Detailed Asymptotics of the Delay-Reliability Tradeoff of Random Linear Streaming Codes

Pin-Wen Su*, Yu-Chih Huang†, Shih-Chun Lin‡, I-Hsiang Wang‡, and Chih-Chun Wang*
*School of ECE, Purdue University, USA, {su173, chihw}@purdue.edu
†Institute of CE, National Yang Ming Chiao Tung University, Taiwan, jerryhuang@nycu.edu.tw
‡Department of EE and GICE, National Taiwan University, Taiwan, {sclin2, ihwang}@ntu.edu.tw

***Abstract*—*Streaming codes* eliminate the queueing delay and are an appealing candidate for low latency communications. This work studies the tradeoff between error probability p_e and decoding deadline Δ of infinite-memory random linear streaming codes (RLSCs) over i.i.d. symbol erasure channels (SECs). The contributions include (i) Proving $p_e(\Delta) \sim \rho\Delta^{-1.5}e^{-\eta\Delta}$. The *asymptotic power* term $\Delta^{-1.5}$ of RLSCs is a strict improvement over the $\Delta^{-0.5}$ term of random linear block codes; (ii) Deriving a pair of upper and lower bounds on the *asymptotic constant* ρ, which are tight (i.e., identical) for one specific class of SECs; (iii) For any $c > 1$ and any decoding deadline Δ, the *c-optimal memory length* $\alpha_c^*(\Delta)$ is defined as the minimal memory length α needed for the resulting p_e to be within a factor of c of the best possible p_e^* under any α, an important piece of information for practical implementation. This work studies and derives new properties of $\alpha_c^*(\Delta)$ based on the newly developed asymptotics.**

I. Introduction

Low latency communication aims to minimize the end-to-end (e2e) delay, and is a key enabler for various applications ranging from autonomous vehicles to remote healthcare. One promising approach of e2e delay reduction is the *streaming codes* [1]–[6], which take a string of message symbols sequentially as input and output a string of coded symbols instantaneously with zero queueing delay (no waiting).

Our previous work [7], [8] considered random linear streaming codes (RLSCs) over i.i.d. symbol erasure channels (SECs) with a sufficiently large finite field size for each symbol. We studied the error probability p_e in the finite memory length α and finite decoding deadline Δ regime, and derived a closed-form expression of $p_e(\Delta, \alpha)$ for any given coding rate R.

Contrary to the finite length results [7], [8], two asymptotic results (with some parameters being asymptotically large) have been developed. Specifically, [9] assumed fixed rate R and *infinite* deadline $\Delta = \infty$, and characterized the *detailed asymptotics* of p_e versus α when $\alpha \to \infty$. Additionally, [10] assumed fixed p_e but *infinite* memory $\alpha = \infty$, and characterized the tradeoff between R and Δ, the so-called second-order achievability regime [11].

Related to our previous asymptotic results [9], [10], in this work we assume fixed R and infinite memory $\alpha = \infty$, and study the tradeoff between p_e and Δ when Δ is sufficiently large. That is, assuming the complexity (memory) is not a concern, we characterize the *detailed asymptotics* of $p_e(\Delta)$ when $\Delta \to \infty$, arguably one of the most important coding-theoretic subjects that depicts the limiting behavior of the codes, which was first studied by Fano in 1961 [12] and further improved in the following half of century [13]–[16]. Specifically, we focus on the tradeoff between p_e and delay x, where x is the blocklength n if under a block code setting or x is the decoding deadline Δ if under a streaming code setting. Our goal is to derive the detailed asymptotics of $p_e(x)$ (also see the discussion in [17]):

$$p_e(x) = \rho x^\beta e^{-\eta x} + o(x^\beta e^{-\eta x}) \tag{1}$$

of which ρ, β and η are called *asymptotic constant*, *asymptotic power*, and *asymptotic decay rate*, respectively.

In a block code setting [12]–[16], the aim is usually to characterize the η value, which is commonly referred to as the *error exponent*. We refer any analysis that explicitly finds the ρ and β values (not just the η value) as a *detailed asymptotic* result since ρ and β play a critical role for small-to-moderate delay x, thus more "detailed".

With our focus exclusively on the *detailed asymptotics* of RLSCs over i.i.d. SECs with sufficiently large finite field size, we make the following contributions:

(i) Under the setting $\alpha = \infty$, we prove that

$$p_e(\Delta) = \rho\Delta^{-1.5}e^{-\eta\Delta} + o(\Delta^{-1.5}e^{-\eta\Delta}) \tag{2}$$

where the η value is identical to the random coding error exponent of block codes [11], [18]. The implication is that RLSCs improve the asymptotic power $\Delta^{-1.5}$ when compared to the $n^{-0.5}$ term of random linear block codes with blocklength n [18], also see our discussion of Lemma 3 and Theorem 1.

(ii) We derive a pair of lower and upper bounds on the ρ value. Our bounds are tight (i.e., identical) for one class of SECs. For SECs outside the specified class, we derive a *good approximation* of ρ as well. (In this work, we say an approximation is tight if it provably matches the true curve when $x \to \infty$. We say an approximation is good if it is numerically close to the true curve when x is large.)

