An Explicit Rate-Optimal Streaming Code for Channels with Burst and Arbitrary Erasures

Elad Domanovitz
School of Electrical Engineering
Tel Aviv University
Tel Aviv-Yafo, Israel
E-mail: domanovi@post.tau.ac.il

Silas L. Fong and Ashish Khisti
Department of Electrical and Computer Engineering
University of Toronto
Toronto, ON M5S 3G4, Canada
E-mail: {silas.fong, akhisti}@utoronto.ca

***Abstract*—In this paper, we consider transmitting a sequence of messages (a streaming source) over a packet erasure channel, where every source message must be recovered perfectly at the destination subject to a fixed decoding delay. Recently, the capacity of such a channel was established. However, the codes shown to achieve the capacity are either non-explicit constructions (proven to exist) or explicit constructions requiring large field size that scales exponentially with the delay. This work presents an explicit rate-optimal construction for all channel and delay parameters over a field size that scales only quadratically with the delay.**

I. Introduction

One of the fundamental requirements of a communication system is to handle interrupts that occur during the transmission of information. Such interrupts occur either due to physical nature of the channel (for example, fading) or due to packet drops in an interim point of the link (for example, due to congestion or overload). In general, there are two main error control schemes in use: Automatic repeat request (ARQ) and forward error correction (FEC). While ARQ is appealing due to not increasing the overhead added to the payload in advance, when considering low latency applications, it may not offer an acceptable solution.

Designing FEC with latency constraint (streaming setup) was introduced by Martinian and Sunderberg in [1]. This work analyzed only the case of bursts separated by sufficiently long guard intervals. In this paper we consider the sliding-window burst erasure model (which incorporates also arbitrary erasures) originally proposed by Badr et al. in [2]. The exact capacity in this setup was determined independently by Fong et al. in [3] and Krishnan et al. [4]. The proof in [3] is existential (proving the existence of an appropriate generator matrix), while the field size requirements are large ($O\binom{T}{N}$). Recently, Dudzicz et al. [5] showed an explicit (and systematic) construction for all rates greater than or equal $1/2$, albeit the required field size is larger than the non-explicit constructions of [3].

In [4] an explicit construction based on linearized polynomials is presented. However, except for a small range of parameters, the field-size requirements are still large $\exp(T)$. The field size was further addressed in [6] in which a new rate-optimal code construction covering all channel and delay parameters, which requires the field size to grow quadratically $O(T^2)$ is introduced. However, explicit constructions were presented only for specific cases. In this paper we present an explicit construction which requires only a quadratic field size for all channel and delay parameters.

II. Problem Statement

A. Notation

We denote by $\mathbb{F}_q$ the finite field of size q where the elements of the matrix belong to. The extension field is denoted by $\mathbb{F}_{q^m}$. The symbol vectors are represented using the bold characters (e.g., $\mathbf{s}$). A generator matrix is denoted by $\mathbf{G}$. The i^{th} column of $\mathbf{G}$ is denoted as $\mathbf{g}_i$. The identity matrix of size k is denoted by I_k. In a standard manner, we define by $R = \frac{k}{n}$ as the rate of the code.

A $k \times n$ matrix $\mathbf{G}$ over a finite field $\mathbb{F}_q$, with $k \leq n$, will be referred to as a MDS matrix if any k distinct columns of $\mathbf{G}$ form a linearly independent set. In the sequel we use the following definitions

Definition 1. *(Punctured Code) Let $\mathcal{C}$ be an (n, k) linear code over $\mathbb{F}_q$. Given a subset $\mathcal{P}$ of $[0 : n-1]$, the code $\mathcal{C}$ punctured on the coordinates in $\mathcal{P}$, is the linear code of length $(n-|\mathcal{P}|)$ obtained from $\mathcal{C}$ by deleting all the coordinates in $\mathcal{P}$.*

When $\mathcal{C}$ is an (n, k) MDS code, puncturing it over $\mathcal{P}$ results with $(n - |\mathcal{P}|, k)$ MDS code.

