}_k \ \mathbf{P}] \in \mathbb{F}^{k\times n}$ which is (W, B, N)-achievable where $\mathbf{P}$ has the form (6).

Case $k < B$:

By Lemma 3, there exists an $(n, k, T)_{\mathbb{F}}$-block code with generator matrix $\mathbf{G} = [\mathbf{I}_k \ \mathbf{P}] \in \mathbb{F}^{k\times n}$ which is (W, B, N)-achievable where $\mathbf{P}$ has the form (7).

Combining the two cases, there exists an $(n, k, T)_{\mathbb{F}}$-block code which is (W, B, N)-achievable. Based on the $(n, k, T)_{\mathbb{F}}$-block code, we can construct an $(n, k, n-1, T)_{\mathbb{F}}$-convolutional code according to Lemma 1. Since the convolutional code is subject to the delay constraint T, we can reduce the memory of the code from $n-1$ to T by simply truncating the sum in (1) without affecting the (W, B, N)-achievability of the code (a more formal argument can be found in [21, Sec. IV]). This concludes the proof.

VI. Optimal Convolutional Codes with Given Column Distance, Column Span and Delay

In this section, we will use Theorem 1 and existing results to derive the maximum achievable rate for convolutional codes given any column distance, column span and decoding delay. For an $(n, k, m, T)_{\mathbb{F}}$-convolutional code with memory m and generator matrices $\mathbf{G}^{\text{conv}}_0, \mathbf{G}^{\text{conv}}_1, \ldots, \mathbf{G}^{\text{conv}}_m$, define

$$\mathbf{G}^{\text{conv}} \triangleq \begin{bmatrix} \mathbf{G}^{\text{conv}}_0 & \mathbf{G}^{\text{conv}}_1 & \cdots & \mathbf{G}^{\text{conv}}_T \\ \mathbf{0}^{k\times n} & \mathbf{G}^{\text{conv}}_0 & \cdots & \mathbf{G}^{\text{conv}}_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^{k\times n} & \mathbf{0}^{k\times n} & \ldots & \mathbf{G}^{\text{conv}}_0 \end{bmatrix}$$

to be the truncated generator matrix where $\mathbf{G}^{\text{conv}}_\ell \triangleq \mathbf{0}^{k\times n}$ for any $m < \ell \le T$ by convention. The following definition is standard (see, e.g., [18, Appendix A]).

Definition 9: For each $(n, k, m, T)_{\mathbb{F}}$-convolutional code, the column distance and the column span are

$$\mathrm{d}_T \triangleq \min\left\{\mathrm{wt}\Big([\mathbf{s}_0 \ \mathbf{s}_1 \ \ldots \ \mathbf{s}_T]\,\mathbf{G}^{\text{conv}}\Big) \,\middle|\, \begin{array}{l} \mathbf{s}_0 \ne \mathbf{0}^{1\times k}, \mathbf{s}_\ell \in \mathbb{F}^k \\ \text{for each } 1 \le \ell \le T \end{array}\right\}$$

and

$$\mathrm{c}_T \triangleq \min\left\{\mathrm{span}\Big([\mathbf{s}_0 \ \mathbf{s}_1 \ \ldots \ \mathbf{s}_T]\,\mathbf{G}^{\text{conv}}\Big) \,\middle|\, \begin{array}{l} \mathbf{s}_0 \ne \mathbf{0}^{1\times k}, \mathbf{s}_\ell \in \mathbb{F}^k \\ \text{for each } 1 \le \ell \le T \end{array}\right\}$$

respectively, where

$$\mathrm{wt}\Big([\mathbf{x}_0 \ \mathbf{x}_1 \ \ldots \ \mathbf{x}_T]\Big) \triangleq \Big|\left\{i \in \{0, 1, \ldots, T\} \,\middle|\, \mathbf{x}_i \ne \mathbf{0}^{1\times n}\right\}\Big|$$

denotes the weight of $[\mathbf{x}_0 \ \mathbf{x}_1 \ \ldots \ \mathbf{x}_T]$ and

$$\begin{aligned} \mathrm{span}\Big([\mathbf{x}_0 \ \mathbf{x}_1 \ \ldots \ \mathbf{x}_T]\Big) \triangleq \max\left\{i \in \{0, \ldots, T\} \,\middle|\, \mathbf{x}_i \ne \mathbf{0}^{1\times n}\right\} \\ - \min\left\{i \in \{0, \ldots, T\} \,\middle|\, \mathbf{x}_i \ne \mathbf{0}^{1\times n}\right\} \end{aligned}$$

denotes the length of the support of $[\mathbf{x}_0 \ \mathbf{x}_1 \ \ldots \ \mathbf{x}_T]$ for any $[\mathbf{x}_0 \ \mathbf{x}_1 \ \ldots \ \mathbf{x}_T] \in \mathbb{F}^{(T+1)n}$.