(iii) For any fixed R, any Δ and any $c > 1$, we define the *c-optimal memory length* $\alpha_c^*(\Delta)$ by

$$\alpha_c^*(\Delta) = \inf\Big\{\alpha : p_e(\Delta, \alpha) \leq c \cdot \inf_\alpha p_e(\Delta, \alpha)\Big\}. \tag{3}$$

E.g., say $c = 1.1$, $\alpha_{1.1}^*(\Delta)$ is the memory length needed to be within 10% of the best achievable $p_e^* = \inf_\alpha p_e(\Delta, \alpha)$, an

This work was supported in parts by NSF under Grants CCF-1816013, CCF-2008527, and CNS-2107363; National Spectrum Consortium (NSC) under grant W15QKN-15-9-1004; and by MOST Taiwan under Grants 111-2221-E-A49 -069 -MY3 and 111-2221-E-002-099-MY2.

important piece of information dictating how much memory is needed when designing a practical streaming code with performance close to p_e^*. Under some mild assumptions, this work uses the new asymptotics (2) and shows that

$$\alpha_c^*(\Delta) \approx a_1\Delta + a_2\ln(\Delta) + a_3\ln(1/(c-1)) + O(1) \quad (4)$$

where a_1 to a_3 are constants that do not depend on c and Δ. In addition to establishing that $\alpha_c^*(\Delta)$ grows approximately linearly with respect to Δ (the statement will be made more precise in Section IV), we have derived a *good approximation* of $\alpha_c^*(\Delta)$ for small Δ.

II. Problem Formulation and Existing Results

The boldface lower/upper letters denote column vectors/matrices, respectively, e.g., $\mathbf{s}(t)$ denotes a column vector indexed by t. We use $\mathbf{s}_a^b$ to represent the *cumulative* column vector $\mathbf{s}_a^b \triangleq \left[\mathbf{s}^\top(a), \mathbf{s}^\top(a+1), \ldots, \mathbf{s}^\top(b)\right]^\top$. We define $(\cdot)^+ \triangleq \max(0,\cdot)$ as the projection operator.

A. The Model of Random Linear Streaming Codes

Consider a point-to-point communication system. In every time slot $t \geq 1$, K message symbols $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_K(t)]^\top$ arrive at the encoder, where each $s_k(t)$ is drawn independently and uniformly randomly from $\mathsf{GF}(2^q)$. Using the αK symbols in the previous α slots $\{\mathbf{s}(\tau) : \tau \in [t-\alpha, t)\}$ plus the current $\mathbf{s}(t)$ as input, where α is called the *memory length*, the encoder outputs N coded symbols by a linear encoder with the encoding matrix uniformly randomly generated from $\mathsf{GF}(2^q)^{N\times(\alpha+1)K}$.

The N coded symbols then pass through a symbol erasure channel (SEC), during which a random subset will be erased and the rest of them, denoted by $\mathcal{C}_t \subseteq \{1, 2, \cdots, N\}$, will arrive at the destination perfectly. Define $C_t \in [0, N]$ as the (random) number of successfully received symbols, which is assumed to be i.i.d. over time. We further denote the C_t received symbols at time t by $\mathbf{y}(t) = [y_1(t), \ldots, y_{C_t}(t)]^\top$. We can thus write

$$\mathbf{y}(t) = \mathbf{H}_t \mathbf{s}_{\max(t-\alpha,1)}^t \quad (5)$$

where $\mathbf{H}_t$ is the (random) receiver matrix that corresponds to the realization of the SEC. By properly shifting and stacking $\mathbf{s}(t)$, $\mathbf{y}(t)$ and $\mathbf{H}_t$, the system till time t can be expressed as:

$$\mathbf{y}_1^t = \mathbf{H}^{(t)}\mathbf{s}_1^t. \quad (6)$$

See Fig. 1 for illustration. There is a diagonal band of non-zero entries in $\mathbf{H}^{(t)}$, the values of which correspond to the randomly drawn encoding matrix.

For any fixed $K < \mathbb{E}\{C_t\}$, we define the error probability $p_{e,q}(\Delta,\alpha)$ under decoding deadline Δ, memory length α and finite field size $\mathsf{GF}(2^q)$ by

$$p_{e,q}(\Delta,\alpha) \triangleq \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Pr\Big(\mathbf{s}(t) \neq f_{\mathrm{ML},t}\Big(\mathbf{y}_1^{t+\Delta}, \mathbf{H}^{(t+\Delta)}\Big)\Big)$$

where $f_{\mathrm{ML},t}(\mathbf{y}_1^{t+\Delta}, \mathbf{H}^{(t+\Delta)})$ is the ML decoder of $\mathbf{s}(t)$ based on the observation $\mathbf{y}_1^{t+\Delta}$ by the deadline $t+\Delta$. We assume the ML decoder also knows the receiver matrix $\mathbf{H}^{(t+\Delta)}$, the most common setting used in erasure coding literature [1], [3], [11], [19], [20].

