On Coding for Real-Time Streaming under Packet Erasures

Derek Leong, Asma Qureshi, and Tracey Ho
Department of Electrical Engineering
California Institute of Technology
Pasadena, California 91125, USA
{derekleong,aqureshi,tho}@caltech.edu

Abstract—We consider a real-time streaming system where messages created at regular time intervals at a source are encoded for transmission to a receiver over a packet erasure link; the receiver must subsequently decode each message within a given delay from its creation time. We study a bursty erasure model in which all erasure patterns containing erasure bursts of a limited length are admissible. For certain classes of parameter values, we provide code constructions that asymptotically achieve the maximum message size among all codes that allow decoding under all admissible erasure patterns. We also study an i.i.d. erasure model in which each transmitted packet is erased independently with the same probability; the objective is to maximize the decoding probability for a given message size. We derive an upper bound on the decoding probability for any time-invariant code, and show that the gap between this bound and the performance of a family of time-invariant intrasession codes is small in the high reliability regime.

Index Terms—Erasure correction, real-time streaming.

I. INTRODUCTION

We consider a real-time streaming system where messages created at regular time intervals at a source are encoded for transmission to a receiver over a packet erasure link; the receiver must subsequently decode each message within a given delay from its creation time.

Two erasure models are studied in this paper. The first is a bursty erasure model in which all erasure patterns containing erasure bursts of a limited length are admissible. The objective is to find a code that achieves the maximum message size, among all codes that allow decoding under all admissible erasure patterns. The second is an i.i.d. erasure model in which each transmitted packet is erased independently with the same probability. The objective here is to maximize the decoding probability for a given message size.

In our previous work [1], we showed that a time-invariant intrasession code is asymptotically optimal over all codes (time-varying and time-invariant, intersession and intrasession) as the number of messages goes to infinity, for a sliding window erasure model, and for the bursty erasure model when the maximum erasure burst length is sufficiently short or long.

Intrasession coding is attractive due to its relative simplicity (it allows coding within the same message but not across different messages), but it is not known in general when intrasession coding is sufficient or when intersession coding is necessary.

Our Contribution: For the bursty erasure model, we show that diagonally interleaved codes derived from specific systematic block codes are asymptotically optimal over all codes for certain classes of parameter values. For the i.i.d. erasure model, we derive an upper bound on the decoding probability for any time-invariant code, and show that the gap between this bound and the performance of a family of time-invariant intrasession codes is small in the high reliability regime.

Related Work: Martinian et al. [2], [3] and Badr et al. [4] provide constructions of streaming codes that minimize the decoding delay for certain types of bursty erasure models. Tree codes or anytime codes, for which the decoding failure probability decays exponentially with delay, are considered in [5]–[7]. Tekin et al. [8] considered erasure correction coding for a non-real-time streaming system where all messages are initially present at the source.

We begin with a formal definition of the problem in Section II. In Sections III and IV, we present our results for the bursty erasure model and the i.i.d. erasure model, respectively. Proofs of theorems are deferred to the appendix.
II. PROBLEM DEFINITION

Consider a discrete-time data streaming system comprising a source and a receiver, with a directed unit-bandwidth packet erasure link from the source to the receiver. Independent messages of uniform size $s > 0$ are created at regular intervals of $c \in \mathbb{Z}^+$ time steps at the source. At each time step $t \in \mathbb{Z}^+$, the source transmits a single data packet of normalized unit size over the packet erasure link; either the entire packet is received instantaneously by the receiver at time step $t$, or the entire packet is erased and never received. The receiver must subsequently decode each message within a delay of $d \in \mathbb{Z}^+$ time steps from its creation time. Fig. 1 depicts this real-time streaming system for an instance of $(c, d)$.

More precisely, let random variable $M_k$ denote message $k$; the random variables $\{M_k\}$ are independent, and $H(M_k) = s$ for each $k \in \mathbb{Z}^+$. To simplify our definition of the encoding functions, we shall further assume that $M_1, M_2, \ldots$ are identically distributed, and nonpositive messages $M_0, M_{-1}, \ldots$ are zeros.

Each message $k \in \mathbb{Z}^+$ is created at time step $(k-1)c+1$, and is to be decoded by time step $(k-1)c+d$. Let $W_k$ be the coding window for message $k$, which we define as the interval of $d$ time steps between its creation time and decoding deadline, i.e., $W_k = ((k-1)c+1, \ldots, (k-1)c+d)$. We shall assume that $d > c$ so as to avoid the degenerate case of nonoverlapping coding windows for which it is sufficient to code individual messages separately.

The unit-size packet transmitted at each time step $t \in \mathbb{Z}^+$ must be a function of messages created at time step $t$ or earlier. Let random variable $X_t$ denote the packet transmitted at time step $t$; we have $H(X_t) \leq 1$ for each $t \in \mathbb{Z}^+$. For brevity, we define $X[A] = \{X_t\}_{t \in A}$.

Because we are dealing with hard message decoding deadlines and fixed-size messages and packets, we consider a given message $k$ to be decodable from the packets received at time steps $t \in A$ if and only if $H(M_k | X[A]) = 0$.

Consider the first $n$ messages $\{1, \ldots, n\}$, and the union of their (overlapping) coding windows $T_n$ given by

$$T_n = W_1 \cup \cdots \cup W_n = \{1, \ldots, n-1\}c+d.$$ 

An erasure pattern $E \subseteq T_n$ specifies a set of erased packet transmissions over the link; the packets transmitted at time steps $t \in E$ are erased, while those transmitted at time steps $t \in T_n \setminus E$ are received. An erasure model essentially describes a distribution of erasure patterns.

For a given pair of positive integers $a$ and $b$, we define the offset quotient $q_{a,b}$ and offset remainder $r_{a,b}$ to be the unique integers satisfying $a = q_{a,b}b + r_{a,b}$, $q_{a,b} \in \mathbb{Z}_a^+$, and $r_{a,b} \in \{1, \ldots, b\}$, where $\mathbb{Z}_a^+$ denotes the set of nonnegative integers, i.e., $\mathbb{Z}^+ \cup \{0\}$. Note that this definition departs from the usual definition of quotient and remainder in that $r_{a,b}$ can be equal to $b$ but not zero.

III. BURSTY ERASURE MODEL

In this section, we look at erasure patterns that contain erasure bursts of a limited length. Consider the first $n$ messages $\{1, \ldots, n\}$, and the union of their (overlapping) coding windows $T_n$. Let $E_n^a$ be the set of erasure patterns in which each erasure burst is an interval of at most $z$ erased time steps, and consecutive erasure bursts are separated by a guard interval or gap of at least $d-z$ unerased time steps, i.e.,

$$E_n^a \triangleq \left\{ E \subseteq T_n : \begin{array}{l}
(t \notin E \land t+1 \in E) \Rightarrow |E \cap \{t+1, \ldots, t+z+1\}| \leq z, \\
(t \in E \land t+1 \notin E) \Rightarrow |E \cap \{t+1, \ldots, t+d-z\}| = 0 \end{array} \right\}.$$ 

The objective is to construct a code that allows all $n$ messages $\{1, \ldots, n\}$ to be decoded by their respective decoding deadlines under any erasure pattern $E \in E_n^a$. Let $s_n^a$ be the maximum message size that can be achieved by such a code, for a given choice of $(n, c, d, z)$.

This model can be seen as an instance of a more general class of bursty erasure models where the maximum erasure burst length and the minimum guard interval length can be arbitrarily specified. In a similar bursty erasure model considered by Martinian et al. [2], [3] and Badr et al. [4], the maximum erasure burst length (given by $B$) is $z$, while the minimum guard interval length (given by $T$) is $d-1$. For the same choice of $(d, z)$, our model captures a larger set of erasure patterns and is therefore stricter (the respective cut-set bounds reflect this comparison).

In [1], we showed that a time-invariant intrasession code is asymptotically optimal when $d$ is a multiple of $c$, or when the maximum erasure burst length $z$ is sufficiently short or long. Here, we present a family of time-invariant intersession codes that are asymptotically optimal in several other cases.

A. Diagonally Interleaved Codes

Consider a systematic block code $C$ that encodes a given vector of $d-\alpha$ information symbols $a = (a[1], \ldots, a[d-\alpha])$ as a codeword vector of $d$ symbols $(a[1], \ldots, a[d-\alpha], b[1], \ldots, b[\alpha])$, where each symbol has a normalized size of $\frac{1}{d}$. For each $i \in \{1, \ldots, \alpha\}$, we define an encoding function $g_i$ so that the parity symbol $b[i]$ is given by $b[i] = g_i(a)$.

For a given choice of $(c, d, \alpha)$, we can derive a time-invariant diagonally interleaved code for a message size of $s = \frac{d-\alpha}{c}$ by interleaving codeword symbols produced by the component systematic block code $C$ in a diagonal pattern.

