# Optimal Streaming Codes for Channels With Burst and Arbitrary Erasures

Silas L. Fong, *Member, IEEE*, Ashish Khisti, *Member, IEEE*, Baochun Li, *Fellow, IEEE*, Wai-Tian Tan, *Senior Member, IEEE*, Xiaoqing Zhu, *Member, IEEE*, and John Apostolopoulos, *Fellow, IEEE*

***Abstract*—This paper considers transmitting a sequence of messages (streaming messages) 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 introduces either one burst erasure or multiple arbitrary erasures in any fixed-sized sliding window. Under this channel loss assumption, we fully characterize the maximum achievable rate by constructing streaming codes that achieve the optimal rate. In addition, our construction of optimal streaming codes 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 and the Fritchman channel.**

***Index Terms*—Burst and arbitrary erasures, channel capacity, convolutional codes, column distance, column span, forward error correction, low-latency, packet erasure channel, sliding window, streaming codes.**

## I. Introduction

LOW-LATENCY video conferencing has been a cornerstone for communication and collaboration for individuals and enterprises. The advent of 5G networks 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 data link and transport layers: Automatic repeat request (ARQ) and forward error correction (FEC).

For long-distance low-latency communication, 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 including transmission, propagation and processing delays is required to be lower than 150 ms for interactive applications such as voice and video according to the International Telecommunication Union [2] (see [3] for an overview of the ubiquitous H.264/AVC video coding standard). This aggregate delay makes ARQ impractical for communication between two distant points with aggregate delay larger than 150 ms. For example, ARQ cannot be used for communication between two diametrically opposite points on the earth's circumference because the corresponding propagation delay alone is at least 200 ms [4].

For short-distance low-latency communications in the Tactile Internet [5], the next evolution of IoT, whose round-trip latency is required to be less than 1 ms [1], using ARQ schemes at the transport layer for error control is an inefficient use of precious time resources because the time budget allocated for retransmissions could instead be used for processing data at end users or data processing servers. Consider the example of remotely controlling a critical device where a sensor wants to communicate with an actuator in real time through a control server with round-trip latency less than 1 ms as illustrated in [1, Fig. 3]. The latency goals for processing delay at the terminals, transmission delay over the air interfaces between the terminals and the control server and data processing delay at the control server are 0.3 ms, 0.2 ms and 0.5 ms respectively. If an ARQ scheme is used for error control, then retransmissions compete the precious time resources with data computation at the terminals and the control server.

On the contrary, FEC schemes are amenable to low-latency communications because no retransmission is required. Instead of using retransmissions to achieve high reliability, FEC schemes increase the correlation among the transmitted symbols by adding redundant information. In other words,

Manuscript received January 15, 2018; revised December 5, 2018; accepted January 2, 2019. Date of publication January 22, 2019; date of current version June 14, 2019. This work was supported by a grant from the Cisco Research Center and matching funds from the NSERC CRD program. This paper was presented in part at the 2018 IEEE International Symposium on Information Theory.

S. L. Fong, A. Khisti, and B. Li are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: silas.fong@utoronto.ca; akhisti@ece.utoronto.ca; bli@ece.utoronto.edu).

W.-T. Tan, X. Zhu, and J. Apostolopoulos are with Enterprise Networking Innovation Labs, Cisco Systems, San Jose, CA 95134 USA.

Communicated by A. Jiang, Associate Editor for Coding Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2019.2894124

FEC schemes avoid the extra round-trip latency needed by retransmissions at the expense of the extra processing time spent on adding and removing redundant information at end users. Since FEC schemes inject redundancy at a constant rate while retransmissions inject redundancy at a highly non-uniform rate, FEC rather than ARQ schemes are more suitable for controlling delay for low-latency communications.

In order to search for FEC codes at the transport layer which are suitable for low-latency communications over the Internet, we are motivated to investigate the fundamental limits of low-latency streaming codes with FEC.

### *A. Motivation of Studying Packet-Erasure Channel*

In practice, packet losses experienced at the network layer can be well approximated by statistical models [6], [7], including the well-known Gilbert-Elliott (GE) channel [8], [9] and its generalization the Fritchman channel [10]. In order to find good FEC codes for error correction at the transport 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, we are motivated to study other simplified channel models that provide useful approximations to practical low-latency communications over the Internet.

In this paper, we focus on a packet-erasure channel model that introduces both burst and arbitrary errors. In any window of a fixed size, we assume that the channel introduces either a burst erasure or multiple arbitrary erasures. Although this channel model is not statistical, it has been shown in [11] that streaming FEC codes that correct both burst and arbitrary erasures can significantly outperform traditional streaming FEC codes that correct only one type of erasures (either burst or arbitrary) for both the GE channel and the Fritchman channel.

### *B. System Model*

In order to describe the existing results for the packet-erasure channel model, we would like to briefly describe the channel model. A formal description will appear later in the paper. The channel consists of a source and a destination. In each time slot, the source chooses a collection of $k$ symbols destined for the destination and encodes the $k$ symbols into a collection of $n$ symbols followed by transmitting the $n$ symbols through the channel. The collection of $n$ symbols transmitted in a time slot are either received perfectly by the destination or erased (lost). The fraction $k/n$ specifies the coding rate. We call the $k$ symbols chosen by the source, the $n$ symbols transmitted by the source and the $n$ symbols received by the destination the *source packet*, the *transmitted packet* and the *received packet* respectively. Since every low-latency application is subject to a tight delay constraint, we assume that every source packet generated in a time slot must be decoded with delay $T$, i.e., within the future $T$ time slots.

In order to capture the packet loss behavior over the Internet, we first consider the simple scenario where either one burst erasure with length no longer than $B$ occurs or multiple arbitrary erasures with total count no larger than $N$ occur on the discrete timeline. Since a channel that introduces any $N$ arbitrary erasures can introduce any burst erasure of length $N$, we assume without loss of generality (wlog) that

$$B \geq N. \tag{1}$$

In order to avoid triviality, we assume wlog that

$$B > 0, \tag{2}$$

or otherwise $B = N = 0$ by (1) in which case no coding is needed to achieve the maximum rate one. Similarly, a channel that introduces any burst erasure of a positive length can introduce one arbitrary erasure, hence we assume wlog that

$$N \geq \begin{cases} 1 & \text{if } B > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

In addition, we assume wlog that

$$T \geq B, \tag{4}$$

or otherwise a burst erasure of length $B$ starting from a certain time slot would prevent the destination from timely recovering the source packet transmitted in the same time slot. Under the erasure channel model described above, we are interested in characterizing the maximum coding rate $k/n$ for sending information over the channel such that every source packet can be perfectly recovered by the destination with delay $T$. In the rest of the paper, we assume wlog the following holds due to (1), (2), (3) and (4):

$$T \geq B \geq N \geq 1. \tag{5}$$

### *C. Related Work*

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 [12]–[15], but the optimality of the convolutional codes under delay constraints was not discussed until the work by Martinian and Sundberg [16] in 2004. In [16], streaming codes for the special case $N = 1$ are considered and the maximum achievable rate for convolutional codes over a channel that introduces only a single burst erasure (because $N = 1$) was proved to be $\frac{T}{T+B}$. Various generalizations of the burst erasure model and the low-latency convolutional codes in [16] have been proposed in [17]–[20].

For a channel that introduces both burst and arbitrary erasures as described in Section I-B, optimal convolutional codes with rate 1/2 were discovered in [19] in 2013. Recently, it was proved by Badr *et al.* [11, Ths. 1 and 2] that the maximum achievable rate is bounded between $\frac{T-N}{T+B-N}$ and $\frac{T-N+1}{T+B-N+1}$ for any $(T, B, N)$.

### *D. Main Contribution*

This paper studies the *sliding window model* suggested in [11] which generalizes the simple system model described in Section I-B. Under this model, we assume that either one burst erasure with length no longer than $B$ occurs or multiple arbitrary erasures with total count no larger than $N$ occur in any sliding window of size $W$. If we set $W = \infty$, then the sliding window model reduces to the simple system model described in Section I-B. Throughout this paper, we assume

$$W \geq T + 1 \tag{6}$$

unless specified otherwise. The assumption of the window size $W \geq T + 1$ can be explained intuitively as follows—A source packet generated in a time slot must be decoded by the destination in $T$ time slots, implying that the "lifespan" of each source packet is $T+1$. Setting the window size no smaller than the lifespan of a source packet enables us to investigate how the erasures within the lifespan of a source packet affects the recovery of the packet. Nevertheless, the case where $W < T + 1$ will also be discussed in the sequel.

Under the sliding window model, Badr *et al.* [11, Ths. 1 and 2] showed that the maximum achievable rate lies between $\frac{T-N}{T+B-N}$ and $\frac{T-N+1}{T+B-N+1}$ for any $(W, T, B, N)$, which is not a satisfactory result because the lower and upper bounds do not coincide for any $(W, T, B, N)$. The main result of this paper shows that the upper bound is indeed achievable, i.e., the maximum achievable rate equals $\frac{T-N+1}{T+B-N+1}$ for any $(W, T, B, N)$. This generalizes the results in [16] and [19] and strengthens the result in [11] (cf. Section I-C). The exact statement of our main result will be stated in Section II-C. The proof of the main result can be divided into the following two steps:

1. Construct an $(n, k)$-block code with $\frac{k}{n} = \frac{T-N+1}{T+B-N+1}$ having the following property: The destination can perfectly recover the $k$ source symbols with decoding delay $T$ as long as the block code is used over the erasure channel in $n$ consecutive time slots.
2. Convert the $(n, k)$-block code into a convolutional code by periodic interleaving [15].

The details of the above two steps can be found in Section IV and Section III.

In addition, our construction of optimal streaming codes implies the full characterization of the maximum achievable rate for convolutional codes with any given column distance, column span and decoding delay, whose details can be found in Section VII. Simulation results in Section IX reveal that our proposed codes outperform all existing practical streaming codes over some instances of the GE channel and the Fritchman channel.

### *E. Paper Outline*

This paper is organized as follows. The notation in this paper is explained in the next subsection. Section II presents the formulation of streaming codes for the packet erasure channel and states the main result. Section III presents the preliminary results — a standard procedure for interleaving a block code into a streaming code and two key lemmas which enable us to construct block codes that can be interleaved to form optimal streaming codes. Section IV contains the proof of the main result, i.e., the existence of optimal streaming codes over the packet erasure channel for all parameters of $(W, T, B, N)$. The optimal streaming codes take the form of convolutional codes obtained by interleaving the block codes as described in the two key lemmas in Section III. Section V and Section VI present the proofs of the two key lemmas respectively. In Section VII, we discuss the column distance and the column span for low-latency convolutional codes, and use the result in Section IV to characterize the maximum achievable rate for convolutional codes with fixed column distance, column span and decoding delay. Section VIII describes a practical random code construction of optimal low-latency convolutional codes. Section IX contains numerical results that compare the performance of the optimal convolutional codes with state-of-the-art schemes over the GE channel and the Fritchman channel. Section X concludes this paper.

### *F. 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 $\mathbf{G}^t$ and rank$(\mathbf{G})$ denote respectively the transpose and the rank 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 $a_\ell$ 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 MDS $(L + B, L)$-code always exists as long as $|\mathbb{F}| \geq L + B$ [21]. For a matrix $\mathbf{G} \in \mathbb{F}^{k\times n}$, the column space of $\mathbf{G}$ is the set space$(\mathbf{G}) \triangleq \{\mathbf{G}\boldsymbol{\alpha} | \boldsymbol{\alpha} \in \mathbb{F}^{n\times 1}\}$. A $W$-dimensional tuple is denoted by $e^W \triangleq (e_0, e_1, \ldots, e_{W-1})$ where $e_i$ denotes the $(i + 1)^{\text{th}}$ element of $e^W$. The $W$-dimensional diagonal matrix with diagonal elements $e^W$ is denoted by

$$\text{diag}(e_0, e_1, \ldots, e_{W-1}) \triangleq \begin{bmatrix} e_0 & 0 & \cdots & 0 \\ 0 & e_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & e_{W-1} \end{bmatrix}.$$

## II. Streaming Codes for Channels With Burst and Arbitrary Erasures

This section formally defines our system model that was described in Sections I-B and I-D, and states the main result.

### *A. Problem Formulation*

The source wants to send a sequence of length-$k$ packets $\{\mathbf{s}_i\}_{i=0}^{\infty}$ 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 source packet $\mathbf{s}_i$ 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 $W \geq T+1$ consecutive time slots, there exists either one burst erasure with length no longer than $B$ or multiple arbitrary erasures with total count no larger than $N$. By the assumptions (5) and (6), we assume

$$W > T \geq B \geq N \geq 1 \tag{7}$$

unless specified otherwise.

### *B. Standard Definitions*

The formal definition of the streaming code described in the previous subsection is stated as follows.

*Definition 1 [11, Sec. II-B]:* An $(n, k, T)_{\mathbb{F}}$-streaming code consists of the following:

1) A sequence of source packets $\{\mathbf{s}_i\}_{i=0}^{\infty}$ where $\mathbf{s}_i \in \mathbb{F}^k$.
2) An encoding function

$$f_i : \underbrace{\mathbb{F}^k \times \ldots \times \mathbb{F}^k}_{i+1 \text{ times}} \to \mathbb{F}^n$$

for each $i \in \mathbb{Z}_+$, where $f_i$ is used by the source at time $i$ to encode $\mathbf{s}_i$ according to

$$\mathbf{x}_i = f_i(\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_i).$$

3) A decoding function

$$\varphi_{i+T} : \underbrace{\mathbb{F}^n \cup \{*\} \times \ldots \times \mathbb{F}^n \cup \{*\}}_{i+T+1 \text{ times}} \to \mathbb{F}^k$$

for each $i \in \mathbb{Z}_+$, where $\varphi_{i+T}$ is used by the destination at time $i+T$ to estimate $\mathbf{s}_i$ according to[1]

$$\hat{\mathbf{s}}_i = \varphi_{i+T}(\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{i+T}). \tag{8}$$

*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 for each $i \in \mathbb{Z}_+$,

$$\mathbf{x}_i = \sum_{\ell=0}^{m} \mathbf{s}_{i-\ell}\, \mathbf{G}_\ell^{\text{conv}} \tag{9}$$

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., [23, Sec. 1.4]), and the role of $T$ specifies the decoding delay associated with the convolutional code (cf. (8)).

*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 $e^\infty$ that satisfies the following: For each $i \in \mathbb{Z}_+$ and any window

$$\mathcal{W}_i \triangleq \{i, i+1, \ldots, i+W-1\}, \tag{10}$$

either $N < \sum_{\ell \in \mathcal{W}_i} e_\ell \leq B$ holds with all the 1's in $(e_i, e_{i+1}, \ldots, e_{i+W-1})$ occupying consecutive positions or $\sum_{\ell \in \mathcal{W}_i} e_\ell \leq N$ holds with no restriction on the positions of 1's. In other words, a $(W, B, N)$-erasure sequence introduces either one burst erasure with length no longer than $B$ or multiple arbitrary erasures with total count no larger than $N$ in any window $\mathcal{W}_i$, $i \in \mathbb{Z}_+$. The set of $(W, B, N)$-erasure sequences is denoted by $\Omega_{(W,B,N)}^{\infty}$.

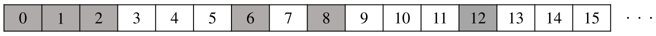

Fig. 1. A periodic (5, 3, 2)-erasure sequence with period 16.

*Example 1:* Suppose $(W, B, N) = (5, 3, 2)$. Consider the periodic sequence with period 16 as shown in Figure 1. The sequence is in $\Omega_{(5,3,2)}^{\infty}$ because in any sliding window of size $W = 5$, there is either a single burst erasure of length no longer than $B = 3$ or no more than $N = 2$ arbitrary erasures.

*Definition 4:* The mapping $g_n : \mathbb{F}^n \times \{0, 1\} \to \mathbb{F}^n \cup \{*\}$ of the erasure channel is defined as

$$g_n(\mathbf{x}, e) = \begin{cases} \mathbf{x} & \text{if } e = 0, \\ * & \text{if } e = 1. \end{cases} \tag{11}$$

For any erasure sequence $e^\infty$ and any $(n, k, T)_{\mathbb{F}}$-streaming code, the following input-output relation holds for the erasure channel for each $i \in \mathbb{Z}_+$:

$$\mathbf{y}_i = g_n(\mathbf{x}_i, e_i). \tag{12}$$

*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 $e^\infty \in \Omega_{(W,B,N)}^{\infty}$: For all $i \in \mathbb{Z}_+$ and all $\mathbf{s}_i \in \mathbb{F}^k$, we have

$$\hat{\mathbf{s}}_i = \mathbf{s}_i$$

where

$$\begin{aligned} \hat{\mathbf{s}}_i &= \varphi_{i+T}\big(\mathbf{y}_0, \ldots, \mathbf{y}_{i+T}\big) \\ &= \varphi_{i+T}\big(g_n(\mathbf{x}_0, e_0), \ldots, g_n(\mathbf{x}_{i+T}, e_{i+T})\big) \end{aligned}$$

due to (8) and (12).