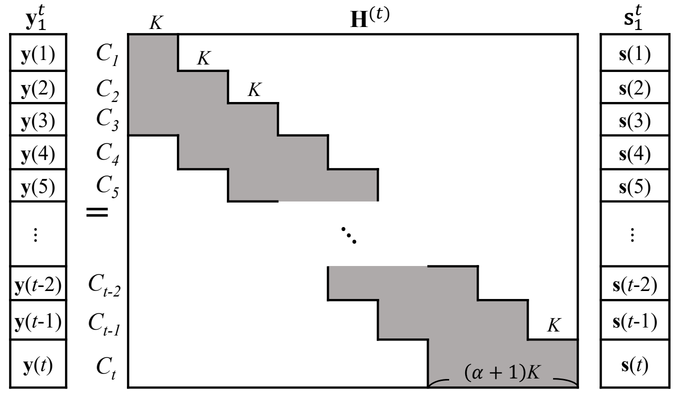

Fig. 1. The matrix-based illustration of the system model.

Focusing exclusively on a sufficiently large finite field $\mathsf{GF}(2^q)$, this work studies the following three quantities:

$$p_e(\Delta,\alpha) \triangleq \lim_{q\to\infty} p_{e,q}(\Delta,\alpha) \quad (7)$$

$$p_e(\Delta) \triangleq p_e(\Delta,\infty) \triangleq \lim_{\alpha\to\infty} p_e(\Delta,\alpha) \quad (8)$$

$$p_e(\infty,\alpha) \triangleq \lim_{\Delta\to\infty} p_e(\Delta,\alpha). \quad (9)$$

B. Existing Results on RLSCs

Define a *one-side bounded random walk* $I_d(t) \geq 0$ by

Definition 1: Set $I_d(0) \triangleq 0$ and iteratively compute

$$I_d(t) \triangleq (X_t + I_d(t-1))^+ \quad (10)$$

where $X_t \triangleq K - C_t$.

Intuitively, X_t is the *degree of uncertainty/freedom* added to the system at time t, a balance between K new message symbols and C_t received coded symbols. The $I_d(t)$ is called the *information debt at time* t of the infinite memory setting [7]–[10], [21], [22].

Definition 2: Since $I_d(0) = 0$, define the *recurrence time of state*-0 by

$$t_{R0} \triangleq \inf\{\tau : \tau > 0, I_d(\tau) = 0\}. \quad (11)$$

The following existing result converts the problem of finding $p_e(\Delta)$ to the analysis of the random walk $I_d(t)$, which will be the foundation of our asymptotic analysis.

Lemma 1: [8, Section IV].

$$p_e(\Delta) = \frac{\mathbb{E}\Big\{(t_{R0} - (\Delta+1))^+\Big\}}{\mathbb{E}\{t_{R0}\}}. \quad (12)$$

We conclude this section by introducing several definitions. Recall that C_t has support $[0, N]$. The support of $X_t = K - C_t$ is thus $[-N+K, K]$.

Definition 3: Define $x_{\mathrm{dw,max}} \triangleq N - K$ and $x_{\mathrm{up,max}} \triangleq K$. That is, $x_{\mathrm{dw,max}}$ (resp. $x_{\mathrm{up,max}}$) denotes the maximum downward (resp. upward) step size that the random walk $I_d(t)$ can make per time slot t.

Definition 4: Define $X_t^+ = \max(0, X_t)$ and define two Laurent series $\Phi(z)$ and $\Phi_+(z)$ as follows:

$$\Phi(z) \triangleq \mathbb{E}\left\{z^{X_t}\right\} \text{ and } \Phi_+(z) \triangleq \mathbb{E}\left\{z^{X_t^+}\right\}. \tag{13}$$

The exact expressions of $\Phi(z)$ and $\Phi_+(z)$ are determined by the probability mass function (PMF) of $X_t = K - C_t$. We use $\Phi'(z)$, $\Phi''(z)$, $\Phi'_+(z)$ and $\Phi''_+(z)$ to denote their first and second order derivatives. Finally, we define

$$z_0 \triangleq \operatorname*{argmin}_{z>0} \Phi(z). \tag{14}$$

III. Main Result #1: Detailed Asymptotics

For any positive integer $a > 0$, we define

$$T_{a\to 0} \triangleq \inf\{t \geq 1 : \sum_{i=1}^{t} X_i \leq -a\}. \tag{15}$$

Intuitively, $T_{a\to 0}$ describes the *hitting time* for the state $I_d(\tau) = a$ to go back to $I_d(\tau') = 0$. We then have

Lemma 2: For any arbitrary positive integer $a > 0$, we have

$$\frac{a}{x_{\text{dw,max}}} \cdot \frac{1}{t} \cdot \Pr\left(\sum_{i=1}^{t} X_i = -a\right) \leq \Pr(T_{a\to 0} = t) \tag{16}$$

$$\leq a \cdot \sum_{\delta=0}^{x_{\text{dw,max}}-1} \frac{1}{t} \cdot \Pr\left(\sum_{i=1}^{t} X_i = -(a+\delta)\right). \tag{17}$$

Some remarks are in order. Firstly, if $x_{\text{dw,max}} = 1$, then the lower and upper bounding values are identical. Lemma 2 thus characterizes the exact value of $\Pr(T_{a\to 0} = t)$ regardless of the $x_{\text{up,max}}$ value. The reason that the bounds are loose in general is that if $x_{\text{dw,max}} \geq 2$, the random walk may sometimes *overshoot* state-0 before the value of $I_d(\tau)$ being shifted back to 0 by the $(\cdot)^+$ operator in (10). This phenomenon significantly complicates the analysis.