First, to facilitate code construction, we represent the derived code by a table of symbols, with each cell in the table assigned one symbol of size $\frac{1}{d}$. Fig. 2 illustrates our construction for an instance of $(c, d, \alpha)$. Let $x[i]$ denote the symbol in column $i$ of $Z$ and row $i \in \{1, \ldots, d\}$. The unit-size packet transmitted at each time step $t$ is composed of the $d$ symbols $x[i][1], \ldots, x[i][d]$ in column $t$ of the table. Rows $\{1, \ldots, d-\alpha\}$ of the table are populated by information symbols, while rows $\{d-\alpha+1, \ldots, d\}$ are populated by parity symbols.

Next, we divide each message $k$ into $(d-\alpha)c$ submessages or information symbols denoted by $M_k[1], \ldots, M_k[(d-\alpha)c]$,
with each symbol having a size of \( \frac{c-z}{d-\alpha}\). The information symbols corresponding to each message \( k \) are assigned evenly to the columns representing the first \( c \) time steps in coding window \( W_k \), so that

\[
x_t[i] = M_{q_t,c+1}[r_t,c-1](d-\alpha) + i
\]

for each \( i \in \{1, \ldots, d-\alpha\} \). To obtain the parity symbols for column \( t \), we apply the component systematic block code \( C \) to the information symbols on each diagonal, so that

\[
x_t[d-\alpha + i] = g_t\left(x_{t-i-(d-\alpha)+\ell}\right)_{\ell=1}^{d-\alpha}
\]

for each \( i \in \{1, \ldots, \alpha\} \). Thus, the \( d \) symbols on each diagonal spanning across \( d \) consecutive time steps constitute one codeword produced by \( C \). Note that the information symbols for nonexistent messages (i.e., nonpositive messages and messages after the actual final message) are assumed to be zeros so that all codeword symbols are well defined.

**B. Optimality of Diagonally Interleaved Codes**

Diagonally interleaved codes that are derived from systematic block codes \( C \) with certain properties turn out to be asymptotically optimal in certain cases. These sufficient code properties are given by the following lemma:

**Lemma 1.** Consider the diagonally interleaved code for a given choice of \( (c, d, \alpha = z) \) satisfying \( c \leq z \leq d - c \).

Suppose that the \( d \) symbols of the codeword vector \( (a[1], \ldots, a[d-z], b[1], \ldots, b[z]) \) produced by the component systematic block code \( C \) are transmitted sequentially across an erasure link, one symbol per time step, over the time interval \( L = \{1, \ldots, d\} \). For each \( j \in \{1, \ldots, d\} \), let \( E_j^L \subseteq L \) be the erasure pattern that contains a single wrap-around erasure burst of exactly \( z \) erased time steps (which may wrap around the last and first time steps in the interval) with the \( j \)th time step in the interval as the “leading” erasure, i.e.,

\[
E_j^L = \{r_j + \ell : d : \ell \in \{0, \ldots, z-1\}\}
\]

Let \( E^2 \) be the set of all such erasure patterns, i.e.,

\[
E^2 = \{E_j^L : j \in \{1, \ldots, d\}\}
\]

If the systematic block code \( C \) satisfies both of the following symbol decoding requirements, then the diagonally interleaved code derived from \( C \) achieves a message size of \( \frac{z}{d-z}c \) for the bursty erasure model:

**D1** For each \( i \in \{1, \ldots, c\} \), the information symbol \( a[i] \) is decodable by the \((d-c+i)\)th time step in interval \( L \) under any erasure pattern \( E_j^L \in E^2 \).

**D2** The information symbols \( a[c+1], \ldots, a[d-z] \) are decodable by the last time step in interval \( L \) under any erasure pattern \( E_j^L \in E^2 \).

The condition \( c \leq z \leq d - c \) is actually implied by the symbol decoding requirements: the first information symbol \( a[1] \) would otherwise be undecodable by its decoding deadline under erasure pattern \( E_j^L \) because by that time step, no parity symbols would have been transmitted if \( c > z \), and no symbols
would have been received if $z > d - c$. Note that the use of a systematic MDS code as the component systematic block code $C$ may not be sufficient here because of the additional decoding deadlines imposed on individual symbols.

The following theorem shows that a degenerate diagonally interleaved code that uses only intrasession coding is asymptotically optimal over all codes for the specified parameter conditions:

**Theorem 1.** Consider the bursty erasure model for a given choice of $(c, d, z)$ satisfying all of the following three conditions:

1) $d$ is not a multiple of $c$;
2) $c \leq z \leq d - c$; and
3) $d$ is a multiple of $d - z$.

Let $C$ be a systematic block code that encodes a given vector of $d - z$ information symbols $\mathbf{a} = (a[1], \ldots, a[d-z])$ as a codeword vector of $d$ symbols $(a[1], \ldots, a[d-z], b[1], \ldots, b[z])$, where each symbol has a normalized size of $\frac{1}{d}$, and the parity symbol $b[i]$ is given by

$$b[i] = g_i(\mathbf{a}) \triangleq a[r_i, d-z]$$

for each $i \in \{1, \ldots, z\}$. The diagonally interleaved code derived from $C$ is asymptotically optimal over all codes in the following sense: it achieves a message size of $\frac{d-z}{d-c}c$, which is equal to the asymptotic maximum achievable message size $\lim_{n \to \infty} S^*_{n}$.
Fig. 5. The $d$ symbols of the codeword vector produced by the systematic block code $C$ of Theorem 3, for (a) $(c, d, z) = (5, 84, 60)$ and (b) $(c, d, z) = (5, 57, 42)$. In (a), because $r' = z'$, there are no virtual information symbols. In (b), because $r' < z'$, we have virtual information symbols on the $(d - z - r'z + 1)$th row (in parentheses). For each $i \in \{1, \ldots, z'\}$, the value of the nondegenerate parity symbol $b[i]$ is given by the bit-wise modulo-2 sum (i.e., exclusive-or) of the actual and virtual information symbols above it in column $i$ of the table.

The systematic block code $C$ of Theorem 1 is illustrated in Fig. 3 for an instance of $(c, d, z)$. Note that all the parity symbols in $C$ are degenerate in the sense that they are just uncoded copies of information symbols.

The following two theorems describe diagonally interleaved codes that are asymptotically optimal over all codes for the specified parameter conditions:

**Theorem 2.** Consider the bursty erasure model for a given choice of $(c, d, z)$ satisfying all of the following five conditions:

1) $d$ is not a multiple of $c$;
2) $c \leq z \leq d - c$;
3) $d$ is not a multiple of $d - z$;
4) $z < d - z$; and
5) $z$ is a multiple of $r'$, where

$$r' \triangleq r_{d-z,z} \in \{1, \ldots, z\}.$$
symbol $b[i]$ is given by

\[ b[i] = g_i(a) \triangleq \left( \bigoplus_{k=1}^{d-z-r'} a[(k-1)z+i] \right) \oplus a[d-z-r'+r_i,r'] \]

for each $i \in \{1, \ldots, z\}$. The diagonally interleaved code derived from $C$ is asymptotically optimal over all codes in the following sense: it achieves a message size of $\frac{d-z}{d}c$, which is equal to the asymptotic maximum achievable message size $\lim_{n \to \infty} s^n_b$.

The systematic block code $C$ of Theorem 2 is illustrated in Fig. 4 for two instances of $(c, d, z)$. The following example demonstrates that in this case, intrasession coding can be strictly suboptimal:

**Example 1** (Suboptimality of Intrasession Coding). Suppose that $(c, d, z) = (2, 5, 2)$. For $n = 9$, the maximum message size that can be achieved by an intrasession code has an upper bound of $s < 1.193$; such a bound can be found by solving a linear program for a subset of erasure patterns in $E^n_s$ (namely, those with alternating intervals of $z$ erased time steps and $d - z$ unerased time steps). The same upper bound also holds for $n > 9$ because any message size that can be achieved for a larger number of messages can also be achieved for a smaller number of messages (we simply apply the same code and ignore the additional messages and packets). On the other hand, the diagonally interleaved code derived from the systematic block code $C$ of Theorem 2 achieves a strictly larger message size of $s = \frac{9}{5} = 1.2$.

**Theorem 3.** Consider the bursty erasure model for a given choice of $(c, d, z)$ satisfying all of the following five conditions:

1. $d$ is not a multiple of $c$;
2. $c \leq z \leq d - c$;
3. $d$ is not a multiple of $d - z$;
4. $z > d - z$; and
5. $z'$ is a multiple of $r'$, where

\[ z' \triangleq r_z, d - z \in \{1, \ldots, d - z - 1\}, \]
\[ r' \triangleq r_{d - z, z'} \in \{1, \ldots, z'\}. \]

Let $C$ be a systematic block code that encodes a given vector of $d - z$ information symbols $a = (a[1], \ldots, a[d-z])$ as a codeword vector of $d$ symbols $(a[1], \ldots, a[d-z], b[1], \ldots, b[z])$, where each symbol has a normalized size of $\frac{1}{d}$, and the parity symbol $b[i]$ is given by

\[ b[i] = g_i(a) \triangleq \left\{ \begin{array}{ll}
  \bigoplus_{k=1}^{d-z-r'} a[(k-1)z+i] \\
  a[d-z-r'+r_i,r'] & \text{if } i \in \{1, \ldots, z'\}, \\
  a[r_i-z',d-z] & \text{if } i \in \{z' + 1, \ldots, z\}
\end{array} \right. \]

for each $i \in \{1, \ldots, z\}$. The diagonally interleaved code derived from $C$ is asymptotically optimal over all codes in the following sense: it achieves a message size of $\frac{d-z}{d}c$, which is equal to the asymptotic maximum achievable message size $\lim_{n \to \infty} s^n_b$.