The following proposition states a well-known fact regarding the column distance and the column span for convolutional codes (see, e.g., [18, Appendix A]).

Proposition 4: Any $(n, k, m, T)_{\mathbb{F}}$-convolutional code with column distance d_T and column span c_T is $(T+1, \mathrm{c}_T - 1, \mathrm{d}_T - 1)$-achievable. Conversely, if an $(n, k, m, T)_{\mathbb{F}}$-convolutional code is $(T + 1, B, N)$-achievable, then $\mathrm{d}_T \ge N + 1$ and $\mathrm{c}_T \ge B + 1$.

Combining Proposition 4 and (4), we conclude that $\frac{k}{n} \le \frac{T - \mathrm{d}_T + 2}{T + \mathrm{c}_T - \mathrm{d}_T + 1}$ for any $(n, k, m, T)_{\mathbb{F}}$-convolutional code with

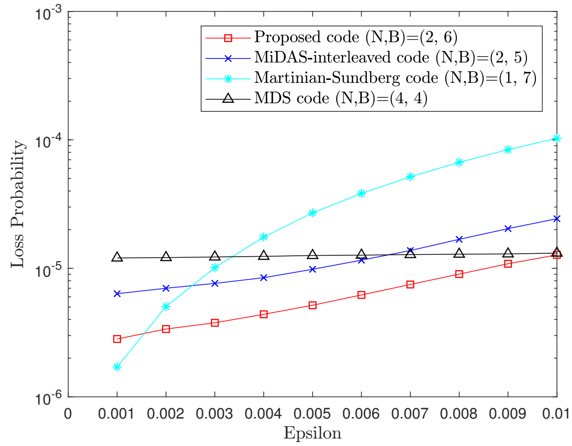

Fig. 1. Loss probabilities over the GE channel with $(\alpha, \beta) = (1\times 10^{-4}, 0.6)$

column distance d_T and column span c_T, which motivates us to define the optimality of a convolutional code as follows.

Definition 10: An $(n, k, m, T)_{\mathbb{F}}$-convolutional code with column distance d_T and column span c_T is said to be *optimal* if $\frac{k}{n} = \frac{T-\mathrm{d}_T+2}{T+\mathrm{c}_T-\mathrm{d}_T+1}$.

The following result regarding d_T and c_T for optimal convolutional codes is a natural consequence of Theorem 1.

Theorem 2 ([21, Th. 2]): Fix any T, d and c where $c \geq d \geq 1$, and let $\mathbb{F}$ be a finite field that satisfies

$$|\mathbb{F}| > 2\left(\binom{T+1}{\mathrm{d}-1} + T - c + 3\right). \tag{8}$$

Then, there exists an optimal $(n, k, T, T)_{\mathbb{F}}$-convolutional code with column distance $\mathrm{d}_T = d$ and column span $\mathrm{c}_T = c$.

VII. Numerical Results

The state-of-the-art MiDAS-interleaved and MiDAS-m-MDS convolutional codes have been proposed in [18, Sec. IV] for the erasure channel, whose constructions involve interleaved block codes and m-MDS codes respectively. In general, convolutional codes that involve m-MDS codes require large field size that grows exponentially in T (as mentioned in [18, Sec. IV-D]), hence they may not be practical for large T. On the other hand, convolutional codes that are based on interleaved block codes can be implemented with practical field size. In Figure 1, we plot the loss probabilities over the GE channel with constant parameters $(\alpha, \beta) = (1\times 10^{-4}, 0.6)$ and the varying good-state error probability ϵ for our random code, the MiDAS-interleaved code, the Martinian-Sundberg code [13] and the random MDS code with rates equal to $1/2$, $21/41 \approx 1/2$, $1/2$ and $1/2$ respectively. As shown in Figure 1, our random code outperforms all the other codes over the GE channel [6], [7] for $0.002 \leq \epsilon \leq 0.01$. For $\epsilon \leq 0.001$, the Martinian-Sundberg code performs the best, which indicates that the loss probability in this case is dominated by burst rather than arbitrary errors. For $\epsilon > 0.01$, the random MDS code performs the best, indicating that the loss probability in this case is dominated by arbitrary rather than burst errors.