Definition 2. *(Shortened Code) Let $\mathcal{C}$ be an (n, k) linear code over $\mathbb{F}_q$. Given a subset $\mathcal{P}$ of $[0 : n-1]$, consider the subcode $\mathcal{C}^*$ achieved when assuming*

$$s_i = 0 \quad \forall\, i \in \mathcal{P}. \tag{1}$$

Then by the phrase $\mathcal{C}$ shortened on the coordinates in $\mathcal{P}$, we will mean the linear code of length $(n - |\mathcal{P}|)$ obtained from $\mathcal{C}^*$ after puncturing on the coordinates given by $\mathcal{P}$.

When $\mathcal{C}$ is an (n, k) MDS code, shortening it over $\mathcal{P}$ results with an $(n-|\mathcal{P}|, k-|\mathcal{P}|)$ MDS code. In the sequel we denote the shortened code as $\mathcal{C}^{|\mathcal{P}|}$.

We note that the field size required to support an (n, k) MDS code is $O(n)$ (as, for example, (n, k) Reed-Solomon code requires field in size n).

B. Streaming Codes

We consider sending a source S which generates at each time instant $t \in \{0,1,2,\ldots\}$ a packet $\mathbf{s}[t] \triangleq [s_0[t], s_1[t], \ldots, s_{k-1}[t]]^T$ with $s_i[t] \in \mathbb{F}_q$ for $i \in \{0,1,2,\ldots,k-1\}$. Note that $\mathbb{F}_q$ is a finite field of size q such that $\mathbf{s}[t] \in \mathbb{F}_q^k$. Each source packet $\mathbf{s}[t]$ is encoded using a causal convolutional encoder $\mathcal{E}$ such that the encoded packet $\mathbf{x}[t] = [x_0[t], x_1[t], \ldots, x_{n-1}[t]]^T = \mathcal{E}(\mathbf{s}[0], \mathbf{s}[1], \ldots, \mathbf{s}[t])$. Where $\mathbf{x}[t] \in \mathbb{F}_q^n$ and $\mathcal{E} : \mathbb{F}_q^{k(t+1)} \to \mathbb{F}_q^n$ is the encoding function.

Each encoded packet $\mathbf{x}[t]$ is transmitted over a channel which introduces erasures on a packet level. The receiver receives at each time $t \in \{0,1,2,\ldots\}$ the packet $\mathbf{y}[t]$ such that

$$\mathbf{y}[t] = \begin{cases} * & \text{if } \mathbf{x}[t] \text{ is erased} \\ \mathbf{x}[t] & \text{Otherwise} \end{cases} \tag{2}$$

At the receiver the decoder $\mathcal{D}$ must reconstruct perfectly the source packet $\mathbf{s}[t]$ within the delay T given the previously received packets $\{y[0], y[1], \ldots, y[t+T]\}$, i.e., $\hat{\mathbf{s}}[t] = \mathcal{D}(y[0], y[1], \ldots, y[t+T]) = \hat{\mathbf{s}}[t]$ with $\mathcal{D}$ being the decoding function and $\hat{\mathbf{s}}[t]$ being the reconstructed source packet $\mathbf{s}[t]$ by the decoder.

C. Channel Model

The channel model considered in this work is the sliding-window burst erasure channel denoted by $\mathcal{C}(W,B,N)$ that was introduced by Badr et al. in [2]. This model introduces up to B consecutive erasures or N arbitrarily positioned arbitrary isolated erasures in any window of size W among the sequence of transmitted packets $\mathbf{x}[t]$.

Since a channel that introduces any N arbitrary erasures can introduce any burst erasure of length N, we assume without loss of generality that $B \geq N$.