The systematic block code $C$ of Theorem 3 is illustrated in Fig. 5 for two instances of $(c, d, z)$. Note that there are two types of parity symbols in $C$: $b[1], \ldots, b[z]$ are nondegenerate parity symbols, while $b[z'+1], \ldots, b[z]$ are degenerate parity symbols which are just uncoded copies of information symbols. The following example demonstrates that in this case, intrasession coding can be strictly suboptimal:

**Example 2** (Suboptimality of Intrasession Coding, cf. Example 1). Suppose that $(c, d, z) = (3, 8, 5)$. For $n \geq 10$, the maximum message size that can be achieved by an intrasession code has an upper bound of $s < 1.118$. On the other hand, the diagonally interleaved code derived from the systematic block code $C$ of Theorem 3 achieves a strictly larger message size of $s = \frac{9}{8} = 1.125$.

**IV. IID Erasure Model**

In this section, we consider a random erasure model in which each packet transmitted over the link is erased independently with the same probability $p_k$. For brevity, let $S_k$ denote the successful event “message $k$ is decodable by its decoding deadline, i.e., time step $(k-1)c + i$” and let $\bar{S}_k$ denote the complementary failure event. We restrict our attention to time-invariant codes here in the interest of practicality.

**Definition** (Time-Invariant Code). A code is time-invariant if there exist causal and deterministic encoding functions $f_1, \ldots, f_c$ and a finite encoder memory size $m_e \in \mathbb{Z}^+$ such that the packet transmitted at each time step $(k-1)c + i$, where $k \in \mathbb{Z}^+, i \in \{1, \ldots, c\}$, is given by the function $f_i$ applied to the $m_e$ most recent messages, i.e.,

\[ X_{(k-1)c+i} = f_i(M_k, M_{k-1}, \ldots, M_{k-m_e+1}). \]

Consider the i.i.d. erasure model for a given choice of $(c, d, p_k, s)$. We shall adopt the decoding probability $\mathbb{P}[S_k]$, i.e., the probability that a given message $k$ is decodable by its decoding deadline, as the primary performance metric. The decoder memory size is assumed to be unbounded so that the decoder has access to all received packets. Let the random subset $U_k \subseteq T_k$ be the unerased time steps that are no later than the decoding deadline for message $k$; the received packets that can be used by the decoder for decoding message $k$ are therefore given by $X[U_k]$. Consequently, the decoding probability $\mathbb{P}[S_k]$, where $k \in \mathbb{Z}^+$, can be expressed in terms of $U_k$ as follows:

\[
\mathbb{P}[S_k] = \mathbb{P}[H(M_k | X[U_k]) = 0] = \sum_{U_k \subseteq T_k} \mathbb{P}[H(M_k | X[U_k]) = 0] \cdot (1 - p_k)^{|U_k|} (p_k)^{|T_k| - |U_k|}.
\]

By combining the proof techniques of [1] and [9, Lemma 1], we can derive an upper bound on the decoding probability $\mathbb{P}[S_k]$ for any time-invariant code:

**Theorem 4.** Consider the i.i.d. erasure model for a given choice of $(c, d, p_k, s)$. For any time-invariant code with encoder
memory size \( m_e \), the probability that a given message \( k \geq m_e \) is decodable by its decoding deadline is upper-bounded as follows:

\[
\mathbb{P}[S_k] \leq \sum_{z=0}^{d} \min\left(\left(\frac{(d-z)c}{d}s, 1\right), \left(\frac{d}{z}\right)\right) \left(1 - p_s\right)^{d-z}(p_s)^z.
\]

(2)

Note that the decoding probability \( \mathbb{P}[S_k] \) for the early messages \( k < m_e \) can potentially be higher than that for the subsequent messages \( k \geq m_e \) because the decoder already knows the nonpositive messages (which are assumed to be zeros).

For real-time streaming applications that are sensitive to bursts of decoding failures, it may be useful to adopt the burstiness of undecodable messages as a secondary performance metric. One way of measuring this burstiness is to compute the conditional probability \( \mathbb{P}[S_k | S_{k-1}] \), i.e., the conditional probability that the next message is undecodable by its decoding deadline given that the current message is undecodable by its decoding deadline. The higher this conditional probability is, the more likely it is to remain "stuck" in a burst of undecodable messages.

### A. Performance of Symmetric Codes

In [1], we introduced a time-invariant intrasession code that divides each packet evenly among all active messages (a message \( k \) is active at time step \( t \) if and only if \( t \) falls within its coding window, i.e., \( t \in W_k \)). An appropriate code (e.g., random linear coding, MDS code) is then applied to the allocation of packet space so that each message can be decoded whenever the total amount of received data that encodes that message is at least the message size \( s \). Here, we generalize this code construction to obtain a family of symmetric time-invariant intrasession codes. For each symmetric code, we define a spreading parameter \( m \in \{c, \ldots, d^e\} \), where \( d^e \triangleq \min(d, m_e) \), so that the effective coding window for message \( k \) is given by

\[
W_k^e \triangleq \{(k-1)c+1, \ldots, (k-1)c+m\}.
\]

We subsequently redefine active messages in terms of the effective coding window instead of the coding window (which corresponds to \( m = d \)).

1) Decoding Probability: Consider the decodability of a given message \( k \in \mathbb{Z}^+ \) for the symmetric code with spreading parameter \( m \). Suppose that \( V_b \) small blocks and \( V_b \) big blocks that encode message \( k \) are received by the decoder; \( V_b \) and \( V_b \) are independent binomial random variables with the following distributions:

\[
V_b \sim \text{Binomial}\left((q_{m,c}+1)r_{m,c}, 1 - p_s\right),
V_b \sim \text{Binomial}\left(q_{m,c}(c - r_{m,c}), 1 - p_s\right).
\]

The decoding probability \( \mathbb{P}[S_k] \) can therefore be expressed in terms of these random variables as follows:

\[
\mathbb{P}[S_k] = \begin{cases} 
\mathbb{P}\left[\frac{1}{q_{m,c}+1}V_b \geq s\right] & \text{if } r_{m,c} = c, \\
\mathbb{P}\left[\frac{1}{q_{m,c}+1}V_b + \frac{1}{q_{m,c}}V_b \geq s\right] & \text{otherwise}
\end{cases}
\]

(\*)

Figs. 6a and 6b show how the family of symmetric codes perform in terms of the decoding probability \( \mathbb{P}[S] \), for an instance of \( (c, d, m_e) \). These plots and other empirical observations suggest that 1) maximal spreading (i.e., \( m = d^e \)) performs well, i.e., achieves a relatively high \( \mathbb{P}[S] \), when the message size \( s \) and the packet erasure probability \( p_s \) are small, while minimal spreading (i.e., \( m = c \)) performs well when \( s \) and \( p_s \) are large (this echoes the analytical findings of [9]); and 2) although this family of codes may not always contain an optimal time-invariant intrasession code, we can find good codes (with decoding probabilities close to the upper bound of Theorem 4) among them when \( s \) and \( p_s \) are small.

2) Burstiness of Undecodable Messages: Consider the decodability of a given pair of consecutive messages \( k \) and \( k+1 \), where \( k \in \mathbb{Z}^+ \), for the symmetric code with spreading parameter \( m \). The \( 2m \) blocks that encode the pair of messages are spread over the \( m + c \) time steps in the union of the two effective coding windows, i.e.,

\[
\{(k-1)c+1, \ldots, kc+m\}.
\]

These time steps can be partitioned into the following three intervals:

1) \( \{(k-1)c+1, \ldots, (k-1)c+c\} \), in which \( c \) blocks that encode message \( k \) and zero blocks that encode message \( k+1 \) are transmitted;
2) \( \{kc+1, \ldots, (k-1)c+m\} \), in which \( m-c \) blocks that encode message \( k \) and \( m-c \) blocks that encode message \( k+1 \) are transmitted; and
3) \( \{(k-1)c+m+1, \ldots, kc+m\} \), in which zero blocks that encode message \( k \) and \( c \) blocks that encode message \( k+1 \) are transmitted.