*Definition 6:* Fix any $(W, T, B, N)$ that satisfies (7). The $(W, T, B, N)$-capacity, denoted by $C_{(W,T,B,N)}$, is the supremum of the rates attained by $(n, k, T)_{\mathbb{F}}$-streaming codes that are $(W, B, N)$-achievable, i.e.,

$$C_{(W,T,B,N)} \triangleq \sup \left\{ \frac{k}{n} \,\middle|\, \begin{array}{l} \text{A } (W, B, N)\text{-achievable } (n, k, T)_{\mathbb{F}}\text{-} \\ \text{streaming code exists for some } \mathbb{F} \end{array} \right\}.$$

It was shown in [11, Ths. 1 and 2] that

$$\frac{T-N}{T+B-N} \leq C_{(W,T,B,N)} \leq \frac{T-N+1}{T+B-N+1} \tag{13}$$

holds for any $(W, T, B, N)$. Our main result stated in the next subsection closes the gap.

[1] Early decoding is not considered in this definition. In practice, early decoding could decrease the average delay of decoding. See [22] for an implementation of streaming codes where early decoding is permitted. However, the theoretical and simulation results in this paper remain unchanged even if early decoding is permitted because this paper focuses on maximum rather than average decoding delay.

### C. Main Result

*Theorem 1:* Fix any $(W, T, B, N)$ that satisfies (7) and fix a finite field $\mathbb{F}$ such that

$$|\mathbb{F}| > 2\left(\binom{T+1}{N} + T - B + 2\right). \tag{14}$$

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

Combining Theorem 1, Definition 6 and (13), we fully characterize the $(W, T, B, N)$-capacity to be

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

for all $(W, T, B, N)$ that satisfies (7), which generalizes the capacity results for the special case $N = 1$ in [16] and for the special case $\frac{T-N+1}{T+B-N+1} = \frac{1}{2}$ in [19]. In particular, the upper bound in (13) obtained in [11] is tight and the supremum in Definition 6 can be replaced with a maximum.

## III. Preliminaries for the Proof of Theorem 1

An important step of the proof of Theorem 1 is to construct streaming codes by periodically interleaving block codes. The definition of a block code is formally stated as follows.

*Definition 7:* An $(n, k, T)_{\mathbb{F}}$-block code consists of the following:

1) A set of $k$ source symbols $\{s[i]\}_{i=0}^{k-1}$ where $s[i] \in \mathbb{F}$.
2) A $k \times n$ generator matrix

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

where $\mathbf{P} \in \mathbb{F}^{k\times(n-k)}$ is some parity-check matrix to be determined later. The codeword is generated as

$$\begin{bmatrix}x[0]\ x[1]\ \ldots\ x[n-1]\end{bmatrix} \triangleq [s[0]\ s[1]\ \ldots\ s[k-1]]\,\mathbf{G}. \tag{15}$$

3) A decoding function

$$\varphi_{i+T} : \underbrace{\mathbb{F}\cup\{*\} \times \ldots \times \mathbb{F}\cup\{*\}}_{i+T+1 \text{ times}} \to \mathbb{F}$$

for each $i \in \{0, 1, \ldots, k-1\}$, where $\varphi_{i+T}$ is used by the destination at time $i + T$ to estimate $s[i]$ according to

$$\hat{s}[i] = \begin{cases} \varphi_{i+T}(y[0], \ldots, y[i+T]) & \text{if } i + T \le n-1, \\ \varphi_{i+T}(\underbrace{y[0], \ldots, y[n-1], *, \ldots, *}_{i+T+1 \text{ symbols}}) & \\ & \text{if } i + T > n-1. \end{cases} \tag{16}$$

The following definition concerns the error-correcting capability of $(n, k, T)_{\mathbb{F}}$-block codes.

*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[2] $e^\infty \in \Omega^\infty_{(W,B,N)}$: Let

$$y[i] \triangleq g_1(x[i], e_i) \tag{17}$$

be the symbol received by the destination at time $i$ for each $i \in \{0, 1, \ldots, n-1\}$ where $g_1$ is as defined in (11). For the $(n, k, T)_{\mathbb{F}}$-block code, we have

$$\hat{s}[i] = s[i]$$

for all $i \in \{0, 1, \ldots, k-1\}$ and all $s[i] \in \mathbb{F}$ where $\hat{s}[i]$ is constructed according to (16) and (17).

The following lemma implies that constructing a $(W, B, N)$-achievable convolutional code is not more difficult than constructing a $(W, B, N)$-achievable block code. The proof of the following lemma is deferred to Appendix A because it follows the standard argument of interleaving a block code into a convolutional code by means of periodic interleaving [15] (see also [16, Sec. IV-A]).

*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. More specifically, given that $\mathbf{G} = \begin{bmatrix}g_{i,j}\end{bmatrix}_{\substack{0\le i\le k-1,\\ 0\le j\le n-1}}$ is the generator matrix of the $(n, k, T)_{\mathbb{F}}$-block code where $g_{i,j}$ is the entry situated in row $i$ and column $j$ of $\mathbf{G}$, we can construct the $n-1$ generator matrices of the $(n, k, n-1, T)_{\mathbb{F}}$-convolutional code as follows: For each $\ell \in \{0, 1, \ldots, n-1\}$, construct $\mathbf{G}_\ell^{\text{conv}}$ according to (18) as shown at the bottom of this page where $\mathbf{G} = \sum_{\ell=0}^{n-1} \mathbf{G}_\ell^{\text{conv}}$. In particular, if we let $\mathbf{s}_i \triangleq [s_i[0]\ s_i[1]\ \cdots\ s_i[k-1]]$ and let

$$\begin{aligned}&\begin{bmatrix}x_i[0]\ x_{i+1}[1]\ \cdots\ x_{i+n-1}[n-1]\end{bmatrix} \\ &\qquad\qquad \triangleq \begin{bmatrix}s_i[0]\ s_{i+1}[1]\ \cdots\ s_{i+k-1}[k-1]\end{bmatrix}\mathbf{G}\end{aligned}$$

for all $i \in \mathbb{Z}_+$, then the symbols generated at time $i$ by the $(n, k, n-1, T)_{\mathbb{F}}$-convolutional code are

$$\mathbf{x}_i \triangleq \begin{bmatrix}x_i[0]\ x_i[1]\ \cdots\ x_i[n-1]\end{bmatrix}. \tag{19}$$

*Example 2:* Suppose we are given a $(5, 3, 2)$-achievable $(6, 3, 4)_{\mathbb{F}}$-block code with generator matrix

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 2 \end{bmatrix}.$$

Let $\{\mathbf{s}_i\}_{i\in\mathbb{Z}_+}$ be a sequence of streaming messages where $\mathbf{s}_i = \begin{bmatrix}s_i[0]\ s_i[1]\ s_i[2]\end{bmatrix} \in \mathbb{F}^3$. From time $i-2$ to $i+5$, the symbols yielded by the $(6, 3, 5, 4)_{\mathbb{F}}$-convolutional code constructed by interleaving the $(6, 3, 4)_{\mathbb{F}}$-block code according Lemma 1 are shown in Table I. The symbols in Table I highlighted in the

$$\mathbf{G}_\ell^{\text{conv}} \triangleq \begin{cases} \begin{bmatrix}\mathbf{0}^{k\times\ell} & \operatorname{diag}\left(g_{0,\ell}, g_{1,\ell+1}, \ldots, g_{k-1,\ell+k-1}\right) & \mathbf{0}^{k\times(n-k-\ell)}\end{bmatrix} & \text{if } 0 \le \ell \le n-k, \\ \left[\begin{array}{c:c} \mathbf{0}^{k\times\ell} & \begin{matrix}\operatorname{diag}\left(g_{0,\ell}, g_{1,\ell+1}, \ldots, g_{n-1-\ell,n-1}\right) \\ \mathbf{0}^{(k-n+\ell)\times(n-\ell)}\end{matrix} \end{array}\right] & \text{if } n-k < \ell \le n-1, \end{cases} \tag{18}$$

[2] Only the first $n$ elements of $e^\infty$ play a role in the definition.

TABLE I

SYMBOLS YIELDED BY A $(6, 3, 5, 4)_{\mathbb{F}}$-CONVOLUTIONAL CODE THROUGH INTERLEAVING A $(6, 3, 4)_{\mathbb{F}}$-BLOCK CODE

| Time<br>Symbol | $i-2$ | $i-1$ | $i$ | $i+1$ | $i+2$ | $i+3$ | $i+4$ | $i+5$ |
|---|---|---|---|---|---|---|---|---|
| 0 | $s_{i-2}[0]$ | $s_{i-1}[0]$ | $s_i[0]$ | $s_{i+1}[0]$ | $s_{i+2}[0]$ | $s_{i+3}[0]$ | $s_{i+4}[0]$ | $s_{i+5}[0]$ |
| 1 | $s_{i-2}[1]$ | $s_{i-1}[1]$ | $s_i[1]$ | $s_{i+1}[1]$ | $s_{i+2}[1]$ | $s_{i+3}[1]$ | $s_{i+4}[1]$ | $s_{i+5}[1]$ |
| 2 | $s_{i-2}[2]$ | $s_{i-1}[2]$ | $s_i[2]$ | $s_{i+1}[2]$ | $s_{i+2}[2]$ | $s_{i+3}[2]$ | $s_{i+4}[2]$ | $s_{i+5}[2]$ |
| 3 | $\ddots$ | $\ddots$ | $\ddots$ | $s_{i-2}[0]$ | $s_{i-1}[0]$ | $s_i[0]$ | $\ddots$ | $\ddots$ |
| 4 | $\ddots$ | $\ddots$ | $\ddots$ | $\ddots$ | $s_{i-2}[0]$<br>$+ s_{i-1}[1]$<br>$+ s_i[2]$ | $s_{i-1}[0]$<br>$+ s_i[1]$<br>$+ s_{i+1}[2]$ | $s_i[0]$<br>$+ s_{i+1}[1]$<br>$+ s_{i+2}[2]$ | $\ddots$ |
| 5 | $\ddots$ | $\ddots$ | $\ddots$ | $\ddots$ | $\ddots$ | $s_{i-1}[1]$<br>$+ 2s_i[2]$ | $s_i[1]$<br>$+ 2s_{i+1}[2]$ | $s_{i+1}[1]$<br>$+ 2s_{i+2}[2]$ |

same color diagonally (in ↘ direction) are encoded using the same $(6, 3, 4)_{\mathbb{F}}$-block code. Given the fact that each $(6, 3, 4)_{\mathbb{F}}$-block code is $(5, 3, 2)$-achievable, we can see from Table I that $\mathbf{s}_i = \big[s_i[0]\ s_i[1]\ s_i[2]\big]$ can be perfectly recovered by time $i+5$ as long as the erasure sequence belongs to $\Omega^{\infty}_{(5,3,2)}$.

Instead of proving Theorem 1 by enumerating all possible $(W, B, N)$-erasure sequences, we will prove Theorem 1 by enumerating a small subset of sequences called *maximal* $(W, B, N)$*-erasure patterns.*

*Definition 9:* A maximal $(W, B, N)$-erasure pattern is a $W$-dimensional binary tuple $\varepsilon^W$ that satisfies either $\sum_{\ell=0}^{W-1} \varepsilon_\ell = B$ with all the 1's in $\varepsilon^W$ occupying consecutive positions or $\sum_{\ell=0}^{W-1} \varepsilon_\ell = N$ with no restriction on the positions of 1's. The set of maximal $(W, B, N)$-erasure patterns is denoted by $\Omega^W_{B,N}$.

Recall the definition of window $\mathcal{W}_i$ in (10) (where $|\mathcal{W}_i| = W$). For any $(W, B, N)$-erasure sequence $e^\infty$ and any $\mathcal{W}_i$, there always exists a maximal $(W, B, N)$-erasure pattern $\varepsilon^W$ such that $e_\ell \le \varepsilon_\ell$ for all $\ell \in \mathcal{W}_i$ by Definition 3 and Definition 9. The following lemma enables us to prove Theorem 1 by considering only maximal $(T+1, B, N)$-erasure patterns in $\Omega^{T+1}_{B,N}$ rather than all possible $(W, B, N)$-erasure sequences in $\Omega^{\infty}_{(W,B,N)}$. Before presenting the lemma, we define the following notations which will be used in the rest of the paper. We let $\mathbf{u}_i^{(k)}$ denote the $k$-dimensional unit column vector $[\mathbf{0}^{1\times i}\ 1\ \mathbf{0}^{1\times(k-i-1)}]^t$ for each $i \in \{0, 1, \ldots, k-1\}$, let

$$\mathbf{I}_j^{(k)} \triangleq \begin{bmatrix} \mathbf{0}^{(k-j)\times(k-j)} & \mathbf{0}^{(k-j)\times j} \\ \mathbf{0}^{j\times(k-j)} & \mathbf{I}_j \end{bmatrix} \tag{20}$$

be the $k \times k$ diagonal matrix which embeds $\mathbf{I}_j$ as a submatrix for each $j \in \{0, 1, \ldots, k\}$, and let

$$\mathbf{E}_{\varepsilon^j} \triangleq \mathbf{I}_j - \operatorname{diag}(\varepsilon^j) \tag{21}$$

be the $j \times j$ diagonal matrix with diagonal elements $(1 - \varepsilon_0)$, $(1 - \varepsilon_1), \ldots, (1 - \varepsilon_{j-1})$ for any length-$j$ binary tuple $\varepsilon^j$. We will always multiply $\mathbf{E}_{\varepsilon^j}$ on the right side of a matrix having $j$ columns, and the multiplication characterizes the erasure operation introduced by $\mathbf{E}_{\varepsilon^j}$ by zeroing the columns of the multiplied matrix according to $\varepsilon^j$. The proof of the lemma is straightforward and hence relegated to Appendix B.

*Lemma 2:* Fix any $(W, T, B, N)$ that satisfies (7). Let $\mathbf{G} = [\mathbf{g}_0\ \mathbf{g}_1\ \ldots\ \mathbf{g}_{n-1}]$ be a $k \times n$ matrix in $\mathbb{F}^{k\times n}$, and let

$$\mathbf{G}_i \triangleq \begin{cases} [\,\mathbf{g}_i\ \ldots\ \mathbf{g}_{i+T}\,] & \text{if } i \le n-T-1, \\ \big[\,\mathbf{g}_i\ \ldots\ \mathbf{g}_{n-1} \,\vdots\, \mathbf{0}\,\big] & \text{if } i > n-T-1 \end{cases} \tag{22}$$

be a submatrix of $\mathbf{G}$ for each $i \in \{0, 1, \ldots, k-1\}$ where $\mathbf{0}$ is the $k \times (i + T - n + 1)$ zero matrix. The $(n, k, T)_{\mathbb{F}}$-block code with generator matrix $\mathbf{G}$ is $(W, B, N)$-achievable if $\mathbf{G}$ satisfies the following sufficient condition:

For each $i \in \{0, 1, \ldots, k-1\}$ and each maximal $(T+1, B, N)$-erasure pattern $\varepsilon^{T+1} \in \Omega^{T+1}_{B,N}$, it is true that

$$\mathbf{u}_i^{(k)} \in \operatorname{space}\big(\mathbf{I}_{k-i}^{(k)}\, \mathbf{G}_i \mathbf{E}_{\varepsilon^{T+1}}\big). \tag{23}$$

*Remark 2:* Lemma 2 transforms the problem of finding optimal $(n, k, T)_{\mathbb{F}}$-block codes that are $(W, B, N)$-achievable into a purely algebraic problem stated in (23). The physical meaning of (23) can be interpreted as follows: Suppose $[x[0]\ x[1]\ \cdots\ x[n-1]] = [s[0]\ s[1]\ \cdots\ s[k-1]]\,\mathbf{G}$. Then, (23) implies that $s_i$ can be perfectly recovered by time $i + T$ as long as $s_0, s_1, \ldots, s_{i-1}$ have been perfectly recovered and the erasure patten in $\mathcal{W}_i$ is in $\Omega^{T+1}_{B,N}$.

The following lemma shows the existence of a generator matrix $\mathbf{G}$ which satisfies the sufficient condition in (23) when $T - N + 1 \ge B$. One component of the generator matrix is an $m \times (N + m)$ *N-diagonal matrix* defined as

$$\mathbf{D}_N^{m\times(N+m)} \triangleq \begin{bmatrix} d_0^{(0)} & \cdots & d_{N-1}^{(0)} & 0 & \cdots & \cdots & 0 \\ 0 & d_0^{(1)} & \cdots & d_{N-1}^{(1)} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & d_0^{(m-1)} & \cdots & d_{N-1}^{(m-1)} & 0 \end{bmatrix} \tag{24}$$

with arbitrary values for $\{d_\ell^{(i)}\}_{\substack{0\le i\le m-1\\ 0\le \ell\le N-1}}$. The proof of the following lemma is tedious and is therefore deferred to Section V.

