Channels with Input-Correlated Synchronization Errors

Roni Con[*] João Ribeiro[†]

Abstract

"Independent and identically distributed" errors do not accurately capture the noisy behavior of real-world data storage and information transmission technologies. Motivated by this, we study channels with *input-correlated* synchronization errors, meaning that the distribution of synchronization errors (such as deletions and insertions) applied to the i-th input x_i may depend on the whole input string x.

We begin by identifying conditions on the input-correlated synchronization channel under which the channel's information capacity is achieved by a stationary ergodic input source and is equal to its coding capacity. These conditions capture a wide class of channels, including channels with correlated errors observed in DNA-based data storage systems and their multi-trace versions, and generalize prior work. To showcase the usefulness of the general capacity theorem above, we combine it with ideas of Pernice-Li-Wootters (ISIT 2022) to obtain explicit capacity-achieving codes for channels with *runlength-dependent deletions*, motivated by error patterns observed in DNA-based data storage systems.

[*]Department of Computer Science, Technion - Israel Institute of Technology, Haifa. roni.con93@gmail.com

[†]Instituto de Telecomunicações and Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa. jribeiro@tecnico.ulisboa.pt

Contents

1 Introduction **3**
- 1.1 Our contributions . . . 3
- 1.2 Related work . . . 4

2 Preliminaries **5**
- 2.1 Notation . . . 5
- 2.2 Channels . . . 5
- 2.3 Entropy and information rates for stochastic processes and notions of capacity . . . 5
- 2.4 A strengthening of Fekete's lemma . . . 7

3 Capacity theorems for channels with input-correlated synchronization errors **7**
- 3.1 Admissible channels . . . 7
- 3.2 Existence of relevant limits for admissible channels . . . 8
- 3.3 Information capacity of admissible channels is achieved by stationary ergodic process 11
- 3.4 Information capacity of admissible channels is achieved by Markov process . . . 14
- 3.5 Coding capacity equals information capacity, and existence of dense codes from stationary ergodic processes . . . 15
  - 3.5.1 Convergence of information density for block-independent process . . . 15
  - 3.5.2 Capacity-achieving codes for admissible channels . . . 17
  - 3.5.3 Dense capacity-achieving codes for admissible channels . . . 19

4 Some special cases of our capacity theorems **20**
- 4.1 The Mao-Diggavi-Kannan ISI model . . . 20
- 4.2 Multi-trace channels with input-correlated synchronization errors . . . 21
- 4.3 Capacity theorems for trimming synchronization channels . . . 23
- 4.4 Channels with runlength-dependent deletions . . . 23

5 Efficient capacity-achieving codes for channels with runlength-dependent deletions **26**

6 Acknowledgments **29**

1 Introduction

Errors which cause loss of synchronization between sender and receiver, such as deletions, insertions, and replications, occur in various communications and data storage technologies, with DNA-based data storage being a notable recent example. Despite considerable effort, pinning down the capacity and designing efficient nearly-optimal codes for channels with synchronization channels remain major problems in information and coding theory. The surveys by Mitzenmacher [Mit09], Mercier, Bhargava, and Tarokh [MBT10], and Cheraghchi and Ribeiro [CR21] provide in-depth discussions of the many challeneges encountered when dealing with synchronization errors.

Most prior work on channels with synchronization errors has focused on i.i.d. errors. However, synchronization errors in real-world systems do not satisfy this assumption. For example, it has been observed in empirical analyses of widely used DNA sequencing technologies [RRC+13, HMG19] that short substrings of DNA strands (which are written using a 4-symbol alphabet A, C, G, T) with either a very high (above 75%) or very low (below 25%) concentration of G's and C's experience higher deletion rates [RRC+13, Figure 4], and that longer runs of the same symbol in DNA strands experience higher deletion rates than shorter runs during sequencing [RRC+13, Figure 5].

Motivated by this, we study a general class of channels with synchronization errors where the error distribution of the i-th input symbol x_i may depend on the whole input x. We also consider "multi-trace" versions of these channels, where the input x is sent through multiple independent channels, generating multiple channel outputs (*traces*) at the receiver end, which is especially relevant in DNA-based data storage.