Secondly, if we further impose $x_{\text{up,max}} = x_{\text{dw,max}} = 1$, then Lemma 2 is a well-known result of the *reflection principle* of Bernoulli random walks [23]. Allowing for arbitrary $x_{\text{up,max}}$ and $x_{\text{dw,max}}$, Lemma 2 can be viewed as a generalization of the reflection principle, also see the related results on the *ballot theorem* and the *Catalan number* [24], [25].

Lemma 3: For any fixed $a > 0$ when $t \to \infty$,

$$\Pr\left(\sum_{i=1}^{t} X_i = -a\right) = z_0^{a-1}\sqrt{\frac{\Phi(z_0)}{2\pi\Phi''(z_0)}}\, t^{-0.5}(\Phi(z_0))^t + o(t^{-0.5}(\Phi(z_0))^t) \tag{18}$$

where z_0 is defined in (14).

Lemma 3 is proved by Laplace's method of asymptotic expansion, see similar results in [18]. It is worth noting that we have a $t^{-0.5}$ term in (18). Since the left-hand side of (18) is highly related to the error probability of a MDS block code with blocklength $n = t$, the p_e of a random linear block code over i.i.d. SECs has an asymptotic power term $n^{-0.5}$.

If we rewrite the numerator of (12) by

$$\mathbb{E}\left\{(t_{R0} - (\Delta+1))^+\right\} = \sum_{t=1}^{\infty}\sum_{a=1}^{K} t \cdot \Pr(I_d(1) = a) \cdot \Pr(T_{a\to 0} = t + \Delta), \tag{19}$$

the resulting expression shows that the value of $p_e(\Delta)$ using RLSCs in (12) is directly related to the hitting time event $\{T_{a\to 0} = t\}$, rather than the block-based event $\left\{\sum_{i=1}^{t} X_i = -a\right\}$. Combining Lemmas 1 to 3, we have

Proposition 1: Define

$$\varrho_{\text{ub}} \triangleq \frac{x_{\text{dw,max}} \cdot z_0^{x_{\text{dw,max}}-1}\Phi'_+(z_0)}{(1-\Phi(z_0))^2}\frac{(\Phi(z_0))^{1.5}}{\sqrt{2\pi\Phi''(z_0)}}, \tag{20}$$

$$\varrho_{\text{lb}} \triangleq \frac{1}{x_{\text{dw,max}}}\frac{\Phi'_+(z_0)}{(1-\Phi(z_0))^2}\frac{(\Phi(z_0))^{1.5}}{\sqrt{2\pi\Phi''(z_0)}} \tag{21}$$

$$\text{and} \quad \eta \triangleq -\ln(\Phi(z_0)). \tag{22}$$

For any $\epsilon > 0$, there exists $\Delta_\circ$ such that

$$(1-\epsilon)\varrho_{\text{lb}} \leq \frac{\mathbb{E}\left\{(t_{R0} - (\Delta+1))^+\right\}}{\Delta^{-1.5}e^{-\eta\Delta}} \leq (1+\epsilon)\varrho_{\text{ub}} \tag{23}$$

for all $\Delta \geq \Delta_\circ$.

Finally, we can quantify the denominator $\mathbb{E}\{t_{R0}\}$ of (12) by standard hitting analysis, and we have

Theorem 1: For RLSCs with infinite memory $\alpha = \infty$,

$$p_e(\Delta) = \rho\Delta^{-1.5}e^{-\eta\Delta} + o(\Delta^{-1.5}e^{-\eta\Delta}) \tag{24}$$

where η is first defined in (22) and ρ is of bounded value.

Remark 1: The η value in (22) is identical to the random coding error exponent of *block codes* [11], [18]. Therefore, RLSCs neither improve nor degrade the error exponent when compared to its block-code counterpart.

Remark 2: The improvement of the asymptotic power from $n^{-0.5}$ of block codes to $\Delta^{-1.5}$ of RLSCs in (24) is a direct result of the $\frac{1}{t}$ term in Lemma 2 that relates the two probabilities $\Pr(\sum_{i=1}^{t} X_i = -a)$ and $\Pr(T_{a\to 0} = t)$.

Remark 3: Theorem 1 shows that the p_e of RLSCs decreases faster than the p_e of random block codes by a linear order of the delay (Δ or n). Therefore, RLSCs not only eliminate the queueing delay completely but also strictly enhance the error protection when compared to random block codes in the asymptotic regime.