*Lemma 3:* Fix any $(W, T, B, N)$ that satisfies (7) and let $k \triangleq T - N + 1$ and $n \triangleq k + B$. Suppose $k \ge B$, which is equivalent to $k/n \ge 1/2$ (high-rate regime). If $\mathbb{F}$ satisfies (14), there exists a $\mathbf{P} \in \mathbb{F}^{k\times B}$ having the form

$$\left[\begin{array}{c|c} \multicolumn{2}{c}{\mathbf{D}_N^{(B-N)\times B}} \\ \hline \mathbf{0}^{N\times(B-N)} & \mathbf{P}_{\text{right}} \\ \hline \multicolumn{2}{c}{\mathbf{V}^{(k-B)\times B}} \end{array}\right] \tag{25}$$

such that $\mathbf{G} = [\,\mathbf{I}_k\ \mathbf{P}]$ satisfies (23) for all $i \in \{0, 1, \ldots, k\}$ and all $\varepsilon^{T+1} \in \mathbf{\Omega}_{B,N}^{T+1}$, where $\mathbf{D}_N^{m\times(N+m)}$ is an $N$-diagonal matrix as defined in (24), $\mathbf{P}_{\text{right}}$ is a $N \times N$ matrix with non-zero entries, and $\mathbf{V}^{(k-B)\times B}$ denotes the $(k-B) \times B$ parity matrix of a systematic MDS code.

*Remark 3:* For the special case $N = 1$ with delay $T = k + N - 1 = k$, the parity-check matrix $\mathbf{P}$ in Lemma 3 reduces to the parity-check matrix of the Martinian-Sundberg scheme [16, 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 $T = k + N - 1 > k$, 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 3 having the form (25) ensures that the minimum row weight of 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. Since $T = k+N-1$ and $n = k+B$, it follows that $B-N$ symbols encoded by $\mathbf{G}$ need to be decoded before the whole block has been received.

*Remark 4:* For the special case $N = B$ with delay $T = k + N - 1 = n - 1$, we can simply choose $\mathbf{P}$ in Lemma 3 to be $\mathbf{V}^{k\times B}$ such that the resultant code is an MDS code. In this case, the decoding of every symbol encoded by $\mathbf{G}$ can be performed after the whole block has been received because $T = n - 1$.

*Example 3:* Suppose $(W, T, B, N) = (6, 5, 3, 2)$ where $k = 4 \geq B$. Fix $\mathbb{F} = \mathrm{GF}(41)$ so that (14) is satisfied. By Lemma 3, there exists a $\mathbf{G} = [\,\mathbf{I}_k\ \mathbf{P}]$ with $\mathbf{P}$ having the form (25) such that $\mathbf{G}$ satisfies (23). A candidate for such a $\mathbf{G}$ is

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

where the minimum row weight of $\mathbf{G}$ equals 3. In particular, condition (23) is satisfied for each $i \in \{0, 1, 2, 3\}$ and each maximal $(6, 3, 2)$-erasure pattern $\varepsilon^6 \in \mathbf{\Omega}_{(3,2)}^6$ due to the following two facts:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \in \text{space}\left(\left[\begin{array}{cccc|cc} 1 & 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{array}\right]\mathbf{E}_{\varepsilon^6}\right)$$

and

$$\text{space}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{I}_3 \end{bmatrix}\right) \subseteq \text{space}\left(\left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]\mathbf{E}_{\varepsilon^6}\right)$$

where $\mathbf{E}_{\varepsilon^6} = \mathrm{diag}(1 - e_0, 1 - e_1, \ldots, 1 - e_5)$ and $\mathbf{0}$ is the $1 \times 3$ zero matrix. The effect of $\mathbf{E}_{\varepsilon^6}$ is to replace the columns of the multiplied matrix whose indices are inside $\{i \in \{0, 1, 2, 3, 4, 5\} | e_i = 1\}$ with $\mathbf{0}^{4\times 1}$, which is equivalent to "erasing" those columns when we evaluate the column space of the multiplied matrix. Since $\varepsilon^6$ is an arbitrary maximal $(6, 3, 2)$-erasure pattern, the erased columns specified by $\mathbf{E}_{\varepsilon^6}$ take the form of any consecutive 3 columns or any 2 arbitrary columns. The intuition behind the idea of finding $\mathbf{G}$ is explained as follows. Consider the baseline Martinian-Sundberg matrix (cf. Remark 3) denoted by

$$\mathbf{G}^* \triangleq \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

which has the same dimension as the desired $\mathbf{G}$ with dimension $(T-N+1)\times(T+B-N+1) = 4\times 7$. Since the minimum row weight of $\mathbf{G}^*$ equals 2, some symbols cannot be recovered if the channel is subject to 2 arbitrary erasures. Therefore, we are motivated to construct a $\mathbf{G}$ with minimum row weight 3 by replacing some zeros in $\mathbf{G}^*$ with non-zeros so that $\mathbf{G}$ would satisfy (23) for each $i \in \{0, 1, 2, 3\}$. This example remains valid if we replace GF(41) by GF(5), which is not surprising because (14) is only a sufficient condition on $\mathbb{F}$.

*Example 4:* Suppose $(W, T, B, N) = (8, 7, 4, 2)$ where $k = 6 \geq B$. Fix $\mathbb{F} = \mathrm{GF}(67)$ so that (14) is satisfied. By Lemma 3, there exists a $\mathbf{G} = [\,\mathbf{I}_k\ \mathbf{P}]$ with $\mathbf{P}$ having the form (25) such that $\mathbf{G}$ satisfies (23). An example for such a $\mathbf{G}$ is

$$\mathbf{G} = \left[\begin{array}{cccccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 6 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 5 & 5^2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 4^2 & 4^3 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 3^2 & 3^3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 2^2 & 2^3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{array}\right].$$

The following lemma shows the existence of a generator matrix $\mathbf{G}$ which satisfies the sufficient condition in (23) when $T - N + 1 < B$. The proof is tedious and is therefore deferred to Section VI.

*Lemma 4:* Fix any $(W, T, B, N)$ that satisfies (7) 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 (14), there exists a $\mathbf{P} \in \mathbb{F}^{k\times B}$ having the form

$$\left[\begin{array}{c|cc} \mathbf{P}_{\text{left}} & \multicolumn{2}{c}{\mathbf{D}_{k-B+N}^{(B-N)\times k}} \\ \cline{2-3} \mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)} & \mathbf{0} & \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)} \end{array}\right] \tag{26}$$

such that $\mathbf{G} = [\,\mathbf{I}_k\ \mathbf{P}]$ satisfies (23) for all $i \in \{0, 1, \ldots, k\}$ and all $\varepsilon^{T+1} \in \mathbf{\Omega}_{B,N}^{T+1}$, where $\mathbf{P}_{\text{left}}$ is a $(B-N) \times (B-k)$ matrix, $\mathbf{D}_{k-B+N}^{(B-N)\times k}$ is a $(k-B+N)$-diagonal matrix as defined in (24),

$\left[\mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)}\ \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)}\right]$ constitutes the $(k-B+N)\times N$ parity matrix of a systematic MDS code, and $\mathbf{0}$ is the $(k-B+N)\times(B-N)$ zero matrix.

*Remark 5:* Suppose $k<B$. Then $N>1$ must hold, and our choice of $\mathbf{P}$ in Lemma 4 having the form (26) ensures that the minimum row weight of the generator matrix is $N+1$. As in the case $k\geq B$ discussed in Remark 3, we see from (26) that the contribution of every source symbol is not completely erased by any choice of $N$ arbitrary erasures, and $B-N$ symbols encoded by $\mathbf{G}$ need to be decoded before the whole block has been received.

*Example 5:* Suppose $(W,T,B,N)=(6,5,4,3)$ where $k=3<B$. Let $\mathbb{F}=\text{GF}(47)$ so that (14) is satisfied. By Lemma 4, there exists a $\mathbf{G}=[\,\mathbf{I}_k\ \mathbf{P}]$ with $\mathbf{P}$ having the form (26) such that $\mathbf{G}$ satisfies (23). A candidate for such a $\mathbf{G}$ is

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

where the minimum row weight of $\mathbf{G}$ equals 4. This example remains valid if we replace GF(47) by GF(5), which is not surprising because (14) is only a sufficient condition on $\mathbb{F}$.

*Example 6:* Suppose $(W,T,B,N)=(8,7,6,4)$ where $k=4<B$. Let $\mathbb{F}=\text{GF}(149)$ so that (14) is satisfied. By Lemma 4, there exists a $\mathbf{G}=[\,\mathbf{I}_k\ \mathbf{P}]$ with $\mathbf{P}$ having the form (26) such that $\mathbf{G}$ satisfies (23). A candidate for such a $\mathbf{G}$ is

$$\left[\begin{array}{cccc:cccccc}1&0&0&0&1&4&4^2&4^3&0&0\\0&1&0&0&1&3&0&3^3&3^4&0\\0&0&1&0&1&2&0&0&2^4&2^5\\0&0&0&1&1&1&0&0&1&1\end{array}\right].$$

## IV. Proof of Theorem 1

Fix any $(W,T,B,N)$ that satisfies (7) and choose a sufficiently large $\mathbb{F}$ which satisfies (14). Let $k\triangleq T-N+1$ and $n\triangleq k+B$. Consider the following two cases:

*Case $k\geq B$:* By Lemma 3 and 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 (25).

*Case $k<B$:* By Lemma 4 and 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 (26).

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. In addition, since $\mathbf{P}$ has the form either (25) or (26), it follows from (18) that $\mathbf{G}_\ell^{\text{conv}}=\mathbf{0}^{k\times n}$ for any $\ell\geq k+N=T+1$, which implies that the $(n,k,n-1,T)_{\mathbb{F}}$-convolutional code is also an $(n,k,T,T)_{\mathbb{F}}$-convolutional code (cf. Definition 2). This concludes the proof.

## V. Proof of Lemma 3

Fix any $(W,T,B,N)$ that satisfies (7) and recall that $k=T-N+1$. Suppose $k\geq B$. Fix any finite field $\mathbb{F}$ that satisfies (14). Our goal is to show that $\mathbf{G}=[\,\mathbf{I}_k\ \mathbf{P}]$ satisfies (23) for some $\mathbf{P}$ having the form (25), where $\mathbf{I}_{k-i}^{(k)}$, $\mathbf{G}_i$ and $\mathbf{E}_{\varepsilon^{T+1}}$ in (23) are as defined in (20), (22) and (21) respectively. To this end, we construct a variable vector

$$\vec{q}^{\,(i)}\triangleq\left[q_0^{(i)}\ q_1^{(i)}\ \ldots\ q_{N-1}^{(i)}\right]\in\mathbb{F}^N$$

for each $i\in\{0,1,\ldots B-1\}$ where the values of the $B$ vectors will be determined later in this proof. In addition, we define

$$\vec{p}^{\,(i)}\triangleq\begin{cases}[\mathbf{0}^{1\times i}\ \vec{q}^{\,(i)}\ \mathbf{0}^{1\times(B-N-i)}] & \text{if } 0\leq i\leq B-N-1,\\ [\mathbf{0}^{1\times(B-N)}\ \vec{q}^{\,(i)}] & \text{if } B-N\leq i\leq B-1.\end{cases}\tag{27}$$

Construct a $(k-B)\times B$ parity matrix of a systematic MDS $(k,k-B)$-code denoted by $\mathbf{V}^{(k-B)\times B}$, which always exists because $|\mathbb{F}|\geq 2(2T-B+1)\geq k$ by (14). Then, let

$$\mathbf{P}\triangleq\left[\begin{array}{c}\vec{p}^{\,(0)}\\ \vdots\\ \vec{p}^{\,(B-1)}\\ \hdashline \mathbf{V}^{(k-B)\times B}\end{array}\right]\tag{28}$$

where $\vec{p}^{\,(i)}$ denotes the $(i+1)^{\text{th}}$ row of $\mathbf{P}$. It can be seen that $\mathbf{P}$ has the form (25). It remains to show that $\mathbf{G}=[\mathbf{I}_k\ \mathbf{P}]$ satisfies (23) for $i=k-1,k-2,\ldots,0$ for some $\{\vec{q}^{\,(i)}\}_{i=0}^{B-1}$. By inspecting (23) and (28), we see that condition (23) depends on only $\{\vec{q}^{\,(B-1-j)}\}_{j=0}^{i}$ for each $i\in\{B-1,B-2,\ldots,0\}$ and does not depend on $\{\vec{q}^{\,(j)}\}_{j=0}^{B-1}$ for each $i\in\{k-1,k-2,\ldots,B\}$. In the rest of the proof, we will verify condition (23) in the order $i=k-1,k-2,\ldots,0$, which means that we will choose $\{\vec{q}^{\,(i)}\}_{i=0}^{B-1}$ by choosing $\vec{q}^{\,(B-1)},\vec{q}^{\,(B-2)},\ldots,\vec{q}^{\,(0)}$ sequentially. Consider the following three mutually exclusive cases which will be investigated in the following three subsections respectively:

### A. Case $i=k-1,k-2,\ldots,B$

In this case, we have the following fact due to (22) and $\mathbf{G}=[\mathbf{I}_k\ \mathbf{P}]$:

$$\begin{aligned}\mathbf{G}_B&=\left[\begin{array}{cccc:c}\mathbf{g}_B & \mathbf{g}_{B+1} & \cdots & \mathbf{g}_{n-1} & \mathbf{0}^{k\times N}\end{array}\right]\\ &=\left[\begin{array}{c:c:c}\begin{array}{c}\mathbf{0}^{B\times(k-B)}\\ \hdashline \mathbf{I}_{k-B}\end{array} & \mathbf{P} & \mathbf{0}^{k\times N}\end{array}\right].\end{aligned}\tag{29}$$

Since

$$\mathbf{I}_{k-B}^{(k)}\,\mathbf{G}_B=\left[\begin{array}{ccc}\multicolumn{3}{c}{\mathbf{0}^{B\times(T+1)}}\\ \hdashline \mathbf{I}_{k-B} & \mathbf{V}^{(k-B)\times B} & \mathbf{0}^{(k-B)\times N}\end{array}\right]\tag{30}$$

by (29) (recall the definition of $\mathbf{P}$ in (28) and the definition of $\mathbf{I}_{k-B}^{(k)}$ in (20)) and any $(k-B)$ columns of $[\mathbf{I}_{k-B}\ \mathbf{V}^{(k-B)\times B}]\in\mathbb{F}^{(k-B)\times k}$ are independent due to the property of systematic MDS codes, it follows that

$$\text{space}\big(\mathbf{I}_{k-B}^{(k)}\,\mathbf{G}_B\,\mathbf{E}_{\varepsilon^{T+1}}\big)=\text{space}\big(\mathbf{I}_{k-B}^{(k)}\,\mathbf{G}_B\big)\tag{31}$$

for any $\varepsilon^{T+1}$ with $B$ positions of 1's (multiplying $E_{\varepsilon^{T+1}}$ on the right side of a matrix has the effect of zeroing $B$ columns of the multiplied matrix). Combining (31) and (30), we conclude that (23) holds for all $i\in\{k-1,k-2,\ldots,B\}$.