1.1 Our contributions

Our first contribution is a capacity theorem for a general class of channels with input-correlated synchronization errors, which we call *admissible channels*. The precise definition of an admissible channel is given in Section 3.1. As we show in Section 4, this class includes as special cases the channel model of Mao, Diggavi, and Kannan [MDK18], multi-trace channels with input-correlated synchronization errors, and, more concretely, channels with runlength-dependent deletions where bits in runs of length ℓ are deleted independently with probability $d(\ell)$, for an arbitrary function $d : \mathbb{N} \to [0, 1]$.

Theorem 1 (Informal, see Theorem 3 for a formal statement)**.** *Let Z be an admissible channel. Then, its information capacity equals its coding capacity, and the information capacity is achieved by stationary ergodic sources.*

As observed by Pernice, Li, and Wootters [PLW22], a standard but quite useful consequence of Theorem 1 is that admissible channels admit capacity-achieving codes with additional structure. In particular, admissible channels Z with binary input alphabet admit capacity-achieving codes $\mathcal{C}$ such that every short substring of every codeword $c \in \mathcal{C}$ has not-too-small Hamming weight (see Theorem 4 for a formal statement).

To concretely showcase the usefulness of Theorem 1, we combine this consequence with an approach originally applied to channels with i.i.d. synchronization errors in [PLW22] to obtain efficient capacity-achieving codes for channels with "bounded" runlength-dependent deletions. Namely, we will define a class of channels we call BDC-R-L-Bounded(d, μ, M) to be runlength-dependent channels defined with a monotonically increasing function $d(\ell) : \mathbb{N} \to [0, 1]$, number $\mu \in (0, 1)$, and an integer M such that $d(\ell) = d(M) < 1 - \mu$ for all $\ell \geq M$ (see Definition 11).

For these channels, we prove

Theorem 2 (Informal, see Theorem 10 for a formal statement)**.** *Let $\varepsilon > 0$ and let μ and M be constants, and let $d(\ell)$ monotonically increasing such that $d(\ell) = d(M) < 1-\mu$ for all $\ell \geq M$. Then, there exists a family of binary codes $\{C_i\}_{i=1}^{\infty}$ that are robust for the* BDC-R-L-Bounded(d, μ, M) *channel with rate of $C_i >$* Cap(BDC-R-L-Bounded$(d, \mu, M)) - \varepsilon$. *Moreover, C_i is encodable in linear time and decodable in quasi linear time.*

1.2 Related work

Capacity theorems for channels with synchronization errors. The first work to study this topic was by Dobrushin [Dob67], who obtained capacity theorems for channels that apply i.i.d. synchronization errors. Recently, there has been interest in extending such capacity theorems beyond i.i.d. errors. Mao, Diggavi, and Kannan [MDK18] consider channels combining synchronization errors and (bounded) intersymbol interference as a model of nanopore-based sequencing. More precisely, the behavior of the channel on input x_i is some function of $x_i, x_{i-1}, \dots, x_{i-\ell}$, for some memory threshold ℓ. They show that the information and coding capacity of these channels coincide, generalizing Dobrushin's result [Dob67], but do not show that capacity is achieved by a stationary ergodic (or Markov) source. Capacity theorems for a related (more concrete) model of nanopore-based sequencing with noisy duplications have also been studied by McBain, Saunderson, and Viterbo [MVS24, MSV24]. Li and Tan [LT21] consider a channel obtained from the concatenation of a standard deletion channel with i.i.d. deletions and a finite-state discrete memoryless channel. They show that the capacity of this channel is achieved by Markov processes, which implies that the polar codes developed by Tal, Pfister, Fazeli, and Vardy [TPFV22] achieve capacity on the i.i.d. deletion channel (a special case of this result was proved earlier in [TPFV22]). Morozov and Duman [MD24] show that information and coding capacities coincide for channels that introduce deletions and insertions with Markovian memory, in the sense that the behavior of the channel on the i-th input bit depends on the current state of an underlying stationary ergodic finite state Markov chain (whose states are updated independently of past inputs).