We further assume that $W \geq T+1$. In case where $B < W \leq T+1$ we can achieve the capacity by reducing the effective delay to $T_{\text{eff}} = W-1$ as discussed in [7]. Furthermore the capacity is trivially zero if $W \leq B$ as an erasure sequence that erases all the channel packets becomes admissible.

Thus we can assume without loss of generality that

$$W > T \geq B \geq N \geq 1. \tag{3}$$

For further details refer to Section I-B of [3].

D. Capacity

A streaming code with the encoder and decoder definitions in Section II-A is feasible for the $\mathcal{C}(W,B,N)$ sliding window channel if every source packet can be recovered with a delay of T. The maximum achievable rate of a feasible code is the capacity.

Recently, independent works in [3], [4] established that the capacity is given by:

$$C = \frac{T-N+1}{T-N+B+1}. \tag{4}$$

III. Code construction

In [4], it has been shown that designing an optimal streaming code for channels with burst and arbitrary erasures is equivalent to designing a linear (n,k) block which conforms to $\mathcal{C}(W,B,N)$ with delay-constraint T (and thus, the required streaming code can be generated from this block code using diagonal interleaving, see, e.g., [1]). Therefore, in this Section we present the construction of the generator matrix $\mathbf{G}$ of the (n,k) block code $\mathcal{C}$ and show it can decode all the data symbols with maximal delay of T from a burst of length B or N arbitrary erasures symbols in a sliding window of W.

We define

$$\begin{aligned} k &= T-N+1 \\ n &= k+B \end{aligned} \tag{5}$$

in the same manner as was defined in [3]. The code is constructed as follows.

- We start with an (n,k) MDS code $\mathcal{C}''$ over $\mathbb{F}_q$ with the generator matrix[1]

  $$\mathbf{G}'' = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & Y & \cdots & \cdots & \cdots & \cdots & Y \\ 0 & 1 & 0 & 0 & \cdots & 0 & Y & \cdots & \cdots & \cdots & \cdots & Y \\ 0 & 0 & \ddots & 0 & \cdots & 0 & Y & \cdots & \cdots & \cdots & \cdots & Y \\ \vdots & \vdots & & 1 & & \vdots & \vdots & & & & & \vdots \\ \vdots & \vdots & & & \ddots & 0 & \vdots & & & & & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 & Y & \cdots & \cdots & \cdots & \cdots & Y \end{bmatrix}.$$

- We perform row operations to generate code $\mathcal{C}'$ with the generator matrix

  $$\mathbf{G}' = \begin{array}{c} \begin{bmatrix} 1 & X & \cdots & X & 0 & 0 & 0 & \cdots & 0 & X & \cdots & X \\ 0 & 1 & X & \cdots & X & 0 & 0 & \cdots & 0 & \vdots & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & 0 & X & \cdots & X \\ \vdots & \vdots & & 1 & \ddots & \ddots & \ddots & & \vdots & X & & X \\ \vdots & \vdots & & & \ddots & X & \cdots & X & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 & X & \cdots & X & X & \cdots & X \end{bmatrix} \\ \underbrace{\hphantom{1\ X\ \cdots\ X\ 0\ 0\ 0\ 0}}_{k}\underbrace{\hphantom{0\ \cdots\ 0\ 0}}_{N-1}\underbrace{\hphantom{X\ \cdots\ X\ X}}_{B-N+1} \end{array},$$

  where the goal is to "spread" $N-1$ parity symbols diagonally with the data symbols. As it is easy to see that code $\mathcal{C}'$ (and also $\mathcal{C}''$) can recover from a burst of size B starting at time 0 only at time $T' = k+B-1$. It follows that for any $B > N$, $T' > T$ hence this code does not meet the required constraints. Yet, as we show next, it is an important interim step.

[1]We note that Y is not a constant element, but rather a place-holder.