### *B. Case $i = B-1, B-2, \ldots, B-N$*

We will choose $\vec{q}^{\,(i)}$ in a recursive manner for $i = B-1, B-2, \ldots, B-N$. Suppose $i = B-j$ for some $j \in \{1, 2, \ldots, N\}$. Assume $\vec{q}^{\,(B-1)}, \vec{q}^{\,(B-2)}, \ldots, \vec{q}^{\,(B-j+1)}$ have been chosen such that (23) holds for $i = B-1, B-2, \ldots, B-j+1$ for any $\varepsilon^{T+1} \in \boldsymbol{\Omega}_{B,N}^{T+1}$. Our goal is to choose $\vec{q}^{\,(B-j)}$ such that (23) holds for $i = B-j$. To this end, we first recognize the following fact due to (22) and $\mathbf{G} = [\mathbf{I}_k\ \mathbf{P}]$:

$$\begin{aligned}\mathbf{G}_{B-j} &= \left[\mathbf{g}_{B-j}\ \mathbf{g}_{B-j+1}\ \cdots\ \mathbf{g}_{n-1} \,\middle|\, \mathbf{0}^{k\times(N-j)}\right] \\ &= \left[\begin{array}{c|c|c} \begin{array}{c}\mathbf{0}^{(B-j)\times(k-B+j)} \\ \hline \mathbf{I}_{k-B+j}\end{array} & \mathbf{P} & \mathbf{0}^{k\times(N-j)} \end{array}\right]. \end{aligned} \tag{32}$$

Using (32), (27) and (28), we obtain

$$\mathbf{I}_{k-B+j}^{(k)}\,\mathbf{G}_{B-j} = \left[\begin{array}{c|c} \multicolumn{2}{c}{\mathbf{0}^{(B-j)\times(T+1)}} \\ \hline \mathbf{K}^{(k-B+j)\times(k+j)} & \mathbf{0}^{(k-B+j)\times(N-j)} \end{array}\right] \tag{33}$$

where

$$\mathbf{K}^{(k-B+j)\times(k+j)} \triangleq \left[\begin{array}{c|c|c} & \mathbf{0}^{j\times(B-N)} & \begin{matrix}\vec{q}^{\,(B-j)} \\ \vdots \\ \vec{q}^{\,(B-1)}\end{matrix} \\ \cline{2-3} \mathbf{I}_{k-B+j} & \multicolumn{2}{c}{\mathbf{V}^{(k-B)\times B}} \end{array}\right].$$

By definition, we have

$$\mathbf{K}^{(k-B+j)\times(k+j)} = \left[\begin{array}{c|c} 1 & \mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)} \\ \mathbf{0}^{(k-B+j-1)\times 1} & \mathbf{K}^{(k-B+j-1)\times(k+j-1)} \end{array}\right]. \tag{34}$$

Due to the previous case in Section V-A and the assumption in this case, the sufficient condition (23) holds for each $i = k-1, k-2, \ldots, B-j+1$ for any $\varepsilon^{T+1} \in \boldsymbol{\Omega}_{B,N}^{T+1}$, which together with (33) implies that

$$\begin{aligned}\operatorname{rank}\Big(\left[\,\mathbf{K}^{(k-B+j-1)\times(k+j-1)} \,\middle|\, \mathbf{0}^{(k-B+j-1)\times(N-j+1)}\,\right]\mathbf{E}_{\varepsilon^{T+1}}\Big) \\ = k-B+j-1\end{aligned}$$

and hence

$$\operatorname{rank}(\mathbf{K}^{(k-B+j-1)\times(k+j-1)}\,\mathbf{E}_{\varepsilon^{k+j-1}}) = k-B+j-1 \tag{35}$$

for any $\varepsilon^{k+j-1} \in \boldsymbol{\Omega}_{B,N}^{k+j-1}$ (cf. Definition 9). We would like to show the existence of a $\vec{q}^{\,(B-j)} \in \mathbb{F}^N$ such that

$$\mathbf{u}_0^{(k-B+j)} \in \operatorname{space}\big(\mathbf{K}^{(k-B+j)\times(k+j)}\,\mathbf{E}_{\varepsilon^{k+j}}\big) \tag{36}$$

for any $\varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}$, which together with (33) will then imply that (23) holds for $i = B-j$. Fix an arbitrary $\varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}$ and consider the following three subcases:

*Subcase $\varepsilon_0 = 0$:* Using (34) and $\varepsilon_0 = 0$, we conclude that the first column of $\mathbf{K}^{(k-B+j)\times(k+j)}\,\mathbf{E}_{\varepsilon^{k+j}}$ is $\mathbf{u}_0^{(k-B+j)}$, which together with (34) implies that (36) holds for any choice of $\vec{q}^{\,(B-j)}$.

*Subcase $\varepsilon_0 = 1$ and $\sum_{\ell=0}^{k+j-1}\varepsilon_\ell = B$ With All the 1's in $\varepsilon^{k+j}$ Occupying Consecutive Positions:* In this case, $\varepsilon^{k+j}$ equals $(\underbrace{1, \ldots, 1}_{B \text{ times}}, 0, \ldots, 0)$, and

$$\mathbf{K}^{(k-B+j-1)\times(k+j-1)}\mathbf{E}_{(\varepsilon_1,\ldots,\varepsilon_{k+j-1})}$$

consists of exactly $(B-1)$ zero column vectors and $k+j-1-(B-1) = k-B+j$ non-zero column vectors, and the non-zero column vectors are denoted by $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{k-B+j}$. In addition,

$$\operatorname{rank}\big(\left[\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_{k-B+j}\right]\big) = k-B+j-1$$

by (35) (due to our induction hypothesis), which implies that there exists a non-zero vector

$$\boldsymbol{\lambda} \triangleq [\lambda_1\ \lambda_2\ \cdots\ \lambda_{k-B+j}]^t \in \mathbb{F}^{(k-B+j)\times 1}$$

such that

$$\left[\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_{k-B+j}\right]\boldsymbol{\lambda} = \mathbf{0}^{(k-B+j-1)\times 1}. \tag{37}$$

Since $\left[\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_{k-B+j}\right]$ contains a $(k-B)\times(k-B+j)$ submatrix of $\mathbf{V}^{(k-B)\times B}$ (which is the parity matrix of some MDS code) where any $k-B$ columns of $\mathbf{V}^{(k-B)\times B}$ are independent, it follows that any $k-B$ columns of $\left[\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_{k-B+j}\right]$ are independent, which implies from (37) that $\boldsymbol{\lambda}$ contains at least $k-B+1$ non-zero elements. Consequently, it follows from (37) that there exists a non-zero vector

$$\boldsymbol{\rho} \triangleq [\mathbf{0}^{1\times(B-1)}\ \boldsymbol{\lambda}]^t \in \mathbb{F}^{(k+j-1)\times 1}$$

which contains at least $k-B+1$ non-zero elements such that

$$\begin{aligned}&\begin{bmatrix}\mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)} \\ \mathbf{K}^{(k-B+j-1)\times(k+j-1)}\end{bmatrix}\mathbf{E}_{(\varepsilon_1,\ldots,\varepsilon_{k+j-1})}\boldsymbol{\rho} \\ &= \begin{bmatrix}\left[\mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)}\right]\mathbf{E}_{(\varepsilon_1,\ldots,\varepsilon_{k+j-1})}\boldsymbol{\rho} \\ \mathbf{0}^{(k-B+j-1)\times 1}\end{bmatrix} \\ &= \begin{bmatrix}\left[\mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)}\right]\boldsymbol{\rho} \\ \mathbf{0}^{(k-B+j-1)\times 1}\end{bmatrix}.\end{aligned} \tag{38}$$

Using the fact that $\vec{q}^{\,(B-j)}$ is a length-$N$ variable vector and $\boldsymbol{\rho}$ contains at least $k-B+1$ non-zero elements, we claim that $\left[\mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)}\right]\boldsymbol{\rho}$ is a non-zero linear function of $(q_0^{(B-j)}, q_1^{(B-j)}, \ldots, q_{N-1}^{(B-j)})$, and we let $\psi_{\varepsilon^{k+j}}^{(B-j)}(\vec{q}^{\,(B-j)})$ denote the non-zero linear function. To see the above claim, we can assume the contrary that $\left[\mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)}\right]\boldsymbol{\rho} \equiv 0$, which implies $\boldsymbol{\lambda}$ contains at least $N$ zeros, which together with the fact that $\boldsymbol{\lambda}$ contains at least $k-B+1$ non-zero elements leads to the conclusion that $\boldsymbol{\lambda}$ contains at least $k-B+1+N > k-B+j$ elements, contradicting that fact that the length of $\boldsymbol{\lambda}$ is $k-B+j$. Combining (34) and (38), we conclude that (36) holds as long as $\vec{q}^{\,(B-j)}$ satisfies $\psi_{\varepsilon^{k+j}}^{(B-j)}(\vec{q}^{\,(B-j)}) \neq 0$.

*Subcase $\varepsilon_0 = 1$ and $\sum_{\ell=0}^{k+j-1} e_\ell = N$ With No Restriction on the Positions of 1's in $\varepsilon^{k+j}$:* In this case,

$$\mathbf{K}^{(k-B+j-1)\times(k+j-1)}\,\mathbf{E}_{(\varepsilon_1,\ldots,\varepsilon_{k+j-1})}$$

consists of exactly $(N-1)$ zero column vectors and $k+j-1-(N-1) = k+j-N$ non-zero column vectors, which we denote as $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{k+j-N} \in \mathbb{F}^{k-B+j-1}$. Construct the following $(k-B+j)\times(k+j-N)$ submatrix of $\mathbf{K}^{(k-B+j)\times(k+j)}$:

$$\mathbf{A} \triangleq \begin{bmatrix} p_1 & p_2 & \cdots & p_{k+j-N} \\ \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_{k+j-N}\end{bmatrix} \tag{39}$$

for some $[p_1\ p_2\ \cdots\ p_{k+j-N}] \in \mathbb{F}^{k+j-N}$ which is a subvector of $\left[\mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)}\right]$. Since the length of

$[p_1\ p_2\ \cdots\ p_{k+j-N}]$ is strictly larger than the number of zeros in $\left[\mathbf{0}^{1\times(k+j-N-1)}\ \vec{q}^{\,(B-j)}\right]$, there exists an $r \in \{1, 2, \ldots, k+j-N\}$ such that $p_r \neq 0$ where $p_r$ is an element of $\vec{q}^{\,(B-j)}$. Since

$$\operatorname{rank}\left(\left[\mathbf{h}_1\ \cdots\ \mathbf{h}_{r-1}\ \mathbf{h}_{r+1}\ \cdots\ \mathbf{h}_{k+j-N}\right]\right) = k - B + j - 1$$

by (35) (due to the induction hypothesis), we have

$$\mathbf{h}_r \in \operatorname{space}\left(\left[\mathbf{h}_1\ \cdots\ \mathbf{h}_{r-1}\ \mathbf{h}_{r+1}\ \cdots\ \mathbf{h}_{k+j-N}\right]\right),$$

which implies that there exists a non-zero vector

$$\boldsymbol{\lambda} \triangleq [\lambda_1\ \ldots \lambda_{r-1}\ 1\ \lambda_{r+1}\ \ldots\ \lambda_{k+j-N}]^t \in \mathbb{F}^{(k+j-N)\times 1}$$

such that

$$\left[\mathbf{h}_1\ \cdots\ \mathbf{h}_{r-1}\ \mathbf{h}_r\ \mathbf{h}_{r+1}\ \cdots\ \mathbf{h}_{k+j-N}\right]\boldsymbol{\lambda} = \mathbf{0}^{(k-B+j-1)\times 1},$$

which together with (39) and the fact $p_r \neq 0$ implies that

$$\mathbf{A}\boldsymbol{\lambda} = \begin{bmatrix} \varpi_{\varepsilon^{k+j}}^{(B-j)}(\vec{q}^{\,(B-j)}) \\ \mathbf{0}^{(k-B+j-1)\times 1} \end{bmatrix} \tag{40}$$

for some non-zero linear function of $(q_0^{(B-j)}, q_1^{(B-j)}, \ldots, q_{N-1}^{(B-j)})$ denoted by $\varpi_{\varepsilon^{k+j}}^{(B-j)}(\vec{q}^{\,(B-j)})$. Using (34), (39), the fact that $\mathbf{A}$ consists of columns of $\mathbf{K}^{(k-B+j)\times(k+j)}$ and (40), we conclude that (36) holds as long as $\vec{q}^{\,(B-j)}$ satisfies $\varpi_{\varepsilon^{k+j}}^{(B-j)}(\vec{q}^{\,(B-j)}) \neq 0$.

Combining the above three subcases, we see that for any $\varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}$, statement (36) holds for all $i \in \{B-1, B-2, \ldots, B-N\}$ as long as $\vec{q}^{\,(i)} \in \mathbb{F}^N$ satisfies $\psi_{\varepsilon^{k+j}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ and $\varpi_{\varepsilon^{k+j}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$. Since the number of $\underline{\vec{q}}^{\,(i)} \in \mathbb{F}^N$ that satisfies either $\psi_{\varepsilon^{k+j}}^{(i)}(\underline{\vec{q}}^{\,(i)}) = 0$ or $\varpi_{\varepsilon^{k+j}}^{(i)}(\underline{\vec{q}^{\,(i)}}) = 0$ is less than $2|\mathbb{F}|^{N-1}$ for each $i$ and each $\varepsilon^{k+j}$ and

$$\left|\boldsymbol{\Omega}_{B,N}^{k+j}\right| \leq \binom{T+1}{N} + T - B + 2,$$

the hypothesis (14) guarantees the following: For each $i = B-1, B-2, \ldots, B-N$ where the vectors $\vec{q}^{\,(B-1)}, \ldots, \vec{q}^{\,(i+1)}$ have been chosen, we can always choose a $\vec{q}^{\,(i)} \in \mathbb{F}^N$ such that $\psi_{\varepsilon^{k+j}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ and $\varpi_{\varepsilon^{k+j}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ for all $\varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}$ because

$$\frac{\left|\left\{\underline{\vec{q}}^{\,(i)} \in \mathbb{F}^N \,\middle|\, \begin{array}{l}\psi_{\varepsilon^{k+j}}^{(i)}(\underline{\vec{q}}^{\,(i)}) = 0 \text{ or } \varpi_{\varepsilon^{k+j}}^{(i)}(\underline{\vec{q}}^{\,(i)}) = 0 \\ \text{for some } \varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}\end{array}\right\}\right|}{\text{total number of } \underline{\vec{q}}^{\,(i)}}$$
$$\leq \frac{2\left(\binom{T+1}{N} + T - B + 2\right)|\mathbb{F}|^{N-1}}{|\mathbb{F}|^N}$$
$$< 1.$$

By induction, there exist $\vec{q}^{\,(B-1)}, \ldots, \vec{q}^{\,(B-N)}$ such that $\psi_{\varepsilon^{k+j}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ and $\varpi_{\varepsilon^{k+j}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ for all $i \in \{B-1, \ldots, B-N\}$ and all $\varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}$. This together with the conclusions made in the above three subcases implies that statement (36) holds for all $i \in \{B-1, B-2, \ldots, B-N\}$ and all $\varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}$, which together with (33) implies that (23) holds for all $i \in \{B-1, B-2, \ldots, B-N\}$ and all $\varepsilon^{k+j} \in \boldsymbol{\Omega}_{B,N}^{k+j}$.

### C. Case $i = B-N-1, B-N-2, \ldots, 0$

Suppose $\vec{q}^{\,(B-1)}, \vec{q}^{\,(B-2)}, \ldots, \vec{q}^{\,(B-N)}$ have been chosen in the previous subcase. We will choose $\vec{q}^{\,(i)}$ in a recursive manner for $i = B-N-1, B-N-2, \ldots, 0$. Suppose $i = B-N-j$ for some $j \in \{1, 2, \ldots, B-N\}$. Assume $\vec{q}^{\,(B-N-1)}, \vec{q}^{\,(B-N-2)}, \ldots, \vec{q}^{\,(B-N-j+1)}$ have been chosen such that (23) holds for $i = B-N-1, B-N-2, \ldots, B-N-j+1$ for any $\varepsilon^{T+1} \in \boldsymbol{\Omega}_{B,N}^{T+1}$. Our goal is to choose $\vec{q}^{\,(B-N-j)}$ such that (23) holds for $i = B-N-j$. To this end, we first use the first clause in (27), (28) and the fact $\mathbf{G} = [\mathbf{I}_k\ \mathbf{P}]$ to obtain

$$\mathbf{I}_{k-B+N+j}^{(k)}\,\mathbf{G}_{B-N-j} = \begin{bmatrix}\mathbf{0}^{(B-N-j)\times(T+1)} \\ \mathbf{J}^{(k-B+N+j)\times(T+1)}\end{bmatrix} \tag{41}$$

where $\mathbf{J}^{(k-B+N+j)\times(T+1)}$ is defined in (42) as shown at the bottom of this page with $\mathbf{V}^{(k-B)\times(B-j)}$ being the matrix consisting of the first $B-j$ columns of $\mathbf{V}^{(k-B)\times B}$. We would like to show the existence of a $\vec{q}^{\,(B-N-j)} \in \mathbb{F}^N$ such that

$$\mathbf{u}_0^{(k-B+N+j)} \in \operatorname{space}\left(\mathbf{J}^{(k-B+N+j)\times(T+1)}\,\mathbf{E}_{\varepsilon^{T+1}}\right) \tag{43}$$

for any $\varepsilon^{T+1} \in \boldsymbol{\Omega}_{B,N}^{T+1}$, which together with (41) will then imply that (23) holds for $i = B-N-j$. Fix an arbitrary $\varepsilon^{T+1} \in \boldsymbol{\Omega}_{B,N}^{T+1}$ and consider the following three subcases:

*Subcase $\varepsilon_0 = 0$:* Using (42) and $\varepsilon_0 = 0$, we conclude that the first column of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\,\mathbf{E}_{\varepsilon^{T+1}}$ is $\mathbf{u}_0^{(k-B+N+j)}$,

$$\mathbf{J}^{(k-B+N+j)\times(T+1)} \triangleq \left[\begin{array}{c:ccccccc}
 & \mathbf{0}^{1\times(B-N-j)} & q_0^{(B-N-j)} & q_1^{(B-N-j)} & \cdots & \cdots & \cdots & q_{N-1}^{(B-N-j)} \\
 & & \mathbf{0}^{1\times(B-N-j+1)} & q_0^{(B-N-j+1)} & q_1^{(B-N-j+1)} & \cdots & \cdots & q_{N-2}^{(B-N-j+1)} \\
 & & \ddots & \ddots & \ddots & & \vdots & \vdots \\
\mathbf{I}_{k-B+N+j} & & \mathbf{0}^{1\times(B-N)} & & q_0^{(B-N)} & q_1^{(B-N)} & \cdots & q_{N-j-1}^{(B-N)} \\
 & & \vdots & & \vdots & \vdots & \cdots & \vdots \\
 & & \mathbf{0}^{1\times(B-N)} & & q_0^{(B-1)} & q_1^{(B-1)} & \cdots & q_{N-j-1}^{(B-1)} \\ \hdashline
 & & & & \mathbf{V}^{(k-B)\times(B-j)} & & &
\end{array}\right] \tag{42}$$

which together with (42) implies that (43) holds for any choice of $\vec{q}^{\,(B-N-j)}$.