In the sequel, we study the asymptotic constant ρ in (24).

Theorem 2: If $N = K + 1$, then we have

$$\rho = \frac{(z_0)^{-2}\Phi(z_0)^{1.5} \cdot |\Phi'(1)|}{(1-\Phi(z_0))^2\sqrt{2\pi\Phi''(z_0)}}. \tag{25}$$

Theorem 2 is proved by noticing that Lemma 2 and Proposition 1 are tight when $x_{\text{dw,max}} = 1 = N - K$. If $N > K+1$, then we derive a pair of upper and lower bounds instead.

Theorem 3: Define

$$\rho_{\text{ub}} \triangleq \varrho_{\text{ub}}(\mathbb{E}\{t_{R0}\})^{-1} \text{ and } \rho_{\text{lb}} \triangleq \varrho_{\text{lb}}(\mathbb{E}\{t_{R0}\})^{-1} \tag{26}$$

where the value of $\mathbb{E}\{t_{R0}\}$ can be found by the results in [9]. For arbitrary N and K values, we always have

$$0 < \rho_{\text{lb}} \leq \rho \leq \rho_{\text{ub}}. \tag{27}$$

In addition to rigorous upper/lower bounds, we have derived a closed-form approximation of ρ for the case of $N \geq K+1$:

$$\widetilde{\rho} = \frac{\Phi'_{+}(z_0)\Phi(z_0)^{1.5} \cdot |\Phi'(1)|}{(1-\Phi(z_0))^2 \sqrt{2\pi\Phi''(z_0)}\big(\Phi'_{+}(1) - \Phi'(1)\big)} \tag{28}$$

which is found by extrapolating the derivation of ρ from the special case of $N = K+1$ to the general case of $N \geq K+1$. By the definitions in (25), (26), and (28), one can easily verify the following corollary:

Corollary 1: If $N = K+1$, then $\rho_{\text{lb}} = \rho_{\text{ub}} = \widetilde{\rho} = \rho$.

A. Numerical Evaluation

Fig. 2 plots the curves governing $p_e(\Delta)$. We first assume $(N, K) = (3, 2)$ and the PMF of channel, $P_i \triangleq \Pr(C_t = i)$, listed in the caption of Fig. 2. The Exact $p_e(\Delta)$ is plotted by the closed-form formulas in [8, Section IV]. We also plot three types of asymptotic curves, from the coarsest $e^{-\eta\Delta}$ to the most detailed $\rho\Delta^{-1.5}e^{-\eta\Delta}$ with the ρ value described in Theorem 2. As can be seen, the coarser asymptotics are far away from $p_e(\Delta)$ while the provably tight asymptotics closely matches the exact $p_e(\Delta)$ for small $\Delta \in [50, 200]$. Similar observations can be made in all our numerical experiments satisfying $N = K+1$.

In Fig 3, we let $(N, K) = (5, 2)$ and C_t be uniformly distributed over $[0, 5]$, i.e., $\Pr(C_t = i) = 1/6$. We again plot the Exact and three types of asymptotics, except that in Type 3, we use the approximation $\widetilde{\rho}$ in (28) since $N > K+1$ in this example. The results once again demonstrate strong prediction power of the $\widetilde{\rho}\Delta^{-1.5}e^{-\eta\Delta}$ even though there is no tightness guarantee. We observe similar behavior for all (N, K) values and channel distributions we have evaluated.

IV. Main Result #2: c-Optimal Memory Length

This section demonstrates how the strong prediction power of the *detailed asymptotics* can be used to study the c-optimal memory length defined in (3). To that end, we first assume $p_e(\Delta, \alpha)$ of finite Δ and α can be approximated as follows.

$$\begin{aligned} p_e(\Delta, \alpha) &\approx \widetilde{p}_e(\Delta, \alpha) \\ &\triangleq p_e(\Delta, \infty) + p_e(\infty, \alpha) = p_e(\Delta) + p_e(\infty, \alpha) \end{aligned} \tag{29}$$

Fig. 4 verifies the closeness of this approximation when $(N, K) = (5, 2)$ and C_t being binomial with parameters $(5, 0.45)$. We first plot $p_e(\infty, \alpha)$, which is the curve that approaches 0 when $\alpha \to \infty$ since longer memory will lower the error probability when delay $\Delta = \infty$ is not a concern. We then plot two sets of curves, one for $\Delta = 100$ and one for $\Delta = 200$. For $\Delta = 100$, we first plot $p_e(100, \infty)$, which is a flat line (constant) as it does not depend on α. We then plot the approximation $\widetilde{p}_e(100, \alpha)$ in (29), the sum of two curves $p_e(100, \infty)$ and $p_e(\infty, \alpha)$. Finally, we plot the true error probability $p_e(100, \alpha)$ using [8].