Suppose that \( V_b^{(1)} \) small blocks and \( V_b^{(1)} \) big blocks that encode message \( k \) are received by the decoder in the first interval, \( V_b^{(2)} \) small blocks and \( V_b^{(2)} \) big blocks that encode message \( k \) are received by the decoder in the second interval (the same numbers of blocks that encode message \( k+1 \) are also received in the same interval), and \( V_b^{(3)} \) small blocks and \( V_b^{(3)} \) big blocks that encode message \( k+1 \) are received by the decoder in the third interval; \( V_b^{(1)} \), \( V_b^{(1)} \), \( V_b^{(2)} \), \( V_b^{(2)} \), \( V_b^{(3)} \), and \( V_b^{(3)} \) are independent binomial random variables with the following distributions:

\[
V_b^{(1)} \sim \text{Binomial}(r_{m,c}, 1 - p_s),
V_b^{(2)} \sim \text{Binomial}(e - r_{m,c}, 1 - p_s),
\]
Fig. 6. Plots of the decoding failure probability $1 - P(S)$ and the burstiness of undecodable messages as measured by the conditional probability $P(S_{k+1} \mid S_k)$, where $k \in \mathbb{Z}^+$, against message size $s$ and packet erasure probability $p_e$ for the family of symmetric time-invariant intrasession codes, for $(c, d, m) = (3, 18, 6)$. In (a) and (c), we set $p_e = 0.05$. In (b) and (d), we set $s = 1$. Spreading parameter $m \in \{c, \ldots, d'\}$, where $d' = \min(d, m c) = 18$, gives the size of the effective coding window for each code. The black curve in (a) and (b) describes a lower bound on the decoding failure probability for any time-invariant code, as given by Theorem 4.

The conditional probability $P(S_{k+1} \mid S_k)$ can therefore be expressed in terms of these random variables as follows:

$$
V_{m,c}^{(2)} \sim \text{Binomial}(q_{m,c}r_{m,c}, 1 - p_k),
$$

$$
V_{m,c}^{(2)} \sim \text{Binomial}((q_{m,c} - 1)(c - r_{m,c}), 1 - p_k),
$$

$$
V_{m,c}^{(3)} \sim \text{Binomial}(r_{m,c}, 1 - p_k),
$$

$$
V_{m,c}^{(3)} \sim \text{Binomial}(c - r_{m,c}, 1 - p_k).
$$

The conditional probability $P(S_{k+1} \mid S_k)$ can therefore be expressed in terms of these random variables as follows:

$$
P(S_{k+1} \mid S_k) = 
\begin{cases}
0 & \text{if } r_{m,c} = c,
\frac{1}{q_{m,c}+1}\left(\frac{V_{m,c}^{(2)}+V_{m,c}^{(3)}}{V_{m,c}^{(1)}}\right) < s
\end{cases}
$$

Figs. 6c and 6d show how the family of symmetric codes perform in terms of the burstiness of undecodable messages as
measured by the conditional probability $\mathbb{P}\left[ S_{k+1} \mid S_k \right]$, for an instance of $(c, d, m, z)$. These plots and other empirical observations suggest that over a wide range of message sizes $s$ and packet erasure probabilities $p_e$, minimal spreading (i.e., $m = c$) performs well, i.e., achieves a relatively low $\mathbb{P}\left[ S_{k+1} \mid S_k \right]$, while maximal spreading (i.e., $m = d$) performs poorly. This agrees with the intuition that for a pair of consecutive messages, a greater overlap in their effective coding windows does not overlap at all.

3) Trade-off between Performance Metrics: Our results show that for the family of symmetric codes, a trade-off exists between the decoding probability $\mathbb{P}[S]$ and the burstiness of undecodable messages as measured by the conditional probability $\mathbb{P}\left[ S_{k+1} \mid S_k \right]$ when the message size $s$ and packet erasure probability $p_e$ are small (this is a regime of interest because it supports a high decoding probability). Specifically, although maximal spreading (i.e., $m = d$) achieves a high decoding probability, it also exhibits a higher burstiness of undecodable messages. Thus, a symmetric code with a suboptimal decoding probability but lower burstiness may be preferred for an application that is sensitive to bursty undecodable messages.

APPENDIX

PROOFS OF THEOREMS

A. Proof of Lemma 1

Suppose that the component systematic block code $C$ satisfies the symbol decoding requirements given by D1 and D2. We will show that the diagonally interleaved code derived from $C$ achieves a message size of $\sum_{i=1}^{\lfloor z \rfloor c} + 1$ and packet erasure probability $p_e$, which is the decoding deadline for message $k$, under any erasure pattern $E \in \mathcal{E}^n_k$. We do this by considering the codewords that contain one or more information symbols from $M_k$. There are a total of $d - z + c - 1$ such codewords, corresponding to the intervals

\[
L(x_{(k-1)c+1}(d-z)), \ldots, L(x_{(k-1)c+1}(1)), \ldots, L(x_{(k-1)c+c}(1)).
\]

We consider two cases separately, depending on whether the entire codeword interval occurs by the message decoding deadline:

Case 1: Consider the codeword corresponding to the interval $L(x_{(k-1)c+1}(i))$, where $i \in \{1, \ldots, d - z\}$. For brevity, we define

\[
L_i \triangleq L(x_{(k-1)c+1}(i)) = \{(k-1)c + 1 - i + 1, \ldots, (k-1)c + 1 - i + d\}.
\]

Observe that the entire interval $L_i$ occurs by the message decoding deadline since $(k-1)c + 1 - i + d \leq (k-1)c + d$. Let $\mathcal{E}^k_i$ be the set of erasure patterns from $\mathcal{E}^k$ that have been time-shifted to align with $L_i$, i.e.,

\[
\mathcal{E}^k_i \triangleq \{(k-1)c + 1 - i + t : t \in \mathcal{E}^k \} : E^k_1 \in \mathcal{E}^k_1.
\]

Under each erasure pattern $E \in \mathcal{E}^n_k$, the interval $L_i$ intersects with either

1) zero erasure bursts, in which case $L_i$ contains zero erased time steps; or
2) exactly one erasure burst, in which case $L_i$ contains at most $z$ erased time steps (i.e., the maximum length of an erasure burst), all in one contiguous subinterval; or
3) two or more erasure bursts, in which case $L_i$ contains a gap of at least $d - z$ unerased time steps between consecutive erasure bursts.

In each of these three cases, there exists some erasure pattern $E^* \in \mathcal{E}^k_i$ that is a superset of the erased time steps in $L_i$, i.e., $(E \cap L_i) \subseteq E^*$. Since the symbol decoding deadlines of D1 and D2 are satisfied under erasure pattern $E^*$, they must also be satisfied under erasure pattern $E \cap L_i$. It follows that all information symbols in the codeword are decodable by the last time step in interval $L_i$, and therefore by the message decoding deadline. Note that nonpositive time steps are always unerased under the erasure patterns in $\mathcal{E}^n_k$; their corresponding codeword symbols are therefore always known (recall that information symbols corresponding to nonpositive messages are assumed to be zeros).
Case 2: Consider the codeword corresponding to the interval $L(x_{(k-1)c+i+1}[1])$, where $i \in \{2, \ldots, c\}$. For brevity, we define

$$L_i \triangleq L(x_{(k-1)c+i+1}[1]) = \{(k-1)c+i-1+1, \ldots, (k-1)c+i-1+d\}.$$  

Observe that one or more time steps at the end of the interval $L_i$ occur after the message decoding deadline since $(k-1)c+i-1+d > (k-1)c+d$. The first minimum $(c+1-i, d-z)$ information symbols in the codeword correspond to message $k$; subsequent information symbols in the codeword (if any) correspond to later messages. Let $E_i^c$ be the set of erasure patterns from $E^c$ that have been time-shifted to align with $L_i$, i.e.,

$$E_i^c \triangleq \{(k-1)c+i-1+t : t \in E^c \} : E^c \in E_i^c\}.$$

As in Case 1, under each erasure pattern $E \in \mathcal{E}_{n}^d$, there exists some erasure pattern $E^c \in E_i^c$ that is a superset of the erased time steps in $L_i$, i.e., $(E \cap L_i) \subseteq E_i^c$. Since the symbol decoding deadlines of $D_1$ and $D_2$ are satisfied under erasure pattern $E^c$, they must also be satisfied under erasure pattern $E \cap L_i$. In particular, since

$$\min((c+1-i, d-z)) \leq c+1-i < c,$$

it follows from $D_1$ that the first minimum $(c+1-i, d-z)$ information symbols in the codeword, which correspond to message $k$, are decodable by the $(d-c \min((c+1-i, d-z))$th time step in interval $L_i$, which is time step

$$(k-1)c+i-1+d-c \min((c+1-i, d-z))$$

and therefore by the message decoding deadline. Note that although time steps after $(n-1)c + d$ (which is the final time step in $T_n$) are always unerased under the erasure patterns in $\mathcal{E}_{n}^d$, their corresponding codeword symbols are never used for decoding because all the information symbols in $\mathcal{M}$ have to be decoded by the final message decoding deadline, which is time step $(n-1)c + d$.