*Subcase $\varepsilon_0 = 1$ and $\sum_{\ell=0}^{T} \varepsilon_\ell = B$ With All the 1's in $\varepsilon^{T+1}$ Occupying Consecutive Positions:* In this case, $\varepsilon^{T+1} = (\underbrace{1, \ldots, 1}_{B \text{ times}}, 0, \ldots, 0)$ and $j \leq B - N$. Using (42) and the fact that $(k-B+N+j)-B \leq k-B$, we see that the first $k-B+1$ non-zero columns of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\mathbf{E}_{\varepsilon^{T+1}}$ equals

$$\left[\begin{array}{c:c} \mathbf{0}^{(N+j)\times(k-B)} & \begin{matrix} q_0^{(B-N-j)} \\ \mathbf{0}^{(N+j-1)\times 1} \end{matrix} \\ \hdashline \multicolumn{2}{c}{\mathbf{V}^*} \end{array}\right] \tag{44}$$

where $\mathbf{V}^*$ is a $(k-B)\times(k-B+1)$ submatrix of $[\mathbf{I}_{k-B}\ \ \mathbf{V}^{(k-B)\times B}]$. Since any $(k-B)$ columns of $\mathbf{V}^*$ are independent due to the property of systematic MDS matrices, it follows from (44) that

$$\begin{bmatrix} q_0^{(B-N-j)} \\ \mathbf{0}^{(k-B+N+j-1)\times 1} \end{bmatrix} \in \text{space}\left(\mathbf{J}^{(k-B+N+j)\times(T+1)}\mathbf{E}_{\varepsilon^{T+1}}\right),$$

which implies that (43) holds for any choice of $\vec{q}^{\,(B-N-j)}$ that satisfies $q_0^{(B-N-j)} \neq 0$.

*Subcase $\varepsilon_0 = 1$ and $\sum_{\ell=0}^{T} \varepsilon_\ell = N$ With No Restriction on the Positions of 1's in $\varepsilon^{T+1}$:* In this case,

$$\mathbf{J}^{(k-B+N+j)\times(T+1)}\,\mathbf{E}_{\varepsilon^{T+1}}$$

consists of exactly $N$ zero column vectors and $k-B+j$ non-zero column vectors. Consider

$$\begin{aligned} \mathbf{Q} &\triangleq \left[\begin{array}{c:c} \mathbf{0}^{1\times(k-B+N+j-1)} & \vec{q}^{\,(B-N-j)} \\ \mathbf{I}_{k-B+N+j-1} & \mathbf{D} \end{array}\right] \\ &\in \mathbb{F}^{(k-B+N+j)\times(k-B+2N+j-1)} \end{aligned} \tag{45}$$

which consists of the second to the $(k-B+N+j)^{\text{th}}$ columns and the last $N$ columns of $\mathbf{J}^{(k-B+N+j)\times(T+1)}$ where $\mathbf{D}$ is some $(k-B+N+j-1)\times N$ matrix that is determined by (42). Since $\varepsilon_0 = 1$ and $\sum_{\ell=1}^{T}\varepsilon_\ell = N-1$, there exists a $(k-B+N+j)\times(k-B+N+j)$ submatrix of $\mathbf{Q}$ denoted by

$$\mathbf{B} \triangleq \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_{k-B+N+j} \end{bmatrix} \tag{46}$$

such that $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{k-B+N+j}$ are non-zero columns of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\,\mathbf{E}_{\varepsilon^{T+1}}$. By (46) and the linear dependence among the column vectors of the lower $(k-B+N+j-1)\times(k-B+N+j)$ submatrix of $\mathbf{B}$, there exist a non-zero vector

$$\boldsymbol{\lambda} \triangleq [\lambda_1\ \lambda_2\ \ldots\ \lambda_{k-B+N+j}]^t \in \mathbb{F}^{(k-B+N+j)\times 1}$$

and a linear function of $(q_0^{(B-N-j)}, q_1^{(B-N-j)}, \ldots, q_{N-1}^{(B-N-j)})$ denoted by $\chi_{\varepsilon^{T+1}}^{(B-N-j)}(\vec{q}^{\,(B-N-j)})$ such that

$$\mathbf{B}\boldsymbol{\lambda} = \begin{bmatrix} \chi_{\varepsilon^{T+1}}^{(B-N-j)}(\vec{q}^{\,(B-N-j)}) \\ \mathbf{0}^{(k-B+N+j-1)\times 1} \end{bmatrix}. \tag{47}$$

In addition, we claim that $\chi_{\varepsilon^{T+1}}^{(B-N-j)}(\vec{q}^{\,(B-N-j)})$ is a non-zero function. To see this claim, we can assume the contrary that $\chi_{\varepsilon^{T+1}}^{(B-N-j)}(\vec{q}^{\,(B-N-j)}) \equiv 0$, which would imply that $\mathbf{B}\boldsymbol{\lambda} \equiv \mathbf{0}$ with $\boldsymbol{\lambda} \neq \mathbf{0}^{(k-B+N+j)\times 1}$, which together with (45) and (46) would imply

$$\left[\begin{array}{c:c} \mathbf{0}^{1\times(k-B+N+j-1)} & \vec{q}^{\,(B-N-j)} \\ \mathbf{I}_{k-B+N+j-1} & \mathbf{D} \end{array}\right]\boldsymbol{\lambda}' = \mathbf{0}^{(k-B+N+j)\times 1}$$

for some $\boldsymbol{\lambda}' \neq \mathbf{0}^{(k-B+2N+j-1)\times 1}$, which together with the fact that $\vec{q}^{\,(B-N-j)}$ does not contain any zero would imply the contradiction that

$$\begin{bmatrix} \mathbf{0}^{1\times(k-B+N+j-1)} \\ \mathbf{I}_{k-B+N+j-1} \end{bmatrix}\boldsymbol{\lambda}^* = \mathbf{0}^{(k-B+N+j)\times 1} \tag{48}$$

for some $\boldsymbol{\lambda}^* \neq \mathbf{0}^{(k-B+N+j-1)\times 1}$. Using (46), the fact that $\mathbf{B}$ consists of columns of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\,\mathbf{E}_{\varepsilon^{T+1}}$ and (47), we conclude that (43) holds as long as $\vec{q}^{\,(B-N-j)}$ satisfies $\chi_{\varepsilon^{T+1}}^{(B-j)}(\vec{q}^{\,(B-j)}) \neq 0$.

Combining the above three subcases, we see that for any $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$, statement (43) holds for all $i \in \{B-N-1, B-N-2, \ldots, 0\}$ as long as $\vec{q}^{\,(i)} \in \mathbb{F}^N$ satisfies $q_0^{(i)} \neq 0$ and $\chi_{\varepsilon^{T+1}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$. Since the number of $\underline{\vec{q}}^{\,(i)} \in \mathbb{F}^N$ that satisfies either $\underline{q}_0^{(i)} = 0$ or $\chi_{\varepsilon^{T+1}}^{(i)}(\underline{\vec{q}}^{\,(i)}) = 0$ is less than $2|\mathbb{F}|^{N-1}$ for each $i$ and each $\varepsilon^{T+1}$ and

$$\left|\Omega_{B,N}^{T+1}\right| \leq \binom{T+1}{N} + T - B + 2,$$

the hypothesis (14) guarantees the following: For each $i = B-N-1, B-N-2, \ldots, 0$ where the vectors $\vec{q}^{\,(B-1)}, \ldots, \vec{q}^{\,(i+1)}$ have been chosen, we can always choose a $\vec{q}^{\,(i)} \in \mathbb{F}^N$ such that $q_0^{(i)} \neq 0$ and $\chi_{\varepsilon^{T+1}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ for all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$ because

$$\begin{aligned} &\frac{\left|\left\{\underline{\vec{q}}^{\,(i)} \in \mathbb{F}^N \,\middle|\, \begin{array}{l} \underline{q}_0^{(i)} = 0 \text{ or } \chi_{\varepsilon^{T+1}}^{(i)}(\underline{\vec{q}}^{\,(i)}) = 0 \text{ for some} \\ \varepsilon^{T+1} \in \Omega_{B,N}^{T+1} \end{array}\right\}\right|}{\text{total number of } \underline{\vec{q}}^{\,(i)}} \\ &\leq \frac{2\left(\binom{T+1}{N} + T - B + 2\right)|\mathbb{F}|^{N-1}}{|\mathbb{F}|^N} \\ &< 1. \end{aligned}$$

By induction, there exist $\vec{q}^{\,(B-N-1)}, \ldots, \vec{q}^{\,(0)}$ such that $q_0^{(i)} \neq 0$ and $\chi_{\varepsilon^{T+1}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ for all $i \in \{B-1, \ldots, B-N\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$. This together with the conclusions made in the above three subcases implies that statement (43) holds for all $i \in \{B-N-1, B-N-2, \ldots, 0\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$, which together with (41) implies that (23) holds for all $i \in \{B-N-1, B-N-2, \ldots, 0\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$.

### D. Combining the Three Cases

Combining the three cases studied in the preceding three subsections, we conclude that there exist $\vec{q}^{\,(B-1)}, \vec{q}^{\,(B-2)}, \ldots, \vec{q}^{\,(0)}$ such that $\mathbf{G} = [\mathbf{I}_k\ \mathbf{P}]$ with $\mathbf{P}$ having the form (28) satisfies (23) for all $i \in \{k-1, k-2, \ldots, 0\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$. In particular, $\mathbf{P}$ has the form (25).

## VI. Proof of Lemma 4

Fix any $(W, T, B, N)$ that satisfies (7) and recall that $k = T - N + 1$. Suppose $k < B$. Fix any finite field $\mathbb{F}$ that satisfies (14). Our goal is to show that $\mathbf{G} = [\,\mathbf{I}_k\ \mathbf{P}]$ satisfies (23) for some $\mathbf{P}$ having the form (26), where $\mathbf{I}_{k-i}^{(k)}$, $\mathbf{G}_i$ and $\mathbf{E}_{\varepsilon^{T+1}}$ in (23) are as defined in (20), (22) and (21) respectively. To this end, we construct a variable vector

$$\vec{q}^{\,(i)} \triangleq [q_0^{(i)}\ q_1^{(i)}\ \ldots\ q_{N-1}^{(i)}] \in \mathbb{F}^N$$

for each $i \in \{0, 1, \ldots B - N - 1\}$ where the values of the $B - N$ vectors will be determined later in this proof. In addition, we let

$$\vec{q}_{\text{left}}^{\,(i)} \triangleq [q_0^{(i)}\ q_1^{(i)}\ \ldots\ q_{B-k-1}^{(i)}]$$

and

$$\vec{q}_{\text{right}}^{\,(i)} \triangleq [q_{B-k}^{(i)}\ q_{B-k+1}^{(i)}\ \ldots\ q_{N-1}^{(i)}]$$

such that $\vec{q}^{\,(i)} \triangleq [\vec{q}_{\text{left}}^{\,(i)}\ \vec{q}_{\text{right}}^{\,(i)}]$, and define $\vec{p}^{\,(i)} \in \mathbb{F}^B$ as

$$\vec{p}^{\,(i)} \triangleq [\vec{q}_{\text{left}}^{\,(i)}\ \mathbf{0}^{1\times i}\ \vec{q}_{\text{right}}^{\,(i)}\ \mathbf{0}^{1\times(B-N-i)}] \tag{49}$$

for each $i \in \{0, 1, \ldots, B - N - 1\}$. Construct a $(k - B + N) \times N$ parity matrix of a systematic MDS $(k - B + 2N,\ k-B+N)$-code denoted by $\mathbf{V}^{(k-B+N)\times N}$, which always exists because $|\mathbb{F}| \ge 2(2T - B + 1) \ge k - B + 2N$ by (14). Let $\mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)}$ be the $(k - B + N) \times (B - k)$ matrix formed by collecting the first $B - k$ columns of $\mathbf{V}^{(k-B+N)\times N}$ and let $\mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)}$ be the $(k - B + N) \times (k - B + N)$ matrix formed by collecting the last $k - B + N$ columns of $\mathbf{V}^{(k-B+N)\times N}$ such that

$$\left[\mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)}\ \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)}\right] = \mathbf{V}^{(k-B+N)\times N}.$$

Then, let

$$\mathbf{P} \triangleq \left[\begin{array}{ccc} & \vec{p}^{\,(0)} & \\ & \vdots & \\ & \vec{p}^{\,(B-N-1)} & \\ \hdashline \mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)} & \mathbf{0} & \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)} \end{array}\right] \tag{50}$$

where $\mathbf{0}$ is the $(k-B+N)\times(B-N)$ zero matrix. It can be seen that $\mathbf{P}$ has the form (26). It remains to show that $\mathbf{G} = [\mathbf{I}_k\ \mathbf{P}]$ satisfies (23) for $i = k-1, k-2, \ldots, 0$ for some $\{\vec{q}^{\,(i)}\}_{i=0}^{B-N-1}$. By inspecting (23) and (50), we see that condition (23) depends on only $\{\vec{q}^{\,(B-N-1-j)}\}_{j=0}^{i}$ for each $i \in \{B - N - 1,\ B - N - 2, \ldots, 0\}$ and does not depend on $\{\vec{q}^{\,(i)}\}_{i=0}^{B-N-1}$ for each $i \in \{k-1, k-2, \ldots, B - N\}$. In the rest of the proof, we will verify condition (23) in the order $i = k-1, k-2, \ldots, 0$, which means that we will choose $\{\vec{q}^{\,(i)}\}_{i=0}^{B-N-1}$ by choosing $\vec{q}^{\,(B-N-1)}, \vec{q}^{\,(B-N-2)}, \ldots, \vec{q}^{\,(0)}$ sequentially. Consider the following two mutually exclusive cases which will be investigated in the following two subsections respectively:

### A. Case $i = k - 1, k - 2, \ldots, B - N$

In this case, we have the following fact due to (22) and $\mathbf{G} = [\mathbf{I}_k\ \mathbf{P}]$:

$$\begin{aligned}\mathbf{G}_{B-N} &= \left[\,\mathbf{g}_{B-N}\quad \mathbf{g}_{B-N+1}\quad \ldots\quad \mathbf{g}_{n-1}\,\right] \\ &= \left[\begin{array}{c:c} \mathbf{0}^{(B-N)\times(k-B+N)} & \\ \hdashline \mathbf{I}_{k-B+N} & \mathbf{P} \end{array}\right].\end{aligned} \tag{51}$$

Using (51), the definition of $\mathbf{P}$ in (50) and the definition of $\mathbf{I}_{k-B+N}^{(k)}$ in (20), we have

$$\begin{aligned}&\mathbf{I}_{k-B+N}^{(k)}\,\mathbf{G}_{B-N} \\ &\quad= \left[\begin{array}{cccc} & \mathbf{0}^{(B-N)\times(T+1)} & & \\ \hdashline \mathbf{I}_{k-B+N} & \mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)} & \mathbf{0} & \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)} \end{array}\right].\end{aligned} \tag{52}$$

Since any $(k - B + N)$ columns of

$$\left[\mathbf{I}_{k-B+N}\ \mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)}\ \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)}\right] \in \mathbb{F}^{(k-B+N)\times(k-B+2N)}$$

are independent due to the property of systematic MDS codes and

$$\left[\begin{array}{cccc} & \mathbf{0}^{(B-N)\times(T+1)} & & \\ \hdashline \mathbf{I}_{k-B+N} & \mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)} & \mathbf{0} & \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)} \end{array}\right]\mathbf{E}_{\varepsilon^{T+1}}$$

contains at least $(k-B+N)$ non-zero columns for any $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$ with $N$ arbitrary positions of 1 and exactly $(k-B+N)$ non-zero columns for any $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$ with $B$ consecutive positions of 1, it follows from (52) that

$$\text{space}\left(\mathbf{I}_{k-B+N}^{(k)}\,\mathbf{G}_{B-N}\mathbf{E}_{\varepsilon^{T+1}}\right) = \text{space}\left(\begin{bmatrix} \mathbf{0}^{(B-N)\times(k-B+N)} \\ \mathbf{I}_{k-B+N} \end{bmatrix}\right)$$

for any $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$, which then implies (23) for all $i \in \{k - 1, k - 2, \ldots, B - N\}$.