This "spreading" is achieved via successive row cancellation. Equivalently, it can be denoted as $\mathbf{G}' = \mathbf{M}\mathbf{G}''$ where matrix $\mathbf{M}$ is an upper triangular matrix which is denoted as

$$\mathbf{M} = \begin{array}{c} \qquad\quad \overbrace{\hphantom{Y' \quad \cdots \quad Y'}}^{N-1} \qquad\qquad\qquad\qquad \\ \begin{bmatrix} 1 & Y' & \cdots & Y' & 0 & \cdots & \cdots & 0 \\ 0 & 1 & Y' & \cdots & Y' & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & Y' & \cdots & Y' & 0 \\ 0 & \cdots & 0 & 0 & 1 & Y' & \cdots & Y' \\ 0 & \cdots & 0 & 0 & 0 & \ddots & \ddots & Y' \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 & Y' \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array},$$

where Y' denotes a function of one of the Y symbols. Since all entries in $\mathbf{M}$ are linear combinations of elements from $\mathbf{G}'$ it is also over $\mathbb{F}_q$. Further, $\mathbf{M}$ is a full-rank matrix hence it is invertible. Since an erasure of any l columns in $\mathbf{G}'$ can be translated to an erasure of l columns in $\mathbf{G}''$ (by multiplying with the inverse of $\mathbf{M}$), the following property holds.

Property 1. *block code $\mathcal{C}'$ with generator matrix $\mathbf{G}'$ is an (n, k) MDS code over $\mathbb{F}_q$.*

- Finally, we replace the $(B-N+1)\times(B-N+1)$ upper right matrix with $\alpha \cdot \mathbf{I}_{B-N+1}$ where $\alpha \in \mathbb{F}_{q^2} \setminus \mathbb{F}_q$ to generate code $\mathcal{C}$ with the generator matrix

$$\mathbf{G} =$$

$$\begin{bmatrix} 1 & X & \cdots & X & 0 & 0 & 0 & \cdots & 0 & \alpha & \cdots & 0 \\ 0 & 1 & X & \cdots & X & 0 & 0 & \cdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & 0 & 0 & \cdots & \alpha \\ \vdots & \vdots & & 1 & \ddots & \ddots & \ddots & & \vdots & X & \cdots & X \\ \vdots & \vdots & & & \ddots & X & \cdots & X & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 & X & \cdots & X & X & \cdots & X \end{bmatrix}.$$
$$\underbrace{\hphantom{1 \quad X \quad \cdots \quad X \quad 0 \quad 0}}_{k}\ \underbrace{\hphantom{0 \quad \cdots \quad 0}}_{N-1}\ \underbrace{\hphantom{\alpha \quad \cdots \quad 0}}_{B-N+1}$$

The generator matrix $\mathbf{G}$ is composed of the following three blocks

- $\mathbf{H}_1$ - The left $k \times (k + N - 1)$ matrix.
- $\mathbf{H}_2$ - The lower right $(k - (B - N + 1)) \times (n - (B - N + 1))$ matrix.
- $\mathbf{H}_3$ - The upper right $(B-N+1)\times(B-N+1)$ matrix.

These blocks are depicted below:

$$\mathbf{G} =$$

$$\begin{array}{c} \qquad \mathbf{H}_1 \qquad\qquad\qquad\qquad\qquad\qquad \mathbf{H}_3 \\ \begin{bmatrix} 1 & X & \cdots & X & 0 & 0 & 0 & \cdots & 0 & \alpha & \cdots & 0 \\ 0 & 1 & X & \cdots & X & 0 & 0 & \cdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & 0 & 0 & \cdots & \alpha \\ \vdots & \vdots & & 1 & \ddots & \ddots & \ddots & & \vdots & X & \cdots & X \\ \vdots & \vdots & & & \ddots & X & \cdots & X & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 & X & \cdots & X & X & \cdots & X \end{bmatrix} \\ \mathbf{H}_2 \end{array}.$$

Before showing that block code $\mathcal{C}$ with generator matrix $\mathbf{G}$ is a capacity-achieving streaming code we show the following properties of $\mathbf{H}_1$ and $\mathbf{H}_2$.