The models we study are incomparable to those of [LT21, MD24], and our models and capacity theorems generalize those of [MDK18]. We discuss the relationship to the model and results of [MDK18] in more detail. In [MDK18], the channel behavior on the i-th input bit x_i may depend only on a *bounded* window of input bits. In contrast, in our channel model the error distribution for the i-th input bit x_i is some function of *the whole input* x, satisfying some additional assumptions. We show in Section 4.1 that the channel model from [MDK18] satisfies the assumptions required for the application of our capacity theorems, and so our results generalize the corresponding results of [MDK18]. Moreover, we show that stationary ergodic sources achieve the information capacity of these channels, which is particularly relevant for constructing efficient capacity-achieving codes.

Furthermore, our framework also captures interesting scenarios that fall outside the scope of [MDK18]. In Section 4.2 we show that our framework implies capacity theorems for *multi-trace* channels with correlated synchronization errors, where on input x the receiver learns multiple i.i.d. realizations of the channel output $Z(x)$. This setting is especially relevant in the context of DNA-based data storage systems with nanopore-based sequencing [CGMR20, BLS20]. Also, in Section 4.4 we show that our framework includes as special cases deletion channels where the deletion of a bit x_i may depend arbitrarily on the length of the run where x_i is included (in particular, beyond a bounded window around x_i).

Efficient coding for channels with synchronization errors. There has been significant interest in the design of efficiently encodable and decodable codes for channels with synchronization errors. We discuss the progress most relevant to our work. Guruswami and Li [GL19] and later Con and Shpilka [CS22] obtained efficient codes for the i.i.d. binary deletion channel with rate $\Theta(1-d)$, where d is the deletion probability. Later, Tal, Pfister, Fazeli, and Vardy [TPFV22, PT21] and later Tal and Arava [AT23] (combined with a result from [LT21]) designed efficient polar codes achieving the capacity of a family of channels with i.i.d. insertions, deletions, and substitutions, generalizing the i.i.d. deletion channel. Other constructions of efficient capacity-achieving codes for channels with i.i.d. synchronization errors were presented in [Rub22, PLW22], which achieve slightly faster decoding and slightly smaller decoding error probability than the polar coding constructions. Of particular note, the general framework of Pernice, Li, and Wootters [PLW22] yields efficient capacity-achieving codes for a large class of *repeat channels* – these are channels that independently replicate each input bit according to some replication distribution over the naturals (e.g., Bernoulli, Poisson, geometric).[1] This was accomplished by combining a marker-based construction with the capacity theorem of Dobrushin [Dob67], which applies to i.i.d. repeat channels.

A common feature of the works discussed above is that they only consider channels with i.i.d. synchronization errors. In contrast, we study channels with correlated synchronization errors. In particular, we obtain a capacity theorem that applies to a wide class of channels with correlated synchronization errors, which we show can be used to design efficient capacity-achieving codes for synchronization channels with relevant correlations.

2 Preliminaries

2.1 Notation

We denote random variables by uppercase roman letters such as X, Y, and Z. In this work, we will only work with random variables supported on discrete sets. We use $X \to Y \to Z$ to denote the fact that these three random variables form a Markov chain (i.e., Z is conditionally independent of X given Y), and write $X \sim Y$ if random variables X and Y follow the same distribution. We use $\mathbb{E}[X]$ to denote the expected value of a random variable X supported on a subset of $\mathbb{R}$, and $H(X)$ to denote its Shannon entropy.

For a sequence $x = (x_i)_{i\in\mathbb{N}}$, we use x_m^n to denote the subsequence $x_m, x_{m+1}, \dots, x_n$. We use log to denote the base-2 logarithm. For an integer $n \geq 1$, we write $[n] = \{1, 2, \dots, n\}$.