### B. Case $i = B - N - 1, B - N - 2, \ldots, 0$

We will choose $\vec{q}^{\,(i)}$ in a recursive manner for $i = B - N - 1, B - N - 2, \ldots, 0$. Suppose $i = B - N - j$ for some $j \in \{1, 2, \ldots, B - N\}$. Assume $\vec{q}^{\,(B-N-1)}, \vec{q}^{\,(B-N-2)}, \ldots, \vec{q}^{\,(B-N-j+1)}$ have been chosen such that (23) holds for $i = B-N-1, B-N-2, \ldots, B-N-j+1$ for any $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$. Our goal is to choose $\vec{q}^{\,(B-N-j)}$ such that (23) holds for $i = B - N - j$. To this end, we first use (49), (50) and the fact $\mathbf{G} = [\mathbf{I}_k\ \mathbf{P}]$ to obtain

$$\mathbf{I}_{k-B+N+j}^{(k)}\,\mathbf{G}_{B-N-j} = \begin{bmatrix} \mathbf{0}^{(B-N-j)\times(T+1)} \\ \mathbf{J}^{(k-B+N+j)\times(T+1)} \end{bmatrix} \tag{53}$$

where $\mathbf{J}^{(k-B+N+j)\times(T+1)}$ is defined in (54) as shown at the top of the next page with $\mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N-j)}$ being the matrix consisting of the first $k - B + N - j$ columns of $\mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N)}$. We would like to show the existence of a $\vec{q}^{\,(B-N-j)} \in \mathbb{F}^N$ such that

$$\mathbf{u}_0^{(k-B+N+j)} \in \text{space}\big(\mathbf{J}^{(k-B+N+j)\times(T+1)}\,\mathbf{E}_{\varepsilon^{T+1}}\big) \tag{55}$$

$$\mathbf{J}^{(k-B+N+j)\times(T+1)} \triangleq \left[\begin{array}{c:c:cccccc} \mathbf{I}_{k-B+N+j} & \vec{q}_{\text{left}}^{\,(B-N-j)} & \mathbf{0}^{1\times(B-N-j)} & q_{B-k}^{(B-N-j)} & \cdots & \cdots & \cdots & q_{N-1}^{(B-N-j)} \\ & \vdots & & \ddots & & \ddots & \vdots & \vdots \\ & \vec{q}_{\text{left}}^{\,(B-N-1)} & & \mathbf{0}^{1\times(B-N-1)} & & q_{B-k}^{(B-N-1)} & \cdots & q_{N-j}^{(B-N-1)} \\ \hdashline & \mathbf{V}_{\text{left}}^{(k-B+N)\times(B-k)} & & \mathbf{0}^{(k-B+N)\times(B-N)} & & & \mathbf{V}_{\text{right}}^{(k-B+N)\times(k-B+N-j)} & \end{array}\right] \tag{54}$$

for any $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$, which together with (53) will then imply that (23) holds for $i = B - N - j$. Fix an arbitrary $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$ and consider the following three subcases:

*Subcase $\varepsilon_0 = 0$:* Using (42) and $\varepsilon_0 = 0$, we conclude that the first column of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\mathbf{E}_{\varepsilon^{T+1}}$ is $\mathbf{u}_0^{(k-B+N+j)}$, which together with (54) implies that (55) holds for any choice of $\vec{q}^{\,(B-N-j)}$.

*Subcase $\varepsilon_0 = 1$ and $\sum_{\ell=0}^{T}\varepsilon_\ell = B$ With All the 1's in $\varepsilon^{T+1}$ Occupying Consecutive Positions:* In this case, $\varepsilon^{T+1}$ equals $(\underbrace{1,\ldots,1}_{B\text{ times}},0,\ldots,0)$ and $B \geq N + j$. Therefore, it follows from (54) that the first non-zero column of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\mathbf{E}_{\varepsilon^{T+1}}$ equals

$$\begin{bmatrix} q_0^{(B-N-j)} \\ \mathbf{0}^{(k-B+N+j-1)\times 1} \end{bmatrix},$$

which implies that (55) holds for any choice of $\vec{q}^{\,(B-N-j)}$ that satisfies $q_0^{(B-N-j)} \neq 0$.

*Subcase $\varepsilon_0 = 1$ and $\sum_{\ell=0}^{T}\varepsilon_\ell = N$ With No Restriction on the Positions of 1's in $\varepsilon^{T+1}$:* In this case,

$$\mathbf{J}^{(k-B+N+j)\times(T+1)}\mathbf{E}_{\varepsilon^{T+1}}$$

has at least $k - B + N + j$ non-zero column vectors. Consider

$$\mathbf{Q} \triangleq \left[\begin{array}{c:c} \mathbf{0}^{1\times(k-B+N+j-1)} & \vec{q}^{\,(B-N-j)} \\ \mathbf{I}_{k-B+N+j-1} & \mathbf{D} \end{array}\right] \in \mathbb{F}^{(k-B+N+j)\times(k-B+2N+j-1)}$$

which consists of the second to the $(N + j)$th columns and the last $N - B + k$ columns of $\mathbf{J}^{(k-B+N+j)\times(T+1)}$ where $\mathbf{D}$ is some $(k - B + N + j - 1) \times N$ matrix that is determined by (54). Since $\varepsilon_0 = 1$ and $\sum_{\ell=1}^{T}\varepsilon_\ell = N - 1$, there exists a $(k-B+N+j)\times(k-B+N+j)$ submatrix of $\mathbf{Q}$ denoted by

$$\mathbf{B} \triangleq \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_{k-B+N+j} \end{bmatrix} \tag{56}$$

such that $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{k-B+N+j}$ are non-zero columns of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\mathbf{E}_{\varepsilon^{T+1}}$. By (56) and the linear dependence among the column vectors of the lower $(k - B + N + j - 1) \times (k - B + N + j)$ submatrix of $\mathbf{B}$, there exist a non-zero vector

$$\boldsymbol{\lambda} \triangleq [\lambda_1\ \lambda_2\ \ldots\ \lambda_{k-B+N+j}]^t \in \mathbb{F}^{(k-B+N+j)\times 1}$$

and a linear function of $(q_0^{(B-N-j)}, q_1^{(B-N-j)}, \ldots, q_{N-1}^{(B-N-j)})$ denoted by $\chi_{\varepsilon^{T+1}}^{(B-N-j)}(\vec{q}^{\,(B-N-j)})$ such that

$$\mathbf{B}\boldsymbol{\lambda} = \begin{bmatrix} \chi_{\varepsilon^{T+1}}^{(B-N-j)}(\vec{q}^{\,(B-N-j)}) \\ \mathbf{0}^{(k-B+N+j-1)\times 1} \end{bmatrix}. \tag{57}$$

In addition, $\chi_{\varepsilon^{T+1}}^{(B-N-j)}(\vec{q}^{\,(B-N-j)})$ is a non-zero function by very similar arguments used in the proof of Lemma 4 between (47) and (48). Using (56), the fact that $\mathbf{B}$ consists of columns of $\mathbf{J}^{(k-B+N+j)\times(T+1)}\mathbf{E}_{\varepsilon^{T+1}}$ and (57), we conclude that (55) holds as long as $\vec{q}^{\,(B-N-j)}$ satisfies $\chi_{\varepsilon^{T+1}}^{(B-j)}(\vec{q}^{\,(B-j)}) \neq 0$.

Combining the above three subcases and following similar arguments used in the proof of Lemma 4 at the end of Section V-C, we conclude that there exist $\vec{q}^{\,(B-N-1)}, \ldots, \vec{q}^{\,(0)}$ such that $q_0^{(i)} \neq 0$ and $\chi_{\varepsilon^{T+1}}^{(i)}(\vec{q}^{\,(i)}) \neq 0$ for all $i \in \{B - 1, \ldots, B - N\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$. This together with the conclusions made in the above three subcases implies that statement (55) holds for all $i \in \{B-N-1, B-N-2, \ldots, 0\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$, which together with (53) implies that (23) holds for all $i \in \{B - N - 1, B - N - 2, \ldots, 0\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$.

### C. Combining the Two Cases

Combining the two cases studied in the preceding two subsections, we conclude that there exist $\vec{q}^{\,(B-N-1)}, \vec{q}^{\,(B-N-2)}, \ldots, \vec{q}^{\,(0)}$ such that $\mathbf{G} = [\mathbf{I}_k\ \mathbf{P}]$ with $\mathbf{P}$ having the form (50) satisfies (23) for all $i \in \{k-1, k-2, \ldots, 0\}$ and all $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$. In particular, $\mathbf{P}$ has the form (26).

## VII. 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 $\mathbf{G}_0^{\text{conv}}, \mathbf{G}_1^{\text{conv}}, \ldots, \mathbf{G}_m^{\text{conv}}$, define

$$\mathbf{G}^{\text{conv}} \triangleq \begin{bmatrix} \mathbf{G}_0^{\text{conv}} & \mathbf{G}_1^{\text{conv}} & \cdots & \mathbf{G}_T^{\text{conv}} \\ \mathbf{0}^{k\times n} & \mathbf{G}_0^{\text{conv}} & \cdots & \mathbf{G}_{T-1}^{\text{conv}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^{k\times n} & \mathbf{0}^{k\times n} & \cdots & \mathbf{G}_0^{\text{conv}} \end{bmatrix} \tag{58}$$

to be the truncated generator matrix where $\mathbf{G}_\ell^{\text{conv}} \triangleq \mathbf{0}^{k\times n}$ for any $m < \ell \leq T$ by convention. The following definition is standard (see, e.g., [11, Appendix A]).

*Definition 10:* For each $(n, k, m, T)_{\mathbb{F}}$-convolutional code, the column distance and the column span are

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

and

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

respectively, where

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

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

$$\mathrm{span}\big([\mathbf{x}_0\ \mathbf{x}_1\ \dots\ \mathbf{x}_T]\big) \triangleq \max\left\{i \in \{0,1,\dots,T\}\,\middle|\, \mathbf{x}_i \neq \mathbf{0}^{1\times n}\right\} - \min\left\{i \in \{0,1,\dots,T\}\,\middle|\, \mathbf{x}_i \neq \mathbf{0}^{1\times n}\right\}$$

denotes the length of the support of $[\mathbf{x}_0\ \mathbf{x}_1\ \dots\ \mathbf{x}_T]$ for any $[\mathbf{x}_0\ \mathbf{x}_1\ \dots\ \mathbf{x}_T] \in \mathbb{F}^{(T+1)n}$. Obviously, $\mathrm{c}_T \geq \mathrm{d}_T \geq 1$.

The following proposition states a well-known fact regarding the column distance and the column span for convolutional codes (see, e.g., [11, Appendix A]).

*Proposition 5:* Any $(n,k,m,T)_{\mathbb{F}}$-convolutional code whose column distance and column span are $\mathrm{d}_T$ and $\mathrm{c}_T$ respectively 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 $\mathrm{d}_T \geq N+1$ and $\mathrm{c}_T \geq B+1$.

Combining Proposition 5 and (13), we conclude that

$$\frac{k}{n} \leq \frac{T-\mathrm{d}_T+2}{T+\mathrm{c}_T-\mathrm{d}_T+1} \tag{59}$$

for any $(n,k,m,T)_{\mathbb{F}}$-convolutional code with column distance $\mathrm{d}_T$ and column span $\mathrm{c}_T$. Motivated by (59), we define the optimality of a convolutional code as follows.

*Definition 11:* An $(n,k,m,T)_{\mathbb{F}}$-convolutional code is said to be *optimal* if

$$\frac{k}{n} = \frac{T-\mathrm{d}_T+2}{T+\mathrm{c}_T-\mathrm{d}_T+1}.$$

We are ready to invoke Theorem 1 to infer the following result regarding $\mathrm{d}_T$ and $\mathrm{c}_T$ for optimal convolutional codes.

*Theorem 2:* Fix any $T$, $d$ and $c$ where $c \geq d \geq 1$, and let $\mathbb{F}$ be a finite field that satisfies

$$|\mathbb{F}| > 2\left(\binom{T+1}{\mathrm{d}-1} + T - c + 3\right). \tag{60}$$

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

*Proof:* Let $N \triangleq d-1$ and $B \triangleq c-1$. By (60), $\mathbb{F}$ satisfies (14). By Theorem 1, there exists an $(n,k,T,T)_{\mathbb{F}}$-convolutional code that is $(T+1,B,N)$-achievable where $k = T-N+1$ and $n = T+B-N+1$, which implies from Proposition 5 that $\mathrm{d}_T \geq N+1$ and $\mathrm{c}_T \geq B+1$. Since

$$\begin{aligned}\frac{k}{n} &= \frac{T-N+1}{T+B-N+1}\\ &\geq \frac{T-\mathrm{d}_T+2}{T+B-\mathrm{d}_T+2}\\ &\geq \frac{T-\mathrm{d}_T+2}{T+\mathrm{c}_T-\mathrm{d}_T+1}\end{aligned}$$

by all the preceding equations in this proof, it together with (59) implies that

$$\frac{k}{n} = \frac{T-\mathrm{d}_T+2}{T+\mathrm{c}_T-\mathrm{d}_T+1}. \tag{61}$$

In addition, since the equations $\mathrm{d}_T \geq N+1$, $k = T-N+1$, $n = T+B-N+1$ and (61) imply that $\mathrm{c}_T \leq B+1$, it together with the preceding equation $\mathrm{c}_T \geq B+1$ follows that $\mathrm{c}_T = B+1$, which together with (61) implies that $\mathrm{d}_T = N+1$. By Definition 11, this $(n,k,T,T)_{\mathbb{F}}$-convolutional code with column distance $\mathrm{d}_T = N+1 = d$ and column span $\mathrm{c}_T = B+1 = c$ is optimal. ■

*Remark 6:* Regarding Theorem 2, if $c = d$, then the field size requirement can be relaxed to $|\mathbb{F}| \geq T+1$ due to the following. For any $\mathrm{d}_T \in \{1,2,\dots,n\}$, a systematic MDS $(n,k)$-code with $n \triangleq T+1$ and $k \triangleq T-\mathrm{d}_T+2$ (any $\mathrm{d}_T-1$ symbol erasures can be recovered) always exists as long as $|\mathbb{F}| \geq n = T+1$ [21], which together with Lemma 1 implies the existence of an optimal $(n,k,T,T)_{\mathbb{F}}$-convolutional code such that $\mathrm{c}_T = \mathrm{d}_T$ (any $\mathrm{d}_T-1$ packet erasures can be recovered).

## VIII. Random Code Construction

Suppose we are given a channel model which introduces packet erasures, and we would like to communicate through the channel using an optimal $(n,k,T,T)_{\mathbb{F}}$-convolutional code with column distance $\mathrm{d}_T$ and column span $\mathrm{c}_T$ where the optimality is as defined in Definition 11. If $\mathbb{F}$ satisfies

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

Theorem 2 guarantees the existence of such an optimal convolutional code, but does not tell us how to find it efficiently. Therefore, we suggest in this section a practical method of finding optimal convolutional codes efficiently. To this end, we first fix any $(T, \mathrm{d}_T, \mathrm{c}_T)$ such that $T \geq \mathrm{c}_T - 1 \geq \mathrm{d}_T - 1$, and let $W \triangleq T+1$, $B \triangleq \mathrm{c}_T - 1$, $N \triangleq \mathrm{d}_T - 1$, $k \triangleq T-N+1$ and $n \triangleq k+B$. In addition, we fix a finite field $\mathbb{F}$ which does not necessarily satisfy (14). Our goal is to find an optimal $(n,k,T,T)_{\mathbb{F}}$-convolutional code with column distance $\mathrm{d}_T$ and column span $\mathrm{c}_T$. Recall the definition of $\mathbf{G}^{\mathrm{conv}}$ in (58) and the definition of $\mathbf{G}^{\mathrm{conv}}_\ell$ in Definition 2. A method that constructs the generator matrix $\mathbf{G}^{\mathrm{conv}}$ of an optimal convolutional code is described in the following subsection.

### A. Random Encoding

Consider the following two steps of constructing $\{\mathbf{G}^{\mathrm{conv}}_\ell\}_{\ell=0}^T$ and $\mathbf{G}^{\mathrm{conv}}$ in a random manner:

(I) Construct $\mathbf{G} \triangleq [\mathbf{I}_k\ \mathbf{P}]$ through randomly generating $\mathbf{P}$ according to the following rule:

- Depending on whether $k \geq B$ or $k < B$. we generate $\mathbf{P}$ in the form either (25) in Lemma 3 or (26) in Lemma 4 by selecting the non-zero elements in an i.i.d. fashion where each non-zero element is uniformly distributed on $\mathbb{F} \setminus \{0\}$.

Let $\mathcal{C}^{\mathrm{block}}_{\mathrm{i.i.d.}}(\mathbf{G})$ denote the random $(n,k,T)_{\mathbb{F}}$-block code with random generator matrix $\mathbf{G}$ as constructed above.