Property 2. *Block $\mathbf{H}_1$ is a generator matrix of a $(k+N-1, k)$ MDS code over $\mathbb{F}_q$.*

This can be viewed from noting that $\mathbf{H}_1$ is a result of puncturing $(B-N-1)$ columns from G' which is a generator matrix of (n, k) MDS code.

Property 3. *Block $\mathbf{H}_2$ is a generator matrix of an $(n-(B-N+1), k-(B-N+1))$ MDS code over $\mathbb{F}_q$.*

This property holds since one may assume that $\mathbf{H}_2$ is generated from $\mathbf{G}'$ by assuming that the first $(B - N + 1)$ symbols of $\mathbf{G}'$ ($\{x_0, \ldots, x_{B-N}\}$) were received without errors (equivalently assume that $\{s_0, \ldots, s_{B-N}\} = \{0, \ldots, 0\}$). Even though $\mathbf{G}'$ is not systematic, due to its structure receiving $\{x_0, \ldots, x_{B-N}\}$ without errors means that $\{s_0, \ldots, s_{B-N}\}$ can be decoded correctly and hence can be cancelled from the other received symbols. Therefore, the remaining code is an $(n-(B-N+1), k-(B-N+1))$ MDS code over $\mathbb{F}_q$.[2]

Following Properties 2 and 3 we denote the codes induced by $\mathbf{H}_1$ and $\mathbf{H}_2$ as MDS_1 and MDS_2.

Remark 1. *Using (5) it follows that MDS_2 is a $(T, T - B)$ MDS code.*

Remark 2. *In case $T > B$, symbols $\{x_B, \ldots, x_{T-1}\}$ contain information only from $\{s_{B-N+1}, \ldots, s_{k-1}\}$, i.e., these are the only symbols of MDS_2 without interference from $\{s_0, \ldots, s_{B-N}\}$ (which are not data symbols of MDS_2).*

Remark 3. *Although G is not a systematic generator matrix, we note that $\hat{\mathbf{G}} = \mathbf{M}^{-1}\mathbf{G}$ is a systematic generator matrix. Since $\mathbf{M}$ is an invertible upper triangular matrix, it can be shown that $\hat{\mathbf{G}}$ also results in a capacity-achieving streaming code that is systematic.*

The following Theorem is proved in Section V.

[2] This can be also viewed as shortening MDS code $\mathbf{C}'$ by $(B - N + 1)$ data symbols.

Theorem 1. *Block code $\mathcal{C}$ with generator matrix $\mathbf{G}$ is a block code which conforms to $\mathcal{C}(W,B,N)$ with a delay-constraint T and thus a capacity-achieving streaming code of any $\mathcal{C}(W,B,N)$ with delay T and field size that scales quadratically with the delay constraint ($O(T^2)$) can be generated from $\mathcal{C}$ using diagonal interleaving.*

IV. Example

As an example we take the case where $B=4$, $N=3$ and $T=6$. The generator matrix of the resulting code is:

$$\begin{pmatrix} 1 & X & X & 0 & 0 & 0 & \alpha & 0 \\ 0 & 1 & X & X & 0 & 0 & 0 & \alpha \\ 0 & 0 & 1 & X & X & 0 & X & X \\ 0 & 0 & 0 & 1 & X & X & X & X \end{pmatrix},$$

where

- MDS_1 is a $(6,4)$ MDS code.
- MDS_2 is a $(6,2)$ MDS code.

We demonstrate the decoding process for several cases of erasures. We focus on decoding symbol s_0. As decoding symbol s_0 when symbol x_0 is not earased is immediate we focus only on cases where x_0 is erased.