2.2 Channels

We consider channels Z with finite input alphabet Σ_{in} and discrete output alphabet Σ_{out}. On input $x \in \Sigma_{\mathsf{in}}^n$, the channel outputs $Z(x)$ according to some probabilistic rule. We allow the support of $Z(x)$ to be quite general – we will consider settings where $Z(x) \in \Sigma_{\mathsf{out}}^*$ for some discrete output alphabet Σ_{out}, but also settings where $Z(x)$ corresponds to multiple channel traces with input x, in which case $Z(x) \in (\Sigma_{\mathsf{out}}^*)^t$ for some integer $t \geq 1$.

Later on we will impose constraints on the behavior of $Z(x)$ to obtain capacity theorems. We will also write "$Z(X), Z(Y)$" to mean that Z is applied independently to the inputs X and Y.

2.3 Entropy and information rates for stochastic processes and notions of capacity

In this section, we define various types of "channel capacity", and state some basic bounds.

[1] The existence of such efficient capacity-achieving codes does not mean that we now can *determine* the capacity of these channels. We see the contribution of [TPFV22, PT21, Rub22, PLW22] mainly as turning capacity lower bounds into efficient codes with the corresponding rate.

Definition 1 (Entropy rate). *For a process $X = (X_i)_{i\in\mathbb{N}}$, we define the* entropy rate *of X, denoted by $H(X)$, as*

$$H(X) = \lim_{n\to\infty} \frac{H(X_1^n)}{n},$$

if this limit exists.

Definition 2 (Information rate). *For a channel Z and input process $X = (X_i)_{i\in\mathbb{N}}$, we define the* information rate *achievable by X over Z as*

$$I(X; Z(X)) = \liminf_{n\to\infty} \frac{I(X_1^n; Z(X_1^n))}{n}.$$

Definition 3 (Information capacity). *Given a channel Z, we define its* information capacity*, denoted by $\mathsf{ICap}(Z)$, as*

$$\mathsf{ICap}(Z) = \liminf_{n\to\infty} \sup_{P_{X_1^n}} \frac{I(X_1^n; Z(X_1^n))}{n},$$

where the supremum is taken over all distributions of input processes $X = (X_i)_{i\in\mathbb{N}}$.

We now introduce some useful definitions about stochastic processes.

Definition 4 (Block-independent process). *We say that a stochastic process $X = (X_i)_{i\in\mathbb{N}}$ is* block-independent *with blocklength b if for any integer $t \geq 1$ and $n = tb$ we have*

$$\Pr[X_1^n = x_1^n] = \prod_{i=1}^{t} \Pr[X_1^b = x_{(i-1)b+1}^{ib}].$$

Definition 5 (Stationary ergodic process). *We say that a stochastic process $X = (X_i)_{i\in\mathbb{N}}$ is* stationary *if $(X_1, \dots, X_n) \sim (X_{1+\tau}, \dots, X_{n+\tau})$ for any integers $n, \tau \in \mathbb{N}$. Moreover, we say that X is* stationary ergodic *if X is stationary and for any function $f \in L^1$ we almost surely have*

$$\mathbb{E}[f(X_1)] = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(X_j).$$

Definition 6 (Stationary capacity). *Given a channel Z, we define its* stationary capacity*, denoted by $\mathsf{SCap}(Z)$, as*

$$\mathsf{SCap}(Z) = \sup_X I(X; Z(X)),$$

where the supremum is taken over all stationary ergodic input processes $X = (X_i)_{i\in\mathbb{N}}$.

An input process $X = \{X_i\}_{i\in\mathbb{N}}$ is *m-th order Markov* if

$$\Pr[X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_1 = x_1] = \Pr[X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_{n-m} = x_{n-m}]$$

for any n and $x_1, \dots, x_n$.

Definition 7 (m-th order Markov capacity). *Given a channel Z, we define its* m-th order Markov capacity*, denoted by $\mathsf{SCap}^{(m)}(Z)$, as*

$$\mathsf{SCap}^{(m)}(Z) = \sup_X I(X; Z(X)),$$

where the supremum is taken over all m-th order stationary Markov input processes $X = (X_i)_{i\in\mathbb{N}}$.