B. Proof of Theorem 1

We will apply Lemma 1 to show that the diagonally interleaved code derived from the stated systematic block code $C$ achieves a message size of $\frac{d-c}{d}c$ for the specified bursty erasure model. To demonstrate the asymptotic optimality of the code, we will show that this message size matches the maximum achievable message size $s_n^b$ in the limit, i.e.,

$$\lim_{n \to \infty} s_n^b = \frac{d-z}{d}c.$$  

To facilitate our description of the decoding procedure for $C$, we arrange the $d$ symbols of the codeword vector produced by $C$ sequentially across $d-z$ columns, with all the information symbols on the first row, as shown in Fig. 3. Note that each column $i \in \{1, \ldots, d-z\}$ of the table contains exactly $\frac{d-z}{d}c$ parity symbols. For each $i \in \{1, \ldots, d-z\}$, all the (degenerate) parity symbols below the information symbol $a[i]$ in column $i$ of the table have a value of $a[i]$.

Suppose that the $d$ symbols of the codeword vector are transmitted sequentially across an erasure link, one symbol per time step, over the time interval $L \triangleq \{1, \ldots, d\}$. Under each erasure pattern $E^c \in E^c$ (as defined in Lemma 1), exactly one symbol in each column of the table is unerased. Because the degenerate parity symbols take on the values of information symbols in a periodic manner, all the information symbols $a[1], \ldots, a[d-z]$ can be recovered using the $d-z$ unerased symbols. In particular, for each $i \in \{1, \ldots, d-z\}$, the information symbol $a[i]$ can be recovered by time step $d-(d-z)+i$. Since $d-z \geq c$, it follows that the symbol decoding requirements given by $D_1$ and $D_2$ in Lemma 1 are satisfied by $C$. Therefore, according to Lemma 1, the derived code achieves a message size of $\frac{d-z}{d}c$.

To obtain an upper bound for $s_n^b$, we consider the cut-set bound corresponding to a specific periodic erasure pattern $E' \subseteq T_n$ given by

$$E' \triangleq \{j \in T_n : j \in \mathbb{Z}_0^d, i \in \{1, \ldots, z\}\}.$$  

Since $E'$ comprises alternating intervals of $z$ erased time steps and $d-z$ unerased time steps, it is an admissible erasure pattern, i.e., $E' \in \mathcal{E}_{n}^d$. Now, consider a code that achieves the maximum message size $s_n^b$. Such a code must allow all $n$ messages $\{1, \ldots, n\}$ to be decoded under the specific erasure pattern $E'$. We therefore have the following cut-set bound for $s_n^b$:

$$n s_n^b \leq |T_n \setminus E'| \leq \frac{(n-1)c+d+1}{d} (d-z) = \frac{(n-1)c+2d}{d} (d-z) \left(c + \frac{2d-c}{n}\right).$$

Since a message size of $\frac{d-z}{d}c$ is known to be achievable (by the derived code), we have the following upper and lower bounds for $s_n^b$:

$$\frac{d-z}{d}c \leq s_n^b \leq \frac{d-z}{d} c + \frac{2d-c}{n}.$$

These turn out to be matching bounds in the limit as $n \to \infty$:

$$\frac{d-z}{d}c = \lim_{n \to \infty} s_n^b \leq \lim_{n \to \infty} \frac{d-z}{d} c + \frac{2d-c}{n} = \frac{d-z}{d}c.$$

We therefore have (3) as required.

C. Proof of Theorem 2

Our proof technique expands that of Theorem 1. First, we arrange the $d$ symbols of the codeword vector produced by the stated systematic block code $C$ sequentially across $z$ columns, with $r'$ information symbols on the second last row, and all the parity symbols on a separate last row, as shown in Fig. 4. For the case of $r' < d$, we repeat the $r'$ information symbols on the second last row, i.e., $a[d-z-r'+1], \ldots, a[d-z]$, across the row; these repeated virtual information symbols...
are parenthesized to distinguish them from the original actual information symbols of the codeword vector. Note that each column $i \in \{1, \ldots, z\}$ of the table contains exactly $\frac{d-z-r}{z} + 1 \geq 2$ actual and virtual information symbols. For each $i \in \{1, \ldots, z\}$, the value of the parity symbol $b[i]$ is given by the bit-wise modulo-2 sum (i.e., exclusive-or) of the actual and virtual information symbols above it in column $i$ of the table.

Suppose that the $d$ symbols of the codeword vector are transmitted sequentially across an erasure link, one symbol per time step, over the time interval $L = \{1, \ldots, d\}$. To show that the symbol decoding requirements given by D1 and D2 in Lemma 1 are satisfied by $C$, we consider the following four exhaustive cases separately:

Case 1: Consider the case of $r' = z$, for which there are no virtual information symbols. Under each erasure pattern $E^2 \in E^2$ (as defined in Lemma 1), exactly one symbol in each column of the table is erased. For each $i \in \{1, \ldots, z\}$, if the parity symbol $b[i]$ is erased, then all the information symbols in column $i$, which include $a[i]$, are unerased. On the other hand, if $b[i]$ is unerased, then 1) exactly one information symbol in column $i$ is erased; and 2) this information symbol can be recovered by time step $d - z + i$ using the unerased parity symbol $b[i]$ and the unerased and recovered information symbols in the column.

Case 2.1: Consider the case of $r' < z$, with the erasure pattern $E^2_j$, where

$$j \in \{1, \ldots, d - 2z\} \cup \{d - z + 1, \ldots, d\}.$$ 

Recall that index $j$ gives the time step of the “leading” erasure in the burst, which is of length $z$. Under erasure pattern $E^2_j$, the information symbol $a[d-z]$ and the parity symbol $b[1]$ are not simultaneously erased.

For each $i \in \{1, \ldots, r'\}$, if the parity symbol $b[i]$ is erased, then all the information symbols in column $i$ of the table, which include $a[i]$, are unerased. On the other hand, if $b[i]$ is unerased, then 1) exactly one information symbol in the column is erased; and 2) this information symbol can be recovered by time step $d - z + i$ using the unerased parity symbol $b[i]$ and the unerased and recovered information symbols in the column. It follows that all the information symbols on the second last row, i.e., $a[d-z-r'+1], \ldots, a[d-z]$, can be recovered by time step $d - z + r'$.

For each $i \in \{r' + 1, \ldots, z\}$, if the parity symbol $b[i]$ is erased, then all the actual information symbols in column $i$ of the table, which include $a[i]$, are unerased. On the other hand, if $b[i]$ is unerased, then 1) exactly one actual information symbol in the column is erased; and 2) this information symbol can be recovered by time step $d - z + i$ using the unerased parity symbol $b[i]$, the unerased actual information symbols in the column, and the recovered virtual information symbol $a[d-z-r'+r_i,r_i']$.

Case 2.2: Consider the case of $r' < z$, with the erasure pattern $E^2_j$, where

$$j \in \{d - 2z + 1, \ldots, d - z - r'\}. $$

Under erasure pattern $E^2_j$, the information symbols $a[1], \ldots, a[d-r']$, which include $a[1], \ldots, a[r']$, are unerased; 2) the information symbols on the second last row, i.e., $a[d-z-r'+1], \ldots, a[d-z]$, are erased; and 3) the parity symbols $b[d-z-r'+1], \ldots, b[z]$ are unerased.

For each $i \in \{1, \ldots, z - r'\}$, if the information symbol $a[d-2z+i]$ is erased, then 1) the parity symbols $b[i], \ldots, b[z]$ are unerased; 2) the information symbol $a[d-z-r'+u]$ can therefore be recovered by time step $d - z + r' + i$ using the unerased parity symbol $b[z-r'+i]$ and the unerased and recovered information symbols in the column; and 3) the information symbol $a[d-2z+i]$ can subsequently be recovered by time step $d - z + r' + i$ using the unerased parity symbol $b[z-r'+i]$, the unerased actual information symbols in the column, and the recovered virtual information symbol $a[d-z-r'+r_i,r_i'].$ It follows that for each $i \in \{1, \ldots, z\}$, the information symbols $a[1], \ldots, a[d-2z-r'+i]$, which include $a[i]$, can be recovered by time step $d - z + i$.

For each $i \in \{1, \ldots, r'\}$, the information symbol $a[d-z-r'+i]$ can be recovered by time step $d - r' + i$ using the unerased parity symbol $b[z-r'+i]$ and the unerased and recovered information symbols in the column.

Case 2.3: Consider the case of $r' < z$, with the erasure pattern $E^2_j$, where

$$j \in \{d - z - r' + 1, \ldots, d - z\}. $$

Under erasure pattern $E^2_j$, the information symbols $a[1], \ldots, a[d-z-r']$, which include $a[1], \ldots, a[z]$, are unerased.

For each $i \in \{1, \ldots, r'\}$, if the information symbol $a[d-z-r'+i]$ is erased, then 1) the parity symbols $b[z-r'+i], \ldots, b[z]$ are unerased; and 2) the information symbol $a[d-z-r'+i]$ can therefore be recovered by time step $d - r' + i$ using the unerased parity symbol $b[z-r'+i]$ and the unerased and recovered information symbols in the column.