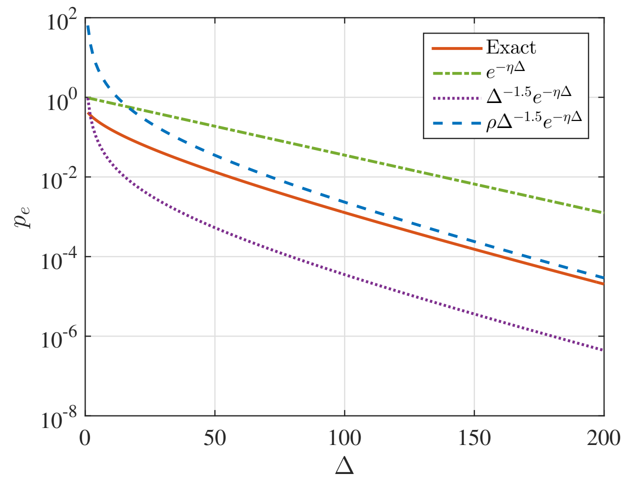

Fig. 2. $p_e(\Delta)$ w. $(N, K) = (3, 2)$ and $(P_0, \ldots, P_3) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3}, \frac{1}{2})$.

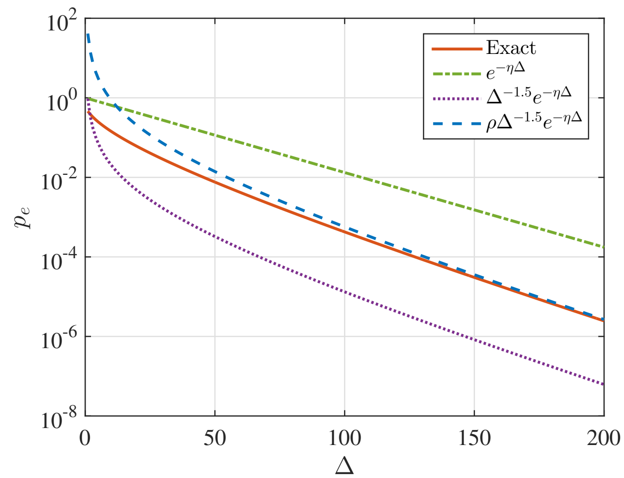

Fig. 3. $p_e(\Delta)$ w. $(N, K) = (5, 2)$ and $C_t \sim U(0, 5)$.

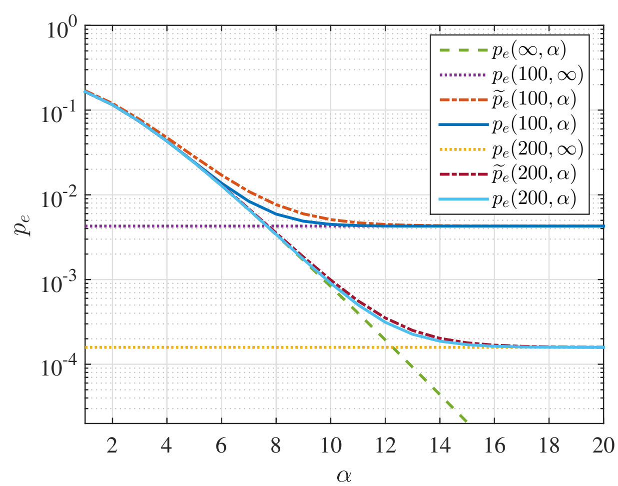

Fig. 4. The p_e vs α when $(N, K) = (5, 2)$ and $C_t \sim B(5, 0.45)$.

One can see that $\widetilde{p}_e(100,\alpha)$ closely approximates the true $p_e(100,\alpha)$, In addition to $\Delta = 100$, we also plot the set of curves for $\Delta = 200$. As can be seen, the gap between $\widetilde{p}_e(\Delta,\alpha)$ and $p_e(\Delta,\alpha)$ becomes even smaller when Δ increases from 100 to 200. The heuristic of our approximation formula (29) is as follows. For fixed Δ, $p_e(\Delta,\alpha)$ first decreases when α increases since the newly added complexity (memory) enhances error protection. However, after more and more memory being added, we face diminishing return and $p_e(\Delta,\alpha)$ is later dominated by the flat line $p_e(\Delta,\infty)$. This phenomenon is captured by (29).

If the approximation (29) is reasonably tight, we will have

$$\alpha_c^*(\Delta) \approx \min\{\alpha : \widetilde{p}_e(\Delta,\alpha) \leq c \cdot \inf_{\alpha} \widetilde{p}_e(\Delta,\alpha)\} \tag{30}$$

$$= \min\{\alpha : p_e(\infty,\alpha) \leq (c-1)p_e(\Delta)\} \tag{31}$$

where (31) holds because (i) by (29) and by the fact that $p_e(\infty,\alpha) \searrow 0$ when $\alpha \to \infty$, we have $\inf_\alpha \widetilde{p}_e(\Delta,\alpha) = p_e(\Delta)$; (ii) we replace $\widetilde{p}_e(\Delta,\alpha)$ by its definition (29) and rearrange the inequality.