Remark 1. We have $\mathsf{ICap}(Z) \geq \mathsf{SCap}(Z) \geq \mathsf{SCap}^{(m)}(Z)$ for any channel Z and any integer $m \geq 0$.

Before we define the coding capacity of a channel Z, we need some auxiliary definitions.

Definition 8 ((n, R, ε)-code for a channel)**.** *Let Z be a channel with finite input alphabet Σ and output alphabet Σ_{out}. We say that $\mathcal{C} \subseteq \Sigma^n$ is an (n, R, ε)*-code for Z *if $|\mathcal{C}| \geq 2^{Rn}$ and there exists a deterministic function $\mathsf{Dec} : \Sigma_{\mathsf{out}}^* \to \Sigma^n$ such that $\Pr[\mathsf{Dec}(Z(C)) \neq C] \leq \varepsilon$, where C is uniformly distributed over $\mathcal{C}$ (i.e., the average decoding error probability of $\mathcal{C}$ is at most ε).*

Definition 9 (Achievable rate)**.** *Let Z be a channel. We say that a real number $R > 0$ is an* achievable rate for Z *if there exists a family of codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ and an integer n_0 such that each $\mathcal{C}_n$ is an (n, R_n, ε_n)-code for Z with $R_n \geq R$ for all $n \geq n_0$ and $\lim_{n\to\infty} \varepsilon_n = 0$.*

Definition 10 (Coding capacity)**.** *Given a channel Z, we define its* coding capacity*, denoted by $\mathsf{CCap}(Z)$, as the supremum of all $R \geq 0$ that are achievable rates for Z.*

Remark 2. It follows via Fano's inequality that $\mathsf{CCap}(Z) \leq \mathsf{ICap}(Z)$ for any channel Z. See, e.g., [PW24, Theorem 19.7].

2.4 A strengthening of Fekete's lemma

We will use the following strengthening of Fekete's lemma due to de Bruijn and Erdős [dBE52] (also used in [MD24]). See [FR20] for an excellent discussion on this topic.

Lemma 1 (Strengthened Fekete's lemma [dBE52])**.** *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence that is "almost" subadditive, in the sense that*

$$a_{n+m} \leq a_n + a_m + f(n+m)$$

for all $n \leq m \leq 2n$ and some f such that $\sum_{n\in\mathbb{N}} f(n)/n^2$ converges. Then, $\lim_{n\to\infty} a_n/n$ exists.

3 Capacity theorems for channels with input-correlated synchronization errors

3.1 Admissible channels

We will show capacity theorems for all channels Z (with finite input alphabet) for which there exists a special associated channel $Z^\star$ with the properties below. We call Z *admissible* (with respect to $Z^\star$).

1. **Bounded output entropy:** There exists a constant $c > 0$ such that for every input process $X = (X_i)_{i\in\mathbb{N}}$ and every integer $n \geq 1$ we have $H(Z^\star(X_1^n)) \leq cn$.
2. **Rate preservation:** For any process $X = \{X_i\}_{i\in\mathbb{N}}$ we have $I(X; Z(X)) = I(X; Z^\star(X))$.
3. **Concatenation:** For any process $X = \{X_i\}_{i\in\mathbb{N}}$ and indices $1 \leq m \leq n$, we have that
$$X_1^n \to Z^\star(X_1^m), Z^\star(X_{m+1}^n) \to Z^\star(X_1^n).$$
4. **Partition:**

(a) **Prefix/suffix-partition:** Fix any integer $\tau \geq 1$. Then, there exists a non-decreasing sequence $\gamma_m = o(m)$ such that $\sum_{m \in \mathbb{N}} \gamma_m / m^2$ converges and for any process $X = \{X_i\}_{i \in \mathbb{N}}$ and integer $n \geq \tau$ there exist random variables W_{pre} and W_{suf} such that (1) $X_1^n \to Z^\star(X_1^n), W_{\text{pre}} \to Z^\star(X_1^\tau), Z^\star(X_{\tau+1}^n)$, (2) $X_1^n \to Z^\star(X_1^n), W_{\text{suf}} \to Z^\star(X_1^{n-\tau}), Z^\star(X_{n-\tau+1}^n)$, and (3) $H(W_{\text{pre}}), H(W_{\text{suf}}) \leq \gamma_n$.