(II) Based on the block code $\mathcal{C}^{\mathrm{block}}_{\mathrm{i.i.d.}}(\mathbf{G})$ constructed above, we construct an $(n,k,T,T)_{\mathbb{F}}$-convolutional code denoted by $\mathcal{C}_{\mathrm{i.i.d.}}(\mathbf{G}^{\mathrm{conv}})$ as outlined in the proof of Theorem 1 in Section IV, where the generator matrix $\mathbf{G}^{\mathrm{conv}}$ is constructed according to (18) in Lemma 1. If $\mathcal{C}^{\mathrm{block}}_{\mathrm{i.i.d.}}(\mathbf{G})$ is $(W,B,N)$-achievable, it then follows from the arguments in the proof of Theorem 2 in

TABLE II

SUCCESS PROBABILITIES $P_{T,\mathrm{c}_T,\mathrm{d}_T}$ OF GENERATING AN OPTIMAL $(n,k,T,T)_{\mathbb{F}}$-CONVOLUTIONAL CODE

| $\lvert\mathbb{F}\rvert$ | $P_{7,8,6}$ | $P_{7,8,2}$ | $P_{7,7,5}$ | $P_{7,7,3}$ | $P_{7,6,4}$ | $P_{7,5,5}$ |
|---|---|---|---|---|---|---|
| 3 | 0 | 0.0617 | 0 | 0.0037 | 0 | 0 |
| 7 | 0.1290 | 0.3473 | 0.0780 | 0.1390 | 0.0437 | 0.0060 |
| 13 | 0.4643 | 0.5713 | 0.3440 | 0.3760 | 0.2263 | 0.1320 |
| 31 | 0.7787 | 0.7910 | 0.6860 | 0.6687 | 0.5773 | 0.4980 |
| 61 | 0.8950 | 0.8897 | 0.8340 | 0.8173 | 0.7667 | 0.7253 |

Section VII that $\mathcal{C}_{\text{i.i.d.}}(\mathbf{G}^{\text{conv}})$ is an optimal $(n,k,T,T)_{\mathbb{F}}$-convolutional code with column distance $\mathrm{d}_T$ and column span $\mathrm{c}_T$.

### B. Numerical Evaluation of a Randomly Constructed Code Being Optimal

In this subsection, we would like to estimate the probability that the random code constructed according to Section VIII-A is optimal. To simplify notation, we let $P_{T,\mathrm{c}_T,\mathrm{d}_T}$ denote the probability of the random code $\mathcal{C}_{\text{i.i.d.}}(\mathbf{G}^{\text{conv}})$ being an optimal $(n,k,T,T)_{\mathbb{F}}$-convolutional code with column distance $\mathrm{d}_T$ and column span $\mathrm{c}_T$. Since characterizing the exact expression of $P_{T,\mathrm{c}_T,\mathrm{d}_T}$ seems intractable, we would like to estimate $P_{T,\mathrm{c}_T,\mathrm{d}_T}$ by simulation. In our simulation, $P_{T,\mathrm{c}_T,\mathrm{d}_T}$ is estimated for the following parameters of $(T,\mathrm{c}_T,\mathrm{d}_T)$: $(7,8,6)$, $(7,8,2)$, $(7,7,5)$, $(7,7,3)$, $(7,6,4)$, and $(7,5,5)$. For each of the aforementioned parameters $(T,\mathrm{c}_T,\mathrm{d}_T)$, we plot the corresponding $P_{T,\mathrm{c}_T,\mathrm{d}_T}$ for $|\mathbb{F}| = 3, 7, 13, 31, 61$ by generating 3000 samples for each $|\mathbb{F}|$, and those $P_{T,\mathrm{c}_T,\mathrm{d}_T}$'s are displayed in Table II. We can see from Table II that all the $P_{T,\mathrm{c}_T,\mathrm{d}_T}$'s are positive for a field size as small as 7 and they are increasing with the field size as expected.

## IX. NUMERICAL STUDIES

The state-of-the-art MiDAS-interleaved and MiDAS-m-MDS convolutional codes have been proposed in [11, 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 [11, 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 particular, the random convolutional codes described in the previous subsection are based on interleaved block codes as illustrated in Table I, which leads to low decoding complexity (comparable to decoding a block code). Since we would like to compare the performance of the random convolutional codes described in the previous subsection with existing practical convolutional codes in real-world systems, only convolutional codes based on interleaved block codes (rather than m-MDS) codes are considered in our numerical studies. More specifically, we will compare the performance of our low-complexity random codes with several practical convolutional codes including MiDAS-interleaved codes [11, Sec. IV-D] and the Martinian-Sundberg code [16] over the following two popular statistical channel models — the GE channel [8], [9] and the Fritchman channel [10].

### A. The Gilbert-Elliott Channel and the Fritchman Channel

In our numerical studies, we consider the GE channel model and the Fritchman channel model as described in [11, Sec. VI], which are introduced below for the sake of completeness.

The GE channel is a two-state Markov model which consists of a good state and a bad state. In the good state, each channel packet is lost with probability $\epsilon \in [0,1)$ whereas in the bad state each channel packet is lost with probability 1. Let $\alpha$ and $\beta$ denote the transition probabilities from the good state to the bad state and vice versa. Then, the average loss rate of the GE channel is given by

$$\frac{\beta}{\alpha+\beta}\cdot\epsilon + \frac{\alpha}{\alpha+\beta}$$

As long as the channel stays in the bad state, the channel behaves as a burst erasure channel. In contrast, the channel behaves like an i.i.d. erasure channel when the channel stays in the good state.

The Fritchman channel model consists of one good state denoted by $G$ and $M$ bad states denoted by $E_1, E_2, \ldots, E_M$. If the state equals $G$ at time $i$, then it will transition to $E_1$ with probability $\alpha$ or stay at state $G$ with probability $1-\alpha$ at time $i+1$. If the state equals $E_M$ at time $i$, it will transition to $G$ with probability $\beta$ or stay at state $E_M$ with probability $1-\beta$ at time $i+1$. If the state equals $E_\ell$ for some $\ell \in \{1,2,\ldots,M-1\}$, then it will transition to $E_{\ell+1}$ with probability $\beta$ or stay at state $E_\ell$ with probability $1-\beta$ at time $i+1$. In the good state, each channel packet is lost with probability $\epsilon$ whereas in the bad state each channel packet is lost with probability 1. Fritchman and related higher-order Markov models are commonly used to model fade durations in mobile links.

### B. Simulation Results

In order to compare our random code with existing codes over practical channels, we plot their loss probabilities over the GE channel and the Fritchman channel where each loss probability is generated by simulating the codes over $10^8$ channel uses. The field size is set to be 997.

In Figure 2(a), we plot the loss probabilities over the GE channel with constant parameters $(\alpha,\beta) = (1\times 10^{-4}, 0.6)$ against the varying parameter $\epsilon$ for our random code, the MiDAS-interleaved code, the Martinian-Sundberg code and the random MDS code with $(W,T,B,N)$ equal to $(8,7,6,2)$, $(8,7,5,2)$, $(8,7,7,1)$ and $(8,7,4,4)$ respectively and rates equal to $1/2$, $21/41 \approx 1/2$, $1/2$ and $1/2$ respectively. The corresponding statistics of the burst length are plotted in Figure 2(b), which shows that the burst histogram follows a geometric distribution with a success probability of $\beta = 0.6$. As shown in Figure 2(a), our random code outperforms all the other codes over the GE channel for $0.003 \le \epsilon \le 0.01$. For $\epsilon \le 0.002$, the Martinian-Sundberg code performs the best, which indicates that the loss probability in this case is

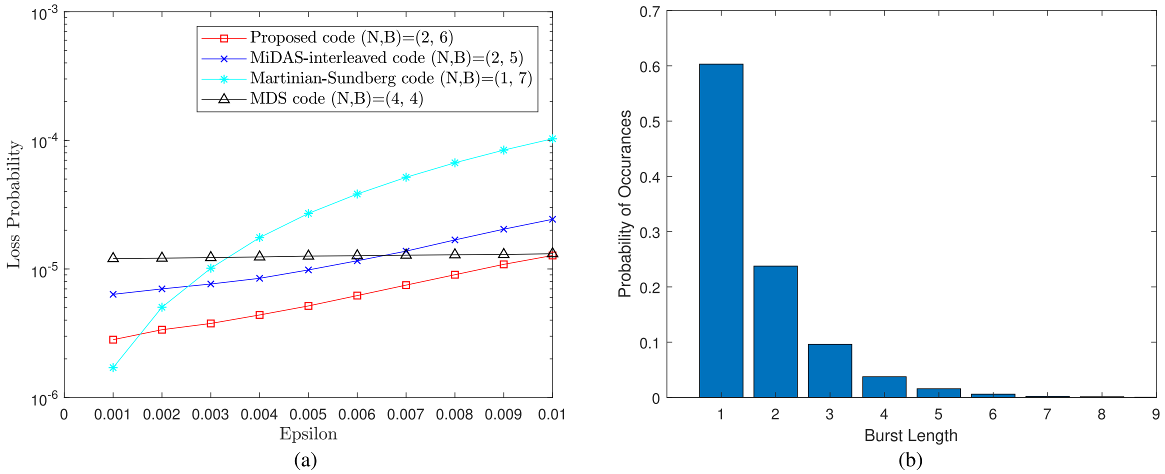

Fig. 2. Simulation for the GE channel with $(\alpha, \beta) = (1 \times 10^{-4}, 0.6)$. (a) Loss probability. (b) Burst histogram.

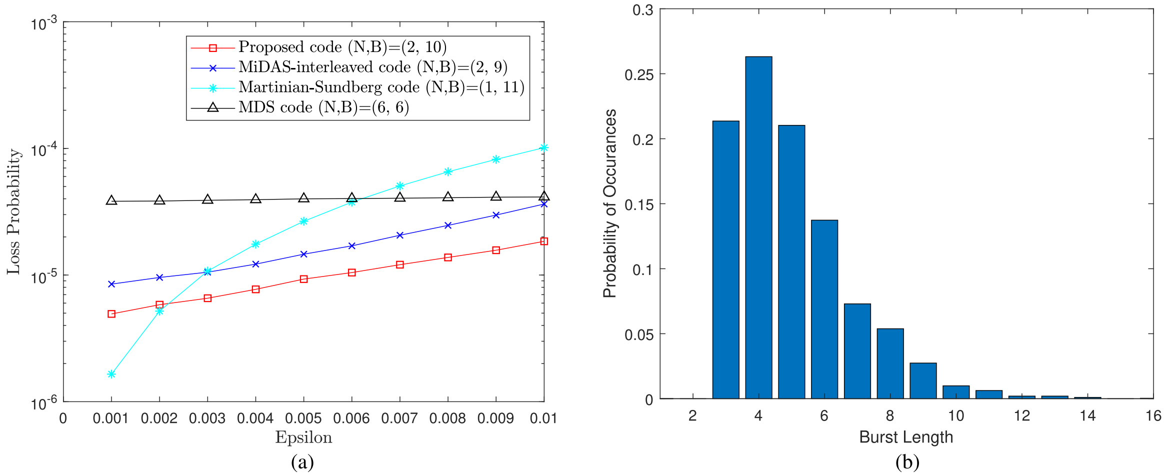

Fig. 3. Simulation for the 4-state Fritchman channel with $(\alpha, \beta, M) = (3 \times 10^{-5}, 0.6, 4)$. (a) Loss probability. (b) Burst histogram.

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. Indeed, our random code, the Martinian-Sundberg code and the random MDS code belong to the class of optimal convolutional codes in the sense of Definition 11. Therefore, it is not surprising that they collectively achieve the best performance as $\epsilon$ varies between 0 and 1.

In Figure 3(a), we plot the loss probabilities over the 4-state Fritchman channel with constant parameters $(\alpha, \beta, M) = (3 \times 10^{-5}, 0.6, 4)$ against the varying parameter $\epsilon$ for our random code, the MiDAS-interleaved code, the Martinian-Sundberg code and the random MDS code with $(W, T, B, N)$ equal to $(12, 11, 10, 2)$, $(12, 11, 9, 2)$, $(12, 11, 11, 1)$ and $(12, 11, 6, 6)$ respectively and rates equal to $11/21 \approx 1/2$, $12/23 \approx 1/2$, $12/22 \approx 1/2$ and $1/2$ respectively. The corresponding statistics of the burst length are plotted in Figure 3(b), which shows that the burst histogram follows a negative binomial distribution of $M - 1 = 3$ failures with a success probability of $\beta = 0.6$. As shown in Figure 3(a), our random code outperforms all the other codes over the 4-state Fritchman channel for $0.003 \leq \epsilon \leq 0.01$. For $\epsilon \leq 0.002$, the Martinian-Sundberg code performs the best, which indicates that the loss probability is dominated by burst rather than arbitrary errors. When $\epsilon$ approaches one, the random MDS code performs the best because the loss probability is dominated by arbitrary rather than burst errors. Our random code, the Martinian-Sundberg code and the random MDS code collectively achieve the best performance as $\epsilon$ varies between 0 and 1, which is consistent with the fact that they belong to the class of optimal convolutional codes in the sense of Definition 11.

## X. Concluding Remarks

In this paper, we study streaming codes over a packet erasure channel whose erasure pattern in every sliding window of size $W \geq T + 1$ is either a burst erasure of maximum length $B$ or multiple arbitrary erasures of maximum total count $N$. Under a fixed tolerable delay constraint $T$ for each transmitted packet, we have shown in Section II-C the existence of convolutional codes that achieve the maximum rate of communication over the erasure channel. In addition, we have characterized in Section VII the maximum achievable rate for convolutional codes with given column distance, column span and decoding delay. In our simulation, our proposed code outperforms all existing practical codes for various packet erasure probabilities over some instances of the GE channel and the Fritchman channel.

Throughout this paper, we have assumed that $W \geq T+1$ and (5) hold (cf. Section I-B) and showed that the maximum achievable rate for streaming codes is $C_{(W,T,B,N)} = \frac{T-N+1}{T+B-N+1}$. For the case where $W < T+1$ and (5) hold, it was shown in [11, Th. 1] that the maximum achievable rate $C_{(W,T,B,N)}$ is bounded as

$$C_{(W,T,B,N)} \leq \frac{W-N}{W+B-N}. \tag{62}$$

On the other hand, it follows from Theorem 1 that

$$C_{(W,W-1,B,N)} = \frac{W-N}{W+B-N}. \tag{63}$$

Since $C_{(W,T,B,N)} \geq C_{(W,W-1,B,N)}$ due to the assumption that $W < T+1$, it follows from (63) that

$$C_{(W,T,B,N)} \geq \frac{W-N}{W+B-N}. \tag{64}$$

Combining (62) and (64), we have $C_{(W,T,B,N)} = \frac{W-N}{W+B-N}$ for the case where $W < T+1$ and (5) hold.

## Appendix A
## Proof of Lemma 1

Suppose we are given a $(W,B,N)$-achievable $(n,k,T)_{\mathbb{F}}$-block code, and let $\mathbf{G} = [\,\mathbf{I}_k\ \mathbf{P}\,] \in \mathbb{F}^{k\times n}$ be the generator matrix. By Definition 8, the $(n,k,T)_{\mathbb{F}}$-block code has the following properties:

(i) The length of the block code is $n$.

(ii) From time 0 to $k-1$, the source symbols
$$[x[0]\ x[1]\ \cdots\ x[k-1]] = [s[0]\ s[1]\ \cdots\ s[k-1]]$$
are transmitted.

(iii) From time $k$ to $n-1$, the parity-check symbols
$$[x[k]\ x[k+1]\ \cdots\ x[n-1]] = [s[0]\ s[1]\ \cdots\ s[k-1]]\,\mathbf{P}$$
are transmitted.

(iv) Upon receiving
$$\begin{aligned}&[y[0]\ y[1]\ \ldots\ y[i+T]]\\ &\quad = [g_1(x[0],e_0)\ g_1(x[1],e_1)\ \ldots\ g_1(x[i+T],e_{i+T})],\end{aligned}$$
the destination can perfectly recover $s[i]$ by time $i+T$ for each $i \in \{0,1,\ldots,k-1\}$ as long as $e^\infty \in \Omega^\infty_{(W,B,N)}$.