Hence, under any erasure pattern $E^2 \in E^2$, all the information symbols $a[1], \ldots, a[d-z]$ are decodable by the last time step in interval $L$: in particular, the information symbol $a[i]$ is decodable by the $(d - z + i)$th time step in interval $L$, for each $i \in \{1, \ldots, z\}$. Since $z \geq c$, it follows that the symbol decoding requirements given by D1 and D2 in Lemma 1 are satisfied by $C$. Therefore, according to Lemma 1, the derived code achieves a message size of $\frac{d - z}{d}zc$.

The rest of the proof leading to the obtainment of (3) is the same as that of Theorem 1.

D. Proof of Theorem 3

Our proof technique expands that of Theorem 2. First, we arrange the $d$ symbols of the codeword vector produced by the stated systematic block code $C$ sequentially across $z'$ columns, with $r'$ information symbols on the $(\frac{d-z-r'}{z})$th row, and all the nondegenerate parity symbols $b[1], \ldots, b[z']$ on a separate row, followed by the degenerate parity symbols, as shown in Fig. 5. For the case of $r' < z'$, we repeat the $r'$ information symbols on the $(\frac{d-z-r'}{z})$th row, i.e., $a[d-z-r'+1], \ldots, a[d-z]$, across the row; these repeated
virtual information symbols are parenthesized to distinguish them from the original actual information symbols of the codeword vector. Note that each column $i \in \{1, \ldots, z\}$ of the table contains exactly $d - z - r'_i + 1 \geq 2$ actual and virtual information symbols. For each $i \in \{1, \ldots, z\}$, the value of the nondegenerate parity symbol $b[i]$ is given by the bit-wise modulo-2 sum (i.e., exclusive-or) of the actual and virtual information symbols above it in column $i$ of the table.

Suppose that the $d$ symbols of the codeword vector are transmitted sequentially across an erasure link, one symbol per time step, over the time interval $L \triangleq \{1, \ldots, d\}$. To show that the symbol decoding requirements given by (1) and (2) in Lemma 1 are satisfied by $C$, we consider the following six exhaustive cases separately:

Case 1: Consider the case of $r' = z'$, for which there are no virtual information symbols. Under each erasure pattern $E^2 \in E^2$ (as defined in Lemma 1), exactly $d - z$ symbols in each column of the table are unerased. Since the $d - z$ unerased symbols in the codeword vector are consecutive (possibly wrapping around symbols $b[z]$ and $a[1]$), it follows that the $d - z$ unerased symbols in each column of the table are on consecutive rows (possibly wrapping around the last and first rows).

For each $i \in \{1, \ldots, z\}$, let $S_i$ be the set of indices corresponding to the erased information symbols in column $i$ of the table. If the nondegenerate parity symbol $b[i]$ is erased, then because the degenerate parity symbols take on the values of information symbols in a periodic manner, each information symbol $a[k]$, where $k \in S_i$, can be recovered by time step $d - (d - z) + k$ using a matching unerased degenerate parity symbol with a value of $a[k]$. On the other hand, if $b[i]$ is unerased, then 1) let $\sigma_i \triangleq \max \{k : k \in S_i\}$; 2) each information symbol $a[k]$, where $k \in S_i \setminus \{\sigma_i\}$, can be recovered by time step $d - z + z' + k$ using the unerased degenerate parity symbol $b[z' + k]$, which has a value of $a[k]$; and 3) the remaining information symbol $a[\sigma_i]$ can be recovered by time step $d - z + \sigma_i$ using the unerased nondegenerate parity symbol $b[i]$ and the unerased and recovered information symbols in the column.

Case 2.1: Consider the case of $r' < z'$, with the erasure pattern $E^2_j$, where

$$j \in \{d - z + d - z' + 1, \ldots, d - z + d - z\}.$$  

Under erasure pattern $E^2_j$, all the nondegenerate parity symbols $b[1], \ldots, b[z']$ are erased. Each of the $d - z$ unerased symbols is therefore either an information symbol or a degenerate parity symbol (which is a copy of an information symbol). Because the degenerate parity symbols take on the values of information symbols in a periodic manner, all the information symbols $a[1], \ldots, a[d - z]$ can be recovered using the unerased symbols. In particular, for each $i \in \{1, \ldots, d - z\}$, the information symbol $a[i]$ can be recovered by time step $d - (d - z) + i$.

Case 2.2: Consider the case of $r' < z'$, with the erasure pattern $E^2_j$, where

$$j \in \{d - z + 1, \ldots, d - z + d - z - r'\}.$$  

Under erasure pattern $E^2_j$, all the information symbols on the $(d - z - r' + 1)$th row, i.e., $a[d - z - r' + 1], \ldots, a[d - z]$, are unerased.

For each $i \in \{1, \ldots, z\}$, let $S_i$ be the set of indices corresponding to the erased information symbols in column $i$ of the table. If $|S_i| = 0$, then all information symbols in the column are unerased. If $|S_i| \geq 1$, then 1) let $\sigma_i \triangleq \max \{k : k \in S_i\}$; 2) each information symbol $a[k]$, where $k \in S_i \setminus \{\sigma_i\}$, can be recovered by time step $d - z + \sigma_i$ using the unerased nondegenerate parity symbol $b[i]$ and the unerased and recovered information symbols in the column.

Case 2.3: Consider the case of $r' < z'$, with the erasure pattern $E^2_j$, where

$$j \in \{d - z + d - z' + 1, \ldots, d - z + d - z\}.$$  

Under erasure pattern $E^2_j$, 1) the information symbols $a[1], \ldots, a[d - z - r']$ are erased; 2) all the nondegenerate parity symbols $b[1], \ldots, b[z']$ are unerased; and 3) the degenerate parity symbols $b[z' + 1], \ldots, b[d - z - r']$ are unerased.

For each $i \in \{1, \ldots, d - z - r'\}$, the information symbol $a[i]$ can be recovered by time step $d - z + z' + i$ using the unerased degenerate parity symbol $b[z' + i]$, which has a value of $a[i]$.

For each $i \in \{1, \ldots, d - z - r'\}$, if the degenerate parity symbol $b[d - z - r' + i]$ has a value of $a[d - z - r' + i]$, is unerased, then 1) the information symbol $a[d - z - r' + i]$ can be recovered by time step $d - z + d - z - r' + i$ using it; and 2) the information symbol $a[d - z - r' + i]$ can subsequently be recovered by time step $d - z + d - z - r' + i$ using the unerased nondegenerate parity symbol $b[i]$ and the recovered information symbols in the column. On the other hand, if $b[d - z - r' + i]$ is erased, then 1) the information symbols $a[d - z - r' + i], \ldots, a[d - z]$ are unerased; and 2) the information symbol $a[d - z - r' + i]$ can therefore be recovered by time step $d - z + d - z - r' + i$ using the unerased nondegenerate parity symbol $b[i]$ and the unerased and recovered information symbols in the column. It follows that all the information symbols on the $(d - z - r' + 1)$th row, i.e., $a[d - z - r' + 1], \ldots, a[d - z]$, can be recovered by time step $d - z + d - z$.

For each $i \in \{1, \ldots, z'\}$, the information symbol $a[d - z - r' + i]$ can be recovered by time step $d - z + d - z$ using the unerased nondegenerate parity symbol $b[r' + i]$ and the recovered information symbols in the column.

Case 2.4: Consider the case of $r' < z'$, with the erasure pattern $E^2_j$, where

$$j \in \{d - z + d - z + 1, \ldots, d - z + d - z + z' - r'\}.$$  

Under erasure pattern $E^2_j$, 1) all the information symbols $a[1], \ldots, a[d - z] are erased; 2) the nondegenerate parity symbols $b[z' - r' + 1], \ldots, b[z'] are unerased; and 3) the degenerate parity symbols $b[z' + 1], \ldots, b[d - z]$ are unerased.
For each \( i \in \{1, \ldots, d - z - z'\} \), the information symbol \( a[i] \) can be recovered by time step \( d - z + z' + i \) using the uneraser degenerate parity symbol \( b[z' + i] \), which has a value of \( a[i] \).

For each \( i \in \{1, \ldots, z' - r'\} \), if the degenerate parity symbol \( b[d - z + i] \), which has a value of \( a[d - z - z' + i] \), is uneraser, then the information symbol \( a[d - z - z' + i] \) can be recovered by time step \( d - z + d - z - r' + i \) using it. On the other hand, if \( b[d - z + i] \) is erased, then 1) the nondegenerate parity symbols \( b[i], \ldots, b[z'] \) are uneraser; 2) the information symbol \( a[d - z - r' + r[i]] \) can therefore be recovered by time step \( d - z + d - z - r' + i \) using the uneraser nondegenerate parity symbol \( b[i] \) and the recovered information symbols in the column; and 3) the information symbol \( a[d - z - z' + i] \) can subsequently be recovered by time step \( d - z + d - z - r' + i \) using the uneraser parity symbol \( b[r' + i] \) and the recovered information symbols in the column.