- A burst of size $B=4$ starting at time 0

$$\begin{pmatrix} 1 & X & X & 0 & 0 & 0 & \alpha & 0 \\ 0 & 1 & X & X & 0 & 0 & 0 & \alpha \\ 0 & 0 & 1 & X & X & 0 & X & X \\ 0 & 0 & 0 & 1 & X & X & X & X \end{pmatrix},$$

  Using MDS_2, s_2 and s_3 can be decoded at time 5 since we have two linear independent symbols from a $(6,2)$ MDS code (x_4 and x_5). These symbols are cancelled from x_6 to recover s_0.
- $N=3$ sporadic erasures where x_6 is erased:

$$\begin{pmatrix} 1 & X & X & 0 & 0 & 0 & \alpha & 0 \\ 0 & 1 & X & X & 0 & 0 & 0 & \alpha \\ 0 & 0 & 1 & X & X & 0 & X & X \\ 0 & 0 & 0 & 1 & X & X & X & X \end{pmatrix}.$$

  Using MDS_1, all data symbols can be decoded at time 4 since we have four linear independent symbols from a $(6,4)$ MDS code.
- $N=3$ sporadic erasures where x_6 is not erased:

$$\begin{pmatrix} 1 & X & X & 0 & 0 & 0 & \alpha & 0 \\ 0 & 1 & X & X & 0 & 0 & 0 & \alpha \\ 0 & 0 & 1 & X & X & 0 & X & X \\ 0 & 0 & 0 & 1 & X & X & X & X \end{pmatrix}.$$

  We note the (3×5) lower right matrix of $\mathbf{H}_1$ (marked as dashed blue) above is a $(5,3)$ MDS code which is the outcome of "shortening" $\mathbf{H}_1$ by one symbol (denoted as MDS_1^1). Receiving symbols $\{x_1,x_2,x_3\}$ with no erasures means that data symbols $\{s_1,s_2,s_3\}$ can be decoded from MDS_1^1 with known interference from s_0 (denoted as $\{\tilde{s}_1,\tilde{s}_2,\tilde{s}_3\}$). Than $\tilde{s}_2$ and $\tilde{s}_3$ can be cancelled from x_6. Since we assume $\alpha\in\mathbb{F}_{q^2}\setminus\mathbb{F}_q$, it is guaranteed that α is not nulled out hence s_0 can be recovered.
  Alternatively, the dashed part of $\mathbf{g}_6$ (interference from $\{s_0,s_1,s_2\}$) is in the span of MDS_1^1. Since we have three symbols from MDS_1^1 ($\{x_1,x_2,x_3\}$) the dashed part of $\mathbf{g}_6$ can be cancelled. Since we assumed $\alpha\in\mathbb{F}_{q^2}\setminus\mathbb{F}_q$, it is guaranteed that α is not nulled out hence s_0 can be recovered.

The decoding of $s_i\in\{s_1,s_2,s_3\}$ is done in a similar manner where we assume by induction that $\{s_0,\ldots,s_{i-1}\}$ have already been recovered by time $T+i$.

V. Proof Of Theorem 1

We prove next that the construction described above conforms to $\mathcal{C}(W,B,N)$ with a delay-constraint T. The field size of this construction is q^2 where $q=n$. Recalling (5) we conclude that the field size of the suggested construction is $O(T^2)$.

- Decoding data symbols $\{s_0,\ldots,s_{B-N}\}$
  We analyze the two different types of erasures:
  - A Burst of length B starting at time i
    Decoding data symbol s_0
    Following Property 3 and Remark 1 we recall that MDS_2 is a $(T,T-B)$ MDS code. A burst of B symbols starting at time 0 means that symbols $\{x_B,\ldots,x_{T-1}\}$ are not erased. Recalling Remark 2, these symbols don't have an interference from data symbols $\{s_0,\ldots,s_{B-N}\}$. Therefore, data symbols $\{s_{B-N+1},\ldots,s_k\}$ can be recovered using MDS_2 and cancelled from symbol x_T to recover data symbol s_0.[3]
    Decoding data symbols $\{s_1,\ldots,s_{B-N}\}$
    we again argue the recovery in two steps:
    * Recovery of $\{s_{B-N+1},\ldots,s_k\}$ using MDS_2.
    * Recovery of s_i using x_{T+i} by canceling the effect of $\{s_{B-N+1},\ldots,s_k\}$.