(b) **Amortized block-partition:** There exists a sequence $\alpha_m = o(m)$ such that for any process $X = \{X_i\}_{i \in \mathbb{N}}$, blocklength b, and number of blocks t there exists a random variable W such that $X_1^{tb} \to Z^\star(X_1^{tb}), W \to Z^\star(X_1^b), \ldots, Z^\star(X_{(t-1)b+1}^{tb})$, and $H(W) \leq t \cdot \alpha_b$.

5. **Amortized preimage size:** There exists a sequence $\beta_m = o(m)$ such that for any process $X = \{X_i\}_{i \in \mathbb{N}}$, blocklength b, and number of blocks t, there exists a random variable Y such that $X_1^{tb} \to Z^\star(X_1^b), \ldots, Z^\star(X_{(t-1)b+1}^{tb}) \to Y$ and a deterministic function ϕ such that

$$Z^\star(X_1^{tb}) = \phi(Z^\star(X_1^b), \ldots, Z^\star(X_{(t-1)b+1}^{tb}), Y)$$

and $\max_z \log |\phi^{-1}(z)| \leq t \cdot \beta_b$.

We will prove the following results for admissible channels.

Theorem 3 (Capacity theorem for admissible channels)**.** *Suppose that Z is an admissible channel. Then,*

$$\mathsf{ICap}(Z) = \mathsf{SCap}(Z) = \lim_{m \to \infty} \mathsf{SCap}^{(m)}(Z) = \mathsf{CCap}(Z).$$

The proof of Theorem 3 proceeds via a series of theorems, where we adapt and generalize approaches from [Dob67, LT21, MD24]. The equality $\mathsf{ICap}(Z) = \mathsf{SCap}(Z)$ corresponds to Theorem 5 in Section 3.3. The equality $\mathsf{ICap}(Z) = \lim_{m \to \infty} \mathsf{SCap}^{(m)}(Z)$ corresponds to Theorem 6 in Section 3.4. The equality $\mathsf{ICap}(Z) = \mathsf{CCap}(Z)$ corresponds to Theorem 8 in Section 3.5.

The next result follows in a standard manner from Theorem 3 using the approach of [PLW22], and is relevant for the construction of efficient capacity-achieving codes for admissible channels. We prove it in Section 3.5.3.

Theorem 4 (Dense capacity-achieving codes for admissible channels)**.** *Let Z be an admissible channel with binary input alphabet. Then, for any $\varepsilon, \zeta > 0$ there exist $\gamma \in (0, 1/2)$ and integers $b(\varepsilon)$ and $t(\varepsilon, \zeta)$ depending only on ε and ζ such that for all $t \geq t(\varepsilon, \zeta)$ there exists a code $\mathcal{C}$ with blocklength $n = t \cdot b(\varepsilon)$, rate $R \geq \mathsf{ICap}(Z) - \varepsilon$, and maximal decoding error probability ε over Z such that for all codewords $c \in \mathcal{C}$ we have $\gamma \zeta n \leq w(c_i^{i+\zeta n}) \leq (1 - \gamma)\zeta n$ for all $i \in [(1 - \zeta)n]$, where $w(\cdot)$ denotes Hamming weight.*

The properties for admissibility laid out above are sufficient for Theorems 3 and 4 to hold, but it is conceivable that they are not necessary. We leave it as an interesting direction for future work to simplify the set of sufficient properties.

3.2 Existence of relevant limits for admissible channels

Let Z be an arbitrary admissible channel with respect to $Z^\star$. Via applications of Fekete's lemma (Lemma 1), we begin by showing that the limits inferior in the definitions of capacities and information rates can be replaced by limits.

Lemma 2. *If Z is an admissible channel with respect to $Z^\star$, then*

$$\mathsf{ICap}(Z) = \mathsf{ICap}(Z^\star) = \lim_{n\to\infty} \sup_{X_1^n} \frac{I(X_1^n; Z^\star(X_1^n))}{n},$$

and the limit on the right-hand side exists.