For each \( i \in \{1, \ldots, r'\} \), the information symbol \( a[d - z - r' + i] \) can be recovered by time step \( d - z + z' + d - z - 2r' + i \) using the uneraser nondegenerate parity symbol \( b[z' + r' + i] \) and the recovered information symbols in the column.

**Case 2.5:** Consider the case of \( r' < z' \), with the erasure pattern \( E'_2 \), where

\[
j \in \{d - z + z' + d - z - r' + 1, \ldots, d - z + z' + d - z\}.
\]

Under erasure pattern \( E'_2 \), 1) all the information symbols \( a[1], \ldots, a[d - z] \) are erased; and 2) the parity symbols \( b[z' + 1], \ldots, b[z' + d - z - r'] \) are uneraser.

For each \( i \in \{1, \ldots, d - z - r'\} \), the information symbol \( a[i] \) can be recovered by time step \( d - z + z' + i \) using the uneraser degenerate parity symbol \( b[z' + i] \), which has a value of \( a[i] \).

For each \( i \in \{1, \ldots, r'\} \), if the degenerate parity symbol \( b[z' + d - z - r' + i] \), which has a value of \( a[d - z - r' + i] \), is uneraser, then the information symbol \( a[d - z - r' + i] \) can be recovered by time step \( d - z + z' + d - z - r' + i \) using it. On the other hand, if \( b[z' + d - z - r' + i] \) is erased, then 1) the nondegenerate parity symbols \( b[i], \ldots, b[z'] \) are uneraser; and 2) the information symbol \( a[d - z - r' + i] \) can therefore be recovered by time step \( d - z + z' + d - z - 2r' + i \) using the uneraser nondegenerate parity symbol \( b[z' + r' + i] \) and the recovered information symbols in the column.

Hence, under any erasure pattern \( E'_2 \in \mathcal{E}' \), the information symbol \( a[i] \) is decodable by the \((d - (d - z))d + (d - z)z\)th step in interval \( L \), for each \( i \in \{1, \ldots, d - z\} \). Since \( d - z \geq c \), it follows that the symbol decoding requirements given by D1 and D2 in Lemma 1 are satisfied by \( C \). Therefore, according to Lemma 1, the derived code achieves a message size of \( \frac{d - z}{d}c \).

The rest of the proof leading to the obtainment of (3) is the same as that of Theorem 1.

**E. Proof of Theorem 4**

By partitioning the set of uneraser time steps \( U_k \subseteq T_k \) into two sets \( U_k^{(d)} \subseteq T_k \setminus W_k \) (i.e., uneraser time steps before the coding window \( W_k \)) and \( U_k^{(d)} \subseteq W_k \) (i.e., uneraser time steps in the coding window \( W_k \)), we can rewrite (1) as follows:

\[
\mathbb{P} [S_k] = \sum_{U_k^{(d)} \subseteq T_k \setminus W_k} \sum_{z = 0}^{d} \sum_{U_k^{(d)} \subseteq W_k: \ |U_k^{(d)}| = d - z} \mathbb{I} \left( H(M_k | X[U_k^{(d)}], X[U_k^{(d)}]) = 0 \right) \cdot (1 - p_k)^{d - z(p_k)^z}.
\]

(4)

Observe that the conditional entropy term appearing in (4) can be lower-bounded as follows:

\[
H \left( M_k \mid X[U_k^{(d)}], X[U_k^{(d)}] \right) \geq H \left( M_k \mid M_k^{k - 1}, X(T_k \setminus W_k), X[U_k^{(d)}] \right) \geq H \left( M_k \mid M_k^{k - 1}, X[U_k^{(d)}] \right) \geq H \left( M_k \mid M_k^{k - 1 - m_e + 1}, X[U_k^{(d)}] \right),
\]

where

(a) follows from the addition of conditioned random variables \( M_k^{k - 1}, X(T_k \setminus W_k) \); (b) follows from the fact that packets \( X(T_k \setminus W_k) \) are independent and \( X[U_k^{(d)}] \) independent of messages \( M_k^{k - m_e} \); (c) follows from the fact that messages are independent, and packets \( X[U_k^{(d)}] \) are independent of messages \( M_k^{k - m_e} \). (we can show this explicitly by considering the conditional mutual information

\[
I \left( M_k: M_k^{k - m_e} \mid M_k^{k - m_e + 1}, X[U_k^{(d)}] \right) = H \left( M_k \mid M_k^{k - m_e + 1}, X[U_k^{(d)}] \right) - H \left( M_k \mid M_k^{k - 1}, X[U_k^{(d)}] \right) = H \left( M_k^{k - m_e} \mid M_k^{k - m_e + 1}, X[U_k^{(d)}] \right).
\]

where both conditional entropy terms on the third line are equal to \( H \left( M_k^{k - m_e} \right) \), which implies that both conditional entropy terms on the second line are equal). As a consequence of (5), we have

\[
\mathbb{I} \left( H \left( M_k \mid X[U_k^{(d)}], X[U_k^{(d)}] \right) = 0 \right) \leq \mathbb{I} \left( H \left( M_k \mid M_k^{k - 1 - m_e + 1}, X[U_k^{(d)}] \right) = 0 \right),
\]

and therefore (4) can be upper-bounded as follows:

\[
\sum_{U_k^{(d)} \subseteq T_k \setminus W_k} \sum_{z = 0}^{d} \sum_{U_k^{(d)} \subseteq W_k: \ |U_k^{(d)}| = d - z} \mathbb{I} \left( H \left( M_k \mid M_k^{k - 1 - m_e + 1}, X[U_k^{(d)}] \right) = 0 \right) \cdot (1 - p_k)^{d - z(p_k)^z}.
\]

(6)