In order to construct an $(n,k,n-1,T)_{\mathbb{F}}$-convolutional code, we first let $\{\mathbf{s}_i\}_{i=0}^\infty$ denote a sequence of length-$k$ packets and let $s_i[j]$ denote the $(j+1)^{\text{th}}$ element of $\mathbf{s}_i$ such that

$$\mathbf{s}_i \triangleq [s_i[0]\ s_i[1]\ \cdots\ s_i[k-1]] \tag{65}$$

for all $i \in \mathbb{Z}_+$. Using the convention that $\mathbf{s}_j \triangleq \mathbf{0}^{1\times k}$ for any $j<0$, we construct

$$\begin{aligned}&\big[x_i[0]\ x_{i+1}[1]\ \cdots\ x_{i+n-1}[n-1]\big]\\ &\qquad \triangleq \big[s_i[0]\ s_{i+1}[1]\ \cdots\ s_{i+k-1}[k-1]\big]\mathbf{G}\end{aligned} \tag{66}$$

for each $i \in \{-n+1,-n+2,\ldots\}$ where $\mathbf{G}$ is the generator matrix of the $(W,B,N)$-achievable $(n,k,T)_{\mathbb{F}}$-block code. In other words, we are coding $\mathbf{s}_i$ diagonally as illustrated in Table I. At each time $i \in \mathbb{Z}_+$, the source transmits

$$\mathbf{x}_i \triangleq \big[x_i[0]\ x_i[1]\ \cdots\ x_i[n-1]\big]. \tag{67}$$

In order to express $\mathbf{x}_i$ in the form of (9), we let $g_{i,j}$ be the entry situated in row $i$ and column $j$ of $\mathbf{G}$ such that

$$\mathbf{G} = \big[g_{i,j}\big]_{\substack{0\leq i\leq k-1,\\ 0\leq j\leq n-1}}$$

and define $\mathbf{G}_\ell^{\text{conv}}$ as in (18) for each $\ell \in \{0,1,\ldots,n-1\}$ such that

$$\mathbf{G} = [\,\mathbf{I}_k\ \mathbf{P}\,] = \sum_{\ell=0}^{n-1}\mathbf{G}_\ell^{\text{conv}}.$$

Following (67), we consider

$$\begin{aligned}\mathbf{x}_i &= \sum_{\ell=0}^{n-1}\big[s_{i-\ell}[0]\ s_{i-1-\ell}[1]\ \cdots\ s_{i-k+1-\ell}[k-1]\big]\\ &\qquad \times\left[\begin{array}{c:c:c} & g_{0,\ell} & \\ \mathbf{0}^{k\times\ell} & \vdots & \mathbf{0}^{k\times(n-\ell-1)} \\ & g_{k-1,\ell} & \end{array}\right] \qquad (68)\\ &= \sum_{\ell=0}^{n-1}\sum_{j=0}^{k-1}\big[\mathbf{0}^{1\times\ell}\ \ s_{i-\ell}[j]\,g_{j,\ell}\ \ \mathbf{0}^{1\times(n-\ell-1)}\big]\\ &= \sum_{\ell=0}^{n-1}\mathbf{s}_{i-\ell}\,\mathbf{G}_\ell^{\text{conv}} \qquad (69)\end{aligned}$$

for each $i \in \mathbb{Z}_+$, where

- (68) is due to (66) and (67).
- (69) is due to (65) and (18) under our convention that $\mathbf{s}_t = \mathbf{0}^{1\times k}$ for all $t<0$.

Based on the $(W,B,N)$-achievable $(n,k,T)_{\mathbb{F}}$-block code which satisfies Properties (i) to (iv) as stated at the beginning of this proof, we construct an $(n,k,n-1,T)_{\mathbb{F}}$-convolutional code whose encoding function at time $i$ is specified by (69), where $\mathbf{x}_i$ and $\mathbf{s}_i$ satisfy (67) and (65) respectively. Our goal is to show that the convolutional code is $(W,B,N)$-achievable. To this end, we fix any $i \in \mathbb{Z}_+$ and any $e^\infty \in \Omega^\infty_{(W,B,N)}$, and would like to show that the destination can perfectly recover $\mathbf{s}_i = \big[s_i[0]\ s_i[1]\ \cdots\ s_i[k-1]\big]$ based on

$$\begin{aligned}&[\mathbf{y}_0\ \mathbf{y}_1\ \ldots\ \mathbf{y}_{i+T}]\\ &\qquad = [g_n(\mathbf{x}_0,e_0)\ g_n(\mathbf{x}_1,e_1)\ \ldots\ g_n(\mathbf{x}_{i+T},e_{i+T})].\end{aligned} \tag{70}$$

According to (67), for each $i \in \{-n+1,-n+2,\ldots\}$, the symbols in $\big[x_i[0]\ x_{i+1}[1]\ \cdots\ x_{i+n-1}[n-1]\big]$ are transmitted between time $i$ to time $i+n-1$. Therefore, it follows from (66), Property (iv) and (70) that for each $i \in \mathbb{Z}_+$ and each $\ell \in \{0,1,\ldots,k-1\}$, the destination can perfectly recover $s_i[\ell]$ by time $i+T$ based on $[\mathbf{y}_i\ \mathbf{y}_{i+1}\ \ldots\ \mathbf{y}_{i+T}]$, which implies that the destination can perfectly recover $\mathbf{s}_i$ time $i+T$ based on $[\mathbf{y}_0\ \mathbf{y}_1\ \ldots\ \mathbf{y}_{i+T}]$. Consequently, for any $i \in \mathbb{Z}_+$ and any $e^\infty \in \Omega^\infty_{(W,B,N)}$, the destination can perfectly recover $\mathbf{s}_i$ by time $i+T$, which implies that the $(n,k,n-1,T)_{\mathbb{F}}$-convolutional code is $(W,B,N)$-achievable. In addition, using (65), (66) and (67), we obtain (19).

## Appendix B
## Proof of Lemma 2

Let $\mathbf{G} = [\mathbf{g}_0\ \mathbf{g}_1\ \ldots\ \mathbf{g}_{n-1}] \in \mathbb{F}^{k\times n}$ be a matrix that satisfies (23) for each $i \in \{0,1,\ldots,k-1\}$ and each maximal $(T+1, B, N)$-erasure pattern $\varepsilon^{T+1} \in \Omega_{B,N}^{T+1}$. First, we would like to construct an $(n,k,T)_{\mathbb{F}}$-block code with generator matrix $\mathbf{G}$ that is $(T+1, B, N)$-achievable (cf. Definition 8), and this lemma will then follow because any $(T+1, B, N)$-achievable $(n,k,T)_{\mathbb{F}}$-block code is also an $(W, B, N)$-achievable $(n,k,T)_{\mathbb{F}}$-block code due to the assumption that $W \ge T+1$. Since the encoding strategy of an $(n,k,T)_{\mathbb{F}}$-block code with generator matrix $\mathbf{G}$ is completely determined by (15), it suffices to show the existence of $\{\varphi_{i+T}\}_{i=0}^{k-1}$ such that

$$s[i] = \begin{cases} \varphi_{i+T}(g_1(x[0], e_0), \ldots, g_1(x[i+T], e_{i+T})) \\ \qquad \text{if } 0 \le i \le B-N, \\ \varphi_{i+T}(g_1(x[0], e_0), \ldots, g_1(x[n], e_n)) \\ \qquad \text{if } B-N+1 \le i \le k-1 \end{cases} \tag{71}$$

holds for any $(T+1, B, N)$-erasure sequence $e^{\infty} \in \Omega_{(T+1,B,N)}^{\infty}$. Recognizing the fact due to (15), (16) and (17) that

$$\begin{aligned} &[g_1(x[0], e_0)\ g_1(x[1], e_1)\ \ldots\ g_1(x[n-1], e_{n-1})]\, \mathbf{E}_{e^n} \\ &\quad = [x[0]\ x[1]\ \ldots\ x[n-1]]\, \mathbf{E}_{e^n} \\ &\quad = [s[0]\ s[1]\ \ldots\ s[k-1]]\, \mathbf{G}\, \mathbf{E}_{e^n}, \end{aligned} \tag{72}$$

we fix an arbitrary $(T+1, B, N)$-erasure sequence $e^{\infty} \in \Omega_{(T+1,B,N)}^{\infty}$ and would like to show that

$$\mathbf{u}_i^{(k)} \in \begin{cases} \text{space}\Big( \big[\mathbf{g}_0\ \cdots\ \mathbf{g}_{i+T}\big]\, \mathbf{E}_{e^{i+T+1}} \Big) & \text{if } 0 \le i \le B-N, \\ \text{space}\Big( \big[\mathbf{g}_0\ \cdots\ \mathbf{g}_{n-1}\big]\, \mathbf{E}_{e^n} \Big) & \text{if } B-N+1 \\ & \quad \le i \le k-1 \end{cases} \tag{73}$$

for each $i$, which together with (72) would then imply the existence of $\{\varphi_{i+T}\}_{i=0}^{k-1}$ that satisfy (71) for each $i$. We will show (73) by induction on $i = 0, 1, \ldots, k-1$. For $i = 0$, (73) follows directly from (23) by setting $\varepsilon^{T+1} = e^{T+1}$. Suppose (73) holds for each $i = 0, 1, \ldots, j$ for some $j < k-1$. Then, showing (73) for $i^* = j+1$ is equivalent to showing

$$\mathbf{u}_{i^*}^{(k)} \in \begin{cases} \text{space}\left( \mathbf{I}_{k-i^*}^{(k)} \big[\mathbf{g}_0\ \cdots\ \mathbf{g}_{i^*+T}\big]\, \mathbf{E}_{e^{i^*+T+1}} \right) \\ \qquad \text{if } 0 \le i^* \le B-N, \\ \text{space}\left( \mathbf{I}_{k-i^*}^{(k)} \big[\mathbf{g}_0\ \cdots\ \mathbf{g}_{n-1}\big]\, \mathbf{E}_{e^n} \right) \\ \qquad \text{if } B-N+1 \le i^* \le k-1, \end{cases}$$

which is a direct consequence of (23). By mathematical induction, we have proved that (73) holds for each $i = 0, 1, \ldots, k-1$.

## ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the three anonymous reviewers for their useful comments which helped us improve the presentation of this work.

## REFERENCES

[1] 5G-PPP. (Feb. 2015). *5G Empowering Vertical Industries*. [Online]. Available: https://5g-ppp.eu/roadmaps/

[2] International Telecommunication Union, "Recommendation G.114," Tech. Rep., May 2003. [Online]. Available: https://www.itu.int/rec/T-REC-G.114

[3] T. Stockhammer and M. M. Hannuksela, "H.264/AVC video for wireless transmission," *IEEE Wireless Commun.*, vol. 12, no. 4, pp. 6–13, Aug. 2005.

[4] A. Badr, A. Khisti, W.-T. Tan, and J. Apostolopoulos, "Perfecting protection for interactive multimedia: A survey of forward error correction for low-delay interactive applications," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 95–113, Mar. 2017.

[5] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile Internet," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[6] G. Hasslinger and O. Hohlfeld, "The Gilbert–Elliott model for packet loss in real time services on the Internet," in *Proc. 14th GI/ITG Conf. Meas., Modelling Eval. Comput. Commun. Syst. (MMB)*, Dortmund, Germany, Mar./Apr. 2008, pp. 1–15.

[7] O. Hohlfeld, R. Geib, and G. Hasslinger, "Packet loss in real-time services: Markovian models generating QoE impairments," in *Proc. 16th Int. Workshop Quality Service*, Enschede, The Netherlands, Jun. 2008, pp. 239–248.

[8] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sep. 1960.

[9] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, no. 5, pp. 1977–1997, Sep. 1963.

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

[11] 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, Jan. 2017.

[12] D. W. Hagelbarger, "Recurrent codes: Easily mechanized, burst-correcting, binary codes," *Bell Syst. Tech. J.*, vol. 38, no. 4, pp. 969–984, Jul. 1959.

[13] A. Wyner and R. Ash, "Analysis of recurrent codes," *IEEE Trans. Inf. Theory*, vol. IT-9, no. 3, pp. 143–156, Jul. 1963.

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

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

[16] 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, Oct. 2004.

[17] D. Leong and T. Ho, "Erasure coding for real-time streaming," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 289–293.

[18] D. Leong, A. Qureshi, and T. Ho, "On coding for real-time streaming under packet erasures," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 1012–1016.

[19] A. Badr, A. Khisti, W.-T. Tan, and J. Apostolopoulos, "Streaming codes for channels with burst and isolated erasures," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 2850–2858.

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

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

[22] A. Badr, A. Khisti, W.-T. Tan, X. Zhu, and J. Apostolopoulos, "FEC for VoIP using dual-delay streaming codes," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, May 2017, pp. 1–9.

[23] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, 2nd ed. Hoboken, NY, USA: Wiley, 2015.

**Silas L. Fong** (M'15) received the B.Eng., M.Phil. and Ph.D. degrees in Information Engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2005, 2007 and 2011 respectively. He is currently a Postdoctoral Fellow with the Department of Electrical and Computer Engineering at University of Toronto, Toronto, ON, Canada. From 2011 to 2013, Dr. Fong was a Postdoctoral Fellow with the Department of Electronic Engineering at City University of Hong Kong, Hong Kong. From 2013 to 2014, he was a Postdoctoral Associate with the Department of Electrical and Computer Engineering at Cornell University, Ithaca, NY. From 2014 to 2017, he was a Research Fellow with the Department of Electrical and Computer Engineering at National University of Singapore (NUS), Singapore. His research interests include information theory and its applications to communication systems such as relay networks, wireless networks, and energy-harvesting channels.

**Ashish Khisti** (S'02–M'08) received his B.ASc. degree in Engineering Sciences (Electrical Option) from University of Toronto, and his S.M. and Ph.D. degrees in Electrical Engineering from the Massachusetts Institute of Technology. Between 2009–2015, he was an assistant professor in the Electrical and Computer Engineering Department at the University of Toronto. He is presently an associate professor, and holds a Canada Research Chair in the same department. He is a recipient of an Ontario Early Researcher Award, the Hewlett-Packard Innovation Research Award and a Cisco Research Center Award. He served as an associate editor for IEEE TRANSACTIONS ON INFORMATION THEORY between 2014–2017.

**Baochun Li** (F'15) received his B.Engr. degree from the Department of Computer Science and Technology, Tsinghua University, China, in 1995 and his M.S. and Ph.D. degrees from the Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, in 1997 and 2000. Since 2000, he has been with the Department of Electrical and Computer Engineering at the University of Toronto, where he is currently a Professor. He holds the Bell Canada Endowed Chair in Computer Engineering since August 2005. His research interests include cloud computing, distributed systems, datacenter networking, and wireless systems. Dr. Li has co-authored more than 360 research papers, with a total of over 17000 citations, an H-index of 75 and an i10-index of 233, according to Google Scholar Citations. He was the recipient of the IEEE Communications Society Leonard G. Abraham Award in the Field of Communications Systems in 2000. In 2009, he was a recipient of the Multimedia Communications Best Paper Award from the IEEE Communications Society, and a recipient of the University of Toronto McLean Award. He is a member of ACM and a Fellow of IEEE.

**Wai-Tian Tan** (SM'12) received BS from Brown University, MSEE from Stanford University, and PhD from University of California, Berkeley, all in electrical engineering. He was a researcher at Hewlett Packard Laboratories from 2000 to 2013 working on various aspects of multimedia communications and systems. He has been with Cisco Systems since 2013, where he is a principal engineer in the Innovations Lab within Enterprise Networking Business. He currently works on various aspects of learning and sensing in wireless networking.

**Xiaoqing Zhu** (M'09) is currently a Sr. Technical Leader at the Innovation Labs of Cisco Systems, Inc. Her research interests include Internet video delivery, real-time interactive multimedia communications, distributed resource optimization, and wireless networking. She holds a B.Eng. in Electronics Engineering from Tsinghua University, Beijing, China. She received both M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, CA, USA. She has previously interned at IBM Almaden Research Center in 2003, and at Sharp Labs of America in 2006. Dr. Zhu has published over 80 peer-reviewed journal and conference papers, receiving the Best Student Paper Award at ACM Multimedia in 2007 and the Best Presentation Award at IEEE Packet Video Workshop in 2013. She is author of 17 issued U.S. patent applications, with a few more pending. Dr. Zhu has served extensively within the multimedia research community, as TPC member and area chair, special issue guest editor, etc. She is chair of the MCDIG (Multimedia Content Distribution: Infrastructure and Algorithms) Interest Group in Multimedia Communication Technical Committee (MMTC) of IEEE. She currently serves as Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA.

**John Apostolopoulos** (F'07) is CTO/VP of Cisco's Enterprise Networking Business (Cisco's largest business) where he drives the technology and architecture direction in strategic areas for the business. This covers the broad Cisco portfolio including Intent-based Networking (IBN), Internet of Things (IoT), wireless (ranging from WiFi to emerging 5G), application-aware networking, multimedia networking, indoor-location-based services, connected car, machine learning and AI applied to the aforementioned areas, and deep learning for visual analytics. Previously, John was Lab Director for the Mobile & Immersive Experience Lab at HP Labs. The MIX Lab conducted research on novel mobile devices and sensing, mobile client/cloud multimedia computing, immersive environments, video & audio signal processing, computer vision & graphics, multimedia networking, glasses-free 3D, next-generation plastic displays, wireless, and user experience design. He is an IEEE Fellow, was an IEEE SPS Distinguished Lecturer, named "one of the world's top 100 young (under 35) innovators in science and technology" (TR100) by MIT Technology Review, received a Certificate of Honor for contributing to the US Digital TV Standard (Engineering Emmy Award 1997), and his work on media transcoding in the middle of a network while preserving end-to-end security (secure transcoding) was adopted in the JPSEC standard. He has published over 100 papers, including receiving 5 best paper awards, and has about 75 granted US patents. John also has strong collaborations with the academic community and was a Consulting Associate Professor of EE at Stanford (2000-09), and frequently lecturers at MIT. He received his B.S., M.S., and Ph.D. degrees in EECS from MIT.