    We first assume by induction that $\{s_0,\ldots,s_{i-1}\}$ have already been recovered by time $T+i$ and claim that $\{x_{i+B},\ldots,x_{i+T-1}\}$, which are non-erased following a burst erasure starting at time i, suffice to recover the symbols in the first step. Then, $\{s_{B-N+1},\ldots,s_k\}$ are cancelled from symbol x_{T+i} to recover data symbol s_i.
  - N arbitrary erasures
    Decoding data symbol s_0
    First, we note that we assume that symbol x_0 is one of the erased symbols otherwise decoding is trivial.[4]
    We further differentiate between the following two cases

[3] In case $T=B$ it can be shown that $\mathbf{H}_3=\alpha\cdot\mathbf{I}_{B-N+1}$ hence s_0 can be recovered directly from x_T.

[4] If $N=1$ it means that the only erasure is that of symbol x_0 and hence decoding is done as described next for the case when symbol x_T is not erased.

* Symbol x_T is erased
  We note that in this case, in MDS_1 we have $N-1$ erasures. Following Property 2, MDS_1 can correct any $N-1$ erasures and hence data symbol s_0 can be recovered.
* Symbol x_T is not erased
  Note that the $(k-1)\times(k-1+N-1)$ lower right submatrix of $\mathbf{H}_1$ (marked as the dashed matrix below) is a $(k-1+N-1, k-1)$ MDS code (can be viewed as "shortening" MDS_1 by one data symbol) and we denote it as MDS_1^1.

$$\mathbf{H}_1$$

$$\begin{array}{c}\left[\begin{array}{c|cccccccc|c}
1 & X & \cdots & X & 0 & 0 & 0 & \cdots & 0 & \alpha \\
0 & 1 & X & \cdots & X & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & & 1 & \ddots & \ddots & \ddots & & \vdots & X \\
\vdots & \vdots & & & \ddots & X & \cdots & X & 0 & \vdots \\
0 & 0 & \cdots & \cdots & 0 & 1 & X & \cdots & X & X
\end{array}\right] \\
\qquad \underbrace{}_{\mathbf{g}_0} \qquad \underbrace{}_{\mathbf{g}_1,\ldots,\mathbf{g}_{T-1}} \qquad \underbrace{}_{\mathbf{g}_T}
\end{array}$$

  Since we may assume that symbols $\{x_1,\ldots,x_{T-1}\}$ have up to $N-1$ erasures, using MDS_1^1, data symbols $\{s_1,\ldots,s_k\}$ can be recovered up to a known interference from s_0 (denoted as $\{\tilde{s}_1,\ldots,\tilde{s}_k\}$). Since $\alpha\in\mathbb{F}_{q^2}\setminus\mathbb{F}_q$, $\{\tilde{s}_1,\ldots,\tilde{s}_{B-N}\}$ can be cancelled from symbol x_T while it is guaranteed that α is not nulled out[5] and thus symbol s_0 can be recovered.
  Alternatively we note that the dashed part of $\mathbf{g}_T$ is in the span of MDS_1^1 (and further can be denoted as linear combination from the base field of the symbols of MDS_1). Since we have enough linear independent columns from MDS_1^1, the dashed part of $\mathbf{g}_T$ can be cancelled while it is guaranteed that $\alpha\in\mathbb{F}_{q^2}\setminus\mathbb{F}_q$, is not nulled out and thus data symbol s_0 can be recovered.