\[
= \sum_{z = 0}^{d} \alpha_k(z) \cdot (1 - p_k)^{d - z(p_k)^z},
\]
where
\[ \alpha_k(z) \triangleq \sum_{U_{\ell_0}^{d-z} \subseteq W_k: |U_{\ell_0}^{d-z}| = d - z} \mathbb{1} \left[ H(M_k \mid M_{k-m_z+1}^{k-1}, X[U_k^{d-z}]) = 0 \right], \]
and (a) follows from a reordering of the sums, and the removal of the factor
\[ \sum_{U_{\ell_0}^{d-z} \subseteq T_k \backslash W_k} (1 - p_k)^{|U_{\ell_0}^{d-z}|} (p_k)^{|T_k| - d - |U_{\ell_0}^{d-z}|} = 1. \]
Consider a fixed choice of subset \( U \subseteq \{1, \ldots, d\} \). Suppose that \( U_{\ell_0}^{d-z} \subseteq W_k \) is the appropriately time-shifted version of \( U \), i.e.,
\[ U_{\ell_0}^{d-z} = \{(k-1)c + i : i \in U\}. \]
According to the definition of time-invariant codes, the packets \( X[U_k^{d-z}] \) can consequently be written in terms of \( U \) as
\[ X[U_k^{d-z}] = \left( f_{z,s} \left( M_k+q_z, \ldots, M_k+q_z-m_z+1 \right) \right)_{i \in U}. \]
The conditional entropy term in the definition of \( \alpha_k(z) \) can therefore be written in terms of the message random variables as
\[ H(M_k \mid M_{k-m_z+1}^{k-1}, X[U_k^{d-z}]) = H \left( M_k \mid M_{k-m_z+1}^{k-1}, \left( f_{z,s} \left( M_k+q_z, \ldots, M_k+q_z-m_z+1 \right) \right)_{i \in U} \right). \]
Since the joint probability distribution of the random variables in this expression is the same for any \( k \geq m_z \), it follows that this conditional entropy term is constant wrt \( k \geq m_z \). Defining \( \alpha(z) \triangleq \alpha_{m_z}(z) \), we therefore have
\[ \alpha(z) = \alpha_{m_z}(z) = \alpha_{m_z+1}(z) = \alpha_{m_z+2}(z) = \cdots \] (7)
for any \( z \in \{0, \ldots, d\} \). To obtain the required upper bound (2), we will show that for any \( z \in \{0, \ldots, d\} \),
\[ \alpha(z) \leq \left[ \min \left( \frac{(d-z)c}{d} + 1, \frac{(d-1)c}{d} \right) \right]. \] (8)
Suppose that \( z \in \{0, \ldots, d\} \). Consider the first \( m_z + n - 1 \) messages \( \{1, \ldots, m_z + n - 1\} \), and the union of their (overlapping) coding windows \( T_{m_z+n-1} \), where \( n \in \mathbb{Z}^+ \). Let \( \mathcal{E}_z \) be the collection of all \( \binom{d}{z} \) possible subsets \( \hat{E} \subseteq \{1, \ldots, d\} \) of size \( z \), i.e.,
\[ \mathcal{E}_z \triangleq \{ \hat{E} \subseteq \{1, \ldots, d\} : |\hat{E}| = z \}. \]
From each \( \hat{E} \in \mathcal{E}_z \), we derive a periodic erasure pattern \( E \subseteq T_{m_z+n-1} \) by concatenating copies of \( \hat{E} \); let \( \mathcal{E}_z \) be the set of these \( \binom{d}{z} \) erasure patterns, i.e.,
\[ \mathcal{E}_z \triangleq \left\{ \{ (j-1)d+i \in T_{m_z+n-1} : j \in \mathbb{Z}^+, i \in \hat{E} \} : \hat{E} \in \mathcal{E}_z \right\}. \]
Note that because of the periodicity of each erasure pattern \( E \in \mathcal{E}_z \), there are exactly \( z \) erased time steps and therefore exactly \( d - z \) unerased time steps in each coding window \( W_k \), i.e.,
\[ |W_k \setminus E| = d - z \quad \forall \ k \in \{1, \ldots, m_z + n - 1\}, E \in \mathcal{E}_z. \] (9)
Furthermore, for a fixed choice of \( k \in \{1, \ldots, m_z + n - 1\} \), the set of \( d - z \) unerased time steps in the coding window for message \( k \), i.e., \( W_k \setminus E \), is distinct under each erasure pattern \( E \in \mathcal{E}_z \); in other words,
\[ (E_1, E_2 \in \mathcal{E}_z) \land (W_k \setminus E_1 = W_k \setminus E_2) \implies E_1 = E_2 \quad \forall k \in \{1, \ldots, m_z + n - 1\}. \] (10)
Suppose that \( k \in \{1, \ldots, m_z + n - 1\} \) and \( E \in \mathcal{E}_z \). From the definition of conditional mutual information, we have
\[ I(M_k : X[W_k \setminus E] \mid M_{k-1}^{k-1}) = H(M_k \mid M_{k-1}^{k-1}, X[W_k \setminus E]) - H(M_k \mid M_{k-1}^{k-1}) = H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) \] (11)
Rearranging terms produces
\[ H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) = H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) + H(M_k \mid M_{k-1}^{k-1}, X[W_k \setminus E]). \] (11)
Since messages are independent, we have
\[ H(M_k \mid M_{k-1}^{k-1}) = s. \] (12)
Now, if
\[ H(M_k \mid M_{k-1}^{k-1}, X[W_k \setminus E]) = 0, \] (13)
then, by substituting (12) and (13) into (11), we obtain
\[ H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) = H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) - s. \] (14)
On the other hand, if condition (13) is not satisfied, then we have the inequality
\[ H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) \leq H(X[W_k \setminus E] \mid M_{k-1}^{k-1}), \] (15)
which is always true.
Suppose that \( k \in \{1, \ldots, m_z + n - 1\} \). According to the definition of \( \alpha_k(z) \), there are \( \alpha_k(z) \) subsets \( U_{\ell_0}^{d-z} \subseteq W_k \) of size \( d - z \) for which
\[ H(M_k \mid M_{k-m_z+1}^{k-1}, X[U_{\ell_0}^{d-z}]) = 0. \]
Equivalently, it follows from properties (9) and (10) of the set of erasure patterns \( \mathcal{E}_z \) that there are \( \alpha_k(z) \) erasure patterns \( E \in \mathcal{E}_z \) for which
\[ H(M_k \mid M_{k-m_z+1}^{k-1}, X[W_k \setminus E]) = 0. \]
Now, since
\[ H(M_k \mid M_{k-1}^{k-1}, X[W_k \setminus E]) \leq H(M_k \mid M_{k-m_z+1}^{k-1}, X[W_k \setminus E]), \] there are therefore at least \( \alpha_k(z) \) erasure patterns \( E \in \mathcal{E}_z \) for which condition (13) is satisfied. Summing over all erasure patterns and applying (14) and (15) the appropriate number of times produces the following inequality:
\[ \sum_{E \in \mathcal{E}_z} H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) \leq \left( \sum_{E \in \mathcal{E}_z} H(X[W_k \setminus E] \mid M_{k-1}^{k-1}) \right) - s \cdot \alpha_k(z). \] (16)
We now proceed to prove by induction that the following inequality holds for any \( k \in \{m_ε, \ldots, m_ε + n - 1\} \):
\[
\sum_{E \in \mathcal{E}_n} H \left( X[W_k \setminus E] \mid M_1^k \right) \\
\leq \left( \sum_{E \in \mathcal{E}_n} |T_k \setminus E| \right) - (k - m_ε + 1)s \cdot α(z). \tag{17}
\]

(Base case) Consider the case of \( k = m_ε \). According to (16), we have
\[
\sum_{E \in \mathcal{E}_n} H \left( X[W_{m_ε} \setminus E] \mid M_1^{m_ε} \right) \\
\leq \left( \sum_{E \in \mathcal{E}_n} H \left( X[W_{m_ε} \setminus E] \mid M_1^{m_ε-1} \right) \right) - s \cdot α_{m_ε}(z) \\
\leq \left( \sum_{E \in \mathcal{E}_n} H \left( X[W_{m_ε} \setminus E] \right) \right) - s \cdot α(z) \\
\leq \left( \sum_{E \in \mathcal{E}_n} |W_{m_ε} \setminus E| \right) - s \cdot α(z) \\
\leq \left( \sum_{E \in \mathcal{E}_n} |T_{m_ε} \setminus E| \right) - s \cdot α(z),
\]
as required, where
(a) follows from the removal of conditioned random variables \( M_1^{m_ε-1} \) in the entropy term, and the application of (7);
(b) follows from the fact that \( H(X_t) \leq 1 \) for any \( t \) because of the unit packet size;
(c) follows from the fact that \( W_{m_ε} \subseteq T_{m_ε} \).

(Inductive step) Suppose that (17) holds for some \( k \in \{m_ε, \ldots, m_ε + n - 2\} \). According to (16), we have
\[
\sum_{E \in \mathcal{E}_n} H \left( X[W_{k+1} \setminus E] \mid M_1^{k+1} \right) \\
\leq \left( \sum_{E \in \mathcal{E}_n} H \left( X[W_{k+1} \setminus E] \mid M_1^k \right) \right) - s \cdot α_{k+1}(z) \\
\leq \left( \sum_{E \in \mathcal{E}_n} H \left( X[W_{k+1} \cup (W_{k+1} \setminus E)] \mid M_1^k \right) \right) - s \cdot α(z) \\
\leq \left( \sum_{E \in \mathcal{E}_n} H \left( X[W_k \setminus E] \mid M_1^k \right) \right) \\
\quad + \left( \sum_{E \in \mathcal{E}_n} H \left( X[(W_{k+1} \setminus E) \setminus (W_k \setminus E)] \right) \right) - s \cdot α(z) \\
\leq \left( \sum_{E \in \mathcal{E}_n} |T_k \setminus E| \right) - (k - m_ε + 1)s \cdot α(z) \\
\quad + \left( \sum_{E \in \mathcal{E}_n} |(W_{k+1} \setminus E) \setminus (W_k \setminus E)| \right) - s \cdot α(z) \\
\quad \equiv \left( \sum_{E \in \mathcal{E}_n} |T_{k+1} \setminus E| \right) - (k - m_ε + 2)s \cdot α(z),
\]
as required, where
(a) follows from the addition of random variables \( X[W_k \setminus E] \) in the entropy term, and the application of (7);
(b) follows from the chain rule for joint entropy, and the removal of conditioned random variables \( X[W_k \setminus E] \), \( M_1^k \) in the second entropy term;
(c) follows from the inductive hypothesis, and the fact that \( H(X_t) \leq 1 \) for any \( t \) because of the unit packet size;
(d) follows from the fact that
\[
|T_k \setminus E| + |(W_{k+1} \setminus E) \setminus (W_k \setminus E)| = |T_{k+1} \setminus E| = |T_k \setminus E|.
\]

Now, since the conditional entropy term in (17) is nonnegative, it follows that for the choice of \( k = m_ε + n - 1 \), we have
\[
0 \leq \left( \sum_{E \in \mathcal{E}_n} |T_{m_ε+n-1} \setminus E| \right) \leq n \cdot s \cdot α(z),
\]
which implies
\[
α(z) \leq \frac{1}{n \cdot s} \sum_{E \in \mathcal{E}_n} |T_{m_ε+n-1} \setminus E| \\
\leq \frac{d}{(m_ε + n - 2)c + d} \left( \frac{d}{d} \right) \left( \frac{c + m_ε c - 2c + 2d}{n} \right).
\]

Furthermore, since \( α(z) \) is independent of \( n \), this upper bound must also hold in the limit \( n \to \infty \), i.e.,
\[
α(z) \leq \frac{(d - z)c}{d \cdot s(z)}.
\]

Finally, taking into account the fact that \( α(z) \) is an integer that is at most \( \binom{d}{c} \), we arrive at (8). Applying (7) and (8) to (6) produces the required upper bound (2) on \( P[S_k] \) for \( k \geq m_ε \).

\section*{Acknowledgment}

The authors would like to thank Ashish Khisti for the interesting discussions.

\section*{References}