Note that the min-based expression in (31) is still difficult to solve. We thus further substitute $p_e(\Delta)$ by our good/tight approximation $\widetilde{\rho}\Delta^{-1.5}e^{-\eta\Delta}$ and substitute $p_e(\infty,\alpha)$ by the following tight approximation first derived in [9]:

$$p_e(\infty,\alpha) = (B_1\alpha + B_2)e^{-B_3\alpha} + o(e^{-B_3\alpha}) \tag{32}$$

where (B_1, B_3) values are provided in [9] and B_2 can be derived by the methods of [9]. We then approximate (31) by

$$\alpha_c^*(\Delta) \approx \widetilde{\alpha}_c^*(\Delta) \triangleq$$
$$\min\big\{\alpha : (B_1\alpha + B_2)e^{-B_3\alpha} \leq (c-1)\widetilde{\rho}\Delta^{-1.5}e^{-\eta\Delta}\big\}. \tag{33}$$

With $\widetilde{\rho}$, η, B_1, B_2 and B_3 all admitting closed-form expressions, the value of our approximation $\widetilde{\alpha}_c^*(\Delta)$ can be efficiently computed by the Lambert W function [26], [27] supported in commercial softwares.

We now analyze the behavior of $\widetilde{\alpha}_c^*(\Delta)$.

Lemma 4: When $\Delta \to \infty$,

$$\widetilde{\alpha}_c^*(\Delta) = \lceil a_1\Delta + a_2\ln(\Delta) + a_3\ln(1/(c-1)) + a_4 + o(1)\rceil$$

where

$$a_1 = \frac{\eta}{B_3},\ a_2 = \frac{2.5}{B_3},\ a_3 = \frac{1}{B_3} \text{ and } a_4 = -\frac{1}{B_3}\ln\left(\frac{\rho}{B_1}\frac{B_3}{\eta}\right).$$

This implies that $\widetilde{\alpha}_c^*(\Delta)$ grows almost linearly with respect to Δ. We conjecture with high confidence that the true $\alpha_c^*(\Delta)$ follows the same linear trend as well because (i) the approximation $\widetilde{p}_e(\Delta,\alpha)$ in (29) is getting tighter when Δ is large, as seen in Fig. 4; and (ii) our detailed asymptotics (24) and the $p_e(\infty,\alpha)$ asymptotics (32) are both provably tight within a constant factor when Δ and α are large.

We further quantify the a_1 value by

Lemma 5: For any fixed (N,K), the value of a_1 satisfies

$$a_1 = \frac{\mathbb{E}\{C_t\} - K}{4K} + o(\mathbb{E}\{C_t\} - K). \tag{34}$$

when the *gap-to-capacity* $\mathbb{E}\{C_t\} - K$ is sufficiently small.

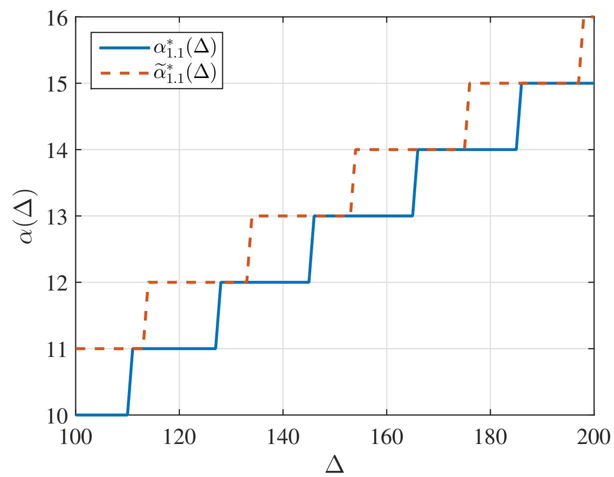

Fig. 5. The c-optimal memory length w. $(N,K) = (5,2)$, $C_t \sim B(5, 0.45)$.

Lemma 5 implies that the growth rate a_1 is small when we operate close to the capacity, a surprising result since one may expect that when $\mathbb{E}\{C_t\} - K$ is small, we need more complex codes (larger α) to protect the data. The intuition is that strong error protection can be achieved by either (i) collecting more observations (longer delay Δ), or (ii) using more complex codes (larger α). However, when $\mathbb{E}\{C_t\} - K$ is small, option (i) becomes more beneficial/critical (in terms of reducing p_e) than option (ii). The ratio $\widetilde{\alpha}_c^*(\Delta)/\Delta$ thus decreases.