Proof. By Lemma 1, it suffices to show that the sequence

$$a_n = \sup_{X_1^n} I(X_1^n; Z^\star(X_1^n))$$

is subadditive, i.e., $a_{n+m} \le a_n + a_m$ for all n, m. Using Item 3, the data processing inequality gives

$$\begin{aligned}
a_{n+m} &= \sup_{X_1^{n+m}} I(X_1^{n+m}; Z^\star(X_1^{n+m})) \\
&\le \sup_{X_1^{n+m}} I(X_1^{n+m}; Z^\star(X_1^n), Z^\star(X_{n+1}^{n+m})) \\
&= \sup_{X_1^{n+m}} [I(X_1^{n+m}; Z^\star(X_1^n)) + I(X_1^{n+m}; Z^\star(X_{n+1}^{n+m})|Z^\star(X_1^n))] \\
&= \sup_{X_1^{n+m}} [I(X_1^n; Z^\star(X_1^n)) + I(X_1^{n+m}; Z^\star(X_{n+1}^{n+m})|Z^\star(X_1^n))] \\
&\le \sup_{X_1^{n+m}} [I(X_1^n; Z^\star(X_1^n)) + I(X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m}))] \\
&\le \sup_{X_1^n} I(X_1^n; Z^\star(X_1^n)) + \sup_{X_{n+1}^{n+m}} I(X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m})) \\
&= a_n + a_m.
\end{aligned}$$

The first inequality uses Item 3. The second inequality uses the fact that $X_1^n, X_{n+1}^{n+m} \to X_{n+1}^{n+m} \to Z^\star(X_{n+1}^{n+m})$, and so

$$\begin{aligned}
I(X_1^n, X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m})|Z^\star(X_1^n)) &= H(Z^\star(X_{n+1}^{n+m})|Z^\star(X_1^n)) - H(Z^\star(X_{n+1}^{n+m})|X_1^n, X_{n+1}^{n+m}, Z^\star(X_1^n)) \\
&= H(Z^\star(X_{n+1}^{n+m})|Z^\star(X_1^n)) - H(Z^\star(X_{n+1}^{n+m})|X_{n+1}^{n+m}) \\
&\le H(Z^\star(X_{n+1}^{n+m})) - H(Z^\star(X_{n+1}^{n+m})|X_{n+1}^{n+m}) \\
&= I(X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m})).
\end{aligned}$$

The third inequality holds because the quantity on the fifth line is maximized by taking X_1^n and X_{n+1}^{n+m} to be independent, since $Z^\star$ is acting independently on each of X_1^n and X_{n+1}^{n+m}. □

Lemma 3. *Let X be an arbitrary stationary process and Z an admissible channel with respect to $Z^\star$. Then,*

$$I(X; Z(X)) = I(X; Z^\star(X)) = \lim_{n\to\infty} \frac{I(X_1^n; Z^\star(X_1^n))}{n},$$

the limit on the right-hand side exists, and $\mathsf{SCap}(Z) = \mathsf{SCap}(Z^\star)$.

Proof. Again, by Lemma 1 it suffices to show that $a_n = I(X_1^n; Z^\star(X_1^n))$ is a subadditive sequence. We have

$$\begin{aligned}
a_{n+m} &= I(X_1^{n+m}; Z^\star(X_1^{n+m})) \\
&\le I(X_1^{n+m}; Z^\star(X_1^n), Z^\star(X_{n+1}^{n+m}))
\end{aligned}$$

$$
\begin{aligned}
&= I(X_1^n, X_{n+1}^{n+m}; Z^\star(X_1^n)) + I(X_1^n, X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m}) | Z^\star(X_1^n)) \\
&= I(X_1^n; Z^\star(X_1^n)) + I(X_1^n, X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m}) | Z^\star(X_1^n)) \\
&\leq I(X_1^n; Z^\star(X_1^n)) + I(X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m})) \\
&= I(X_1^n; Z^\star(X_1^n)) + I(X_1^m; Z^\star(X_1^m)) \\
&= a_n + a_m,
\end{aligned}
$$

as desired. The first inequality uses Item 3. The second equality follows from the chain rule for mutual information. The third equality uses the fact that $X_1^n, X_{n+1}^{n+m} \to X_1^n \to Z^\star(X_1^n)$. The second inequality uses the fact that $X_1^n, X_{n+1}^{n+m} \to X_{n+1}^{n+m} \to Z^\star(X_{n+1}^{n+m})$, analogously to the proof of Lemma 2. The fourth equality uses the fact that X is stationary, and so $X_{n+1}^{n+m} \sim X_1^m$. ☐

Lemma 4. *Let X be a block independent process and suppose that Z is admissible with respect to $Z^\star$. Then,*

$$
I(X; Z(X)) = I(X; Z^\star(X)) = \lim_{n\to\infty} \frac{I(X_1^n; Z^\star(X_1^n))}{n},
$$

and the limit on the right-hand side exists.

Proof. Consider the sequence $a_n = I(X_1^n; Z^\star(X_1^n))$. Let b be the blocklength of X. Then, by Item 3, for any $m, n > b$ we have

$$
\begin{aligned}
a_{n+m} &= I(X_1^{n+m}; Z^\star(X_1^{n+m})) \\
&\leq I(X_1^{n+m}; Z^\star(X_1^n), Z^\star(X_{n+1}^{n+m})) \\
&\leq I(X_1^n; Z^\star(X_1^n)) + I(X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m})).
\end{aligned}
$$

Construct X' by trimming bits from the beginning of X_{n+1}^{n+m} so that $X' = X_{tb+1}^{n+m}$ for some integer t. If m' denotes the length of X', from the block independence of X we get that

$$
I(X'; Z^\star(X')) = I(X_1^{m'}; Z^\star(X_1^{m'})).
$$

Since we trim $r < b$ symbols from X_{n+1}^{n+m} to obtain X', we have

$$
\begin{aligned}
I(X_{n+1}^{n+m}; Z^\star(X_{n+1}^{n+m})) &\leq I(X_{n+1}^{n+m}; Z^\star(X_{n+1}^{tb}), Z^\star(X_{tb+1}^{n+m})) \\
&\leq I(X_{tb+1}^{n+m}; Z^\star(X_{tb+1}^{n+m})) + H(Z^\star(X_{n+1}^{tb})) \\
&\leq I(X_{tb+1}^{n+m}; Z^\star(X_{tb+1}^{n+m})) + cb \\
&= I(X_1^{m'}; Z^\star(X_1^{m'})) + cb \\
&\leq I(X_1^m; Z^\star(X_1^{m'}), Z^\star(X_{m'+1}^m)) + cb \\
&\leq I(X_1^m; Z^\star(X_1^m)) + \gamma_m + cb
\end{aligned}
$$

for a non-decreasing sequence $\gamma_m = o(m)$ such that $\sum_{m\in\mathbb{N}} \gamma_m/m^2$ converges. The first inequality uses Item 3. The third inequality uses Item 1 (more precisely, $H(Z^\star(X_{n+1}^{tb})) \leq c(tb - (n+1)) \leq cb$ for a fixed constant $c > 0$. The first equality uses the block independence of X. The fifth inequality uses Item 4a with $\tau = r < b$.

Therefore, we conclude that for any $b < n \leq m \leq 2n$ we have $a_{n+m} \leq a_n + a_m + f(n+m)$, where $f(n) = \gamma_n + cb$ (here, we use that γ_m is non-decreasing). Since $\sum_{n\in\mathbb{N}} f(n)/n^2$ converges (because $\sum_{n\in\mathbb{N}} \gamma_n/n^2$ converges by Item 4a), Lemma 1 implies that $\lim_{n\to\infty} a_n/n$ exists, as desired. ☐

3.3 Information capacity of admissible channels is achieved by stationary ergodic process

We show the following.

Theorem 5. *Suppose that the channel Z is admissible. Then,*

$$\mathsf{ICap}(