VIII. Concluding Remarks

Throughout this paper, we have assumed that $W \geq T+1$ and showed that the maximum achievable rate is $C_{(W,T,B,N)} = \frac{T-N+1}{T+B-N+1}$.

For the case $W < T+1$, it was shown in [18, Th. 1] that $C_{(W,T,B,N)}$ is bounded as $C_{(W,T,B,N)} \leq \frac{W-N}{W+B-N}$ for any (W, T, B, N). On the other hand, it follows from Theorem 1 that $C_{(W,W-1,B,N)} = \frac{W-N}{W+B-N}$. Since $C_{(W,T,B,N)} \geq C_{(W,W-1,B,N)}$ due to the assumption that $W < T+1$, it follows that $C_{(W,T,B,N)} \geq \frac{W-N}{W+B-N}$. Consequently, $C_{(W,T,B,N)} = \frac{W-N}{W+B-N}$ for the case $W < T+1$.

References

[1] 5G-PPP, "5G empowering vertical industries," Tech. Rep., Feb. 2015. [Online]. Available: https://5g-ppp.eu/roadmaps/

[2] International Telecommunication Union, "Recommendation G.114," May 2003.

[3] A. K. A. Badr, W.-T. Tan, and J. Apostolopoulos, "Perfecting protection for interactive multimedia: A survey of forward error correction for low-delay interactive applications," *IEEE Signal Processing Magazine*, vol. 34, pp. 95 – 113, 2017.

[4] G. Hasslinger and O. Hohlfeld, "The Gilbert-Elliott model for packet loss in real time services on the Internet," in *14th GI/ITG Conference – Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB)*, Dortmund, Germany, Mar./Apr. 2008.

[5] O. Hohlfeld, R. Geib, and G. Hasslinger, "Packet loss in real-time services: Markovian models generating QoE impairments," in *16th Internatioal Workshop on Quality of Service*, Enschede, Netherlands, Jun. 2008.

[6] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, pp. 1253—1265, Sep. 1960.

[7] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell System Technical Journal*, vol. 42, pp. 1977–1997, Sep. 1963.

[8] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 221–227, 1967.

[9] D. W. Hagelbarger, "Recurrent codes: Easily mechanized, burst-correcting, binary codes," *Bell Sys. Tech. J.*, vol. 38, pp. 969 – 984, 1959.

[10] A. D. Wyner and R. Ash, "Analysis of recurrent codes," *IEEE Trans. Inf. Theory*, vol. 9, pp. 143 – 156, 1963.

[11] J. L. Massey, "Implementation of burst-correcting convolutional codes," *IEEE Trans. Inf. Theory*, vol. 11, no. 3, pp. 416 – 422, 1965.

[12] G. D. Forney, "Burst-correcting codes for the classic bursty channel," *IEEE Trans. Inf. Theory*, vol. 19, no. 5, pp. 772 – 781, 1971.

[13] E. Martinian and C.-E. W. Sundberg, "Burst erasure correction codes with low decoding delay," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2494 – 2502, 2004.

[14] D. Leong and T. Ho, "Erasure coding for real-time streaming," in *Proc. IEEE Intl. Symp. Inf. Theory*, Cambridge, MA, Jul. 2012.

[15] D. Leong, A. Qureshi, and T. Ho, "On coding for real-time streaming under packet erasures," in *Proc. IEEE Intl. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013.

[16] A. Badr, A. Khisti, W.-T. Tan, and J. Apostolopoulos, "Streaming codes for channels with burst and isolated erasures," in *IEEE INFOCOM*, Turin, Italy, Apr. 2013.

[17] N. Adler and Y. Cassuto, "Burst-erasure correcting codes with optimal average delay," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2848–2865, 2017.

[18] A. Badr, P. Patil, A. Khisti, W.-T. Tan, and J. Apostolopoulos, "Layered constructions for low-delay streaming codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 111 – 141, 2017.

[19] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, 1st ed. Amsterdam, Holland: North-Holland, Netherlands, 1988.

[20] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, 2nd ed. Hoboken, NY: John Wiley and Sons, Inc., 2015.

[21] S. L. Fong, A. Khisti, B. Li, W.-T. Tan, X. Zhu, and J. Apostolopoulos, "Optimal streaming codes for channels with burst and arbitrary erasures," *submitted to IEEE Trans. Inf. Theory*, Jan. 2018, arXiv:1801.04241 [cs.IT].