A. Numerical Evaluation

Fig. 5 compares the true $\alpha_c^*(\Delta)$ versus our approximation $\widetilde{\alpha}_c^*(\Delta)$. We assume $c = 1.1$, $(N,K) = (5,2)$ and C_t being binomial with parameters $(5, 0.45)$. We numerically find $\alpha_c^*(\Delta)$ using (3) and the $p_e(\Delta,\alpha)$ formulas in [8]. The computation of $\alpha_c^*(\Delta)$ involves exhaustive search plus matrix inversions of sizes growing unboundedly with α, which is feasible only for relatively small (Δ,α). We plot $\widetilde{\alpha}_c^*(\Delta)$ by solving (33). Our $\widetilde{\alpha}_c^*(\Delta)$, which is very easy to compute, closely matches the true $\alpha_c^*(\Delta)$ with the maximum discrepancy being one. In hundreds of data points we examined with different (N,K) values and channel distributions (though unreported due to space constraints), the gap is always ≤ 1 except for the most challenging combinations of small (Δ,α) and very small $(\mathbb{E}\{C_t\} - K)$ for which the approximations (24), (29) and (32) are still loose. In the example of Fig. 5, $a_1 = 0.03088$, which is close to the approximation $(\mathbb{E}\{C_t\} - K)/4K = 0.03125$ in (34). That is, the memory length needed $\widetilde{\alpha}_c^*(\Delta)$ will eventually converge to $\approx 3.1\%$ of the target deadline Δ. This demonstrates the superior encoding efficiency of RLSCs in the sense that when Δ is large, we only need a sparse encoding matrix (similar to $\mathbf{H}^{(t)}$ in Fig. 1 but before random row deletion due to erasures) with 97% entries being 0.

V. Conclusion

We have characterized the detailed asymptotics of the delay-reliability tradeoff of infinite-memory random linear streaming codes over symbol erasure channels and discussed several important implications of the results.

References

[1] E. Martinian and C.-E. W. Sundberg, "Burst Erasure Correction Codes With Low Decoding Delay," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2494–2502, Oct. 2004.

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

[3] S. L. Fong, A. Khisti, B. Li, W. Tan, X. Zhu, and J. Apostolopoulos, "Optimal Streaming Codes for Channels With Burst and Arbitrary Erasures," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4274–4292, Jul. 2019.

[4] M. N. Krishnan, V. Ramkumar, M. Vajha, and P. V. Kumar, "Simple Streaming Codes for Reliable, Low-Latency Communication," *IEEE Communications Letters*, vol. 24, no. 2, pp. 249–253, 2020.

[5] M. N. Krishnan, D. Shukla, and P. V. Kumar, "Rate-Optimal Streaming Codes for Channels With Burst and Random Erasures," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 4869–4891, 2020.

[6] M. Rudow and K. V. Rashmi, "Streaming Codes for Variable-Size Messages," *IEEE Transactions on Information Theory*, vol. 68, no. 9, pp. 5823–5849, 2022.

[7] P.-W. Su, Y.-C. Huang, S.-C. Lin, I.-H. Wang, and C.-C. Wang, "Random Linear Streaming Codes in the Finite Memory Length and Decoding Deadline Regime," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 730–735.

[8] ——, "Random Linear Streaming Codes in the Finite Memory Length and Delay Regime — Part I: Exact Analysis," *IEEE Transactions on Information Theory*, vol. 68, no. 10, pp. 6356–6387, 2022.

[9] ——, "Error Rate Analysis for Random Linear Streaming Codes in the Finite Memory Length Regime," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 491–496.

[10] ——, "Sequentially Mixing Randomly Arriving Packets Improves Channel Dispersion Over Block-Based Designs," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 2321–2326.

[11] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[12] R. M. Fano, *Transmission of information: a statistical theory of communication*. Cambridge, Mass.: MIT Press, 1961.

[13] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, 1965.

[14] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Information and Control, Elsevier*, 1967.

[15] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[16] N. Merhav, "List decoding—random coding exponents and expurgated exponents," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6749–6759, Nov. 2014.

[17] J. Abate, G. L. Choudhury, D. M. Lucantoni, and W. Whitt, "Asymptotic Analysis of Tail Probabilities Based on the Computation of Moments," *The Annals of Applied Probability*, vol. 5, no. 4, pp. 983 – 1007, 1995.

[18] P. Moulin, "The Log-Volume of Optimal Codes for Memoryless Channels, Asymptotically Within a Few Nats," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2278–2313, 2017.

[19] D. Koutsonikolas, C.-C. Wang, and Y. C. Hu, "Efficient Network-Coding-Based Opportunistic Routing Through Cumulative Coded Acknowledgments," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1368–1381, 2011.

[20] C.-C. Wang, "On the Capacity of 1-to-K Broadcast Packet Erasure Channels With Channel Output Feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 931–956, 2012.

[21] E. Martinian, "Dynamic Information and Constraints in Source and Channel Coding," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[22] V. Ramkumar, M. N. Krishnan, M. Vajha, and P. V. Kumar, "On Information-Debt-Optimal Streaming Codes With Small Memory," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1578–1583.

[23] P. Mörters and Y. Peres, *Brownian Motion*. Cambridge University Press, 2010, vol. 30.

[24] L. Addario-Berry and B. A. Reed, *Ballot Theorems, Old and New*. Springer Berlin Heidelberg, 2008, pp. 9–35.

[25] R. P. Stanley, *Catalan Numbers*. Cambridge University Press, 2015.

[26] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W Function," *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

[27] F. Chapeau-Blondeau and A. Monir, "Numerical evaluation of the Lambert W function and application to generation of generalized Gaussian noise with exponent 1/2," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2160–2165, 2002.