Decoding data symbols $\{s_1,\ldots,s_{B-N}\}$
We first assume by induction that $\{s_0,\ldots,s_{i-1}\}$ have already been recovered by time $T+i$. We further assume that their impact on symbols $\{x_T,\ldots,x_{T+i}\}$ is cancelled. After cancelling symbols $\{s_0,\ldots,s_{i-1}\}$ from MDS_1 we are left with a $(k-i+N-1,k-i)$ MDS code (which can recover $N-1$ erasures). Equivalently, this can be viewed as "shortening" MDS_1 by i symbols. We denote it as MDS_1^i.
Further, we may assume that symbol x_i is one of the erased symbols, or otherwise decoding of data symbol s_i is trivial (since we assumed $\{s_0,\ldots,s_{i-1}\}$ have been decoded correctly). We differentiate again between the following two cases:

* Symbol x_{T+i} is erased
  Assuming x_{T+i} is erased means that there are at most $N-1$ erasures in MDS_1^i (which can recover $N-1$ erasures). Thus all data symbols can be decoded up to time $T+i$.
* Symbol x_{T+i} not erased
  We note that the lower right $k-(i+1)\times k-(i+1)+N-1$ sub-matrix of $\mathbf{H}_1$ is also $(k-(i+1)+N-1, k-(i+1))$ MDS code (can be also viewed as "shortening" MDS_1^i by one symbol therefore we denote it as MDS_1^{i+1}). We may assume that symbols $\{x_{i+1},\ldots,x_{T-1}\}$ have up to $N-1$ erasures hence MDS_1^{i+1} can decode data symbols $\{s_{i+1},\ldots,s_k\}$ up to a (known) interference from data symbol s_i (denoted by $\{\tilde{s}_{i+1},\ldots,\tilde{s}_k\}$. Next, $\{\tilde{s}_{i+1},\ldots,\tilde{s}_{B-N}\}$ are cancelled from symbol $T+i$ and, again, it is guaranteed that $\alpha\in\mathbb{F}_{q^2}\setminus\mathbb{F}_q$ is not cancelled, thus s_i can be recovered.

* Decoding data symbol $\{s_{B-N+1},\ldots,s_{k-1}\}$
  We first assume by induction that $\{s_0,\ldots,s_{B-N}\}$ have already been recovered by time $T+i$ and cancelled from the received symbols. This means that we are left with MDS_2. Recalling Property 3 MDS_2 is a $(k-(B-N+1)\times(n-(B-N+1)$ MDS code which means it can correct any B erasures. Recalling that $B\geq N$ it means that either a burst of B erasures or arbitrary N erasures can be correctly decoded at time $T+i$.

[5]Since all elements in MDS_2 belong to $\mathbb{F}_q$, the cancellation is done by multiplying $\{\tilde{s}_1,\ldots,\tilde{s}_{B-N}\}$ with coefficients from the base field.

References

[1] E. Martinian and C.-E. Sundberg, "Burst erasure correction codes with low decoding delay," *IEEE Transactions on Information theory*, vol. 50, no. 10, pp. 2494–2502, 2004.

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

[3] S. L. Fong, A. Khisti, B. Li, W.-T. Tan, X. Zhu, and J. Apostolopoulos, "Optimal streaming codes for channels with burst and arbitrary erasures," *IEEE Transactions on Information Theory*, 2019.

[4] M. N. Krishnan and P. V. Kumar, "Rate-optimal streaming codes for channels with burst and isolated erasures," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1809–1813.

[5] D. Dudzicz, S. L. Fong, and A. Khisti, "An explicit construction of optimal streaming codes for channels with burst and arbitrary erasures," *CoRR*, vol. abs/1903.07434, 2019. [Online]. Available: http://arxiv.org/abs/1903.07434

[6] M. N. Krishnan, D. Shukla, and P. V. Kumar, "Low field-size, rate-optimal streaming codes for channels with burst and random erasures," *CoRR*, vol. abs/1903.06210, 2019. [Online]. Available: http://arxiv.org/abs/1903.06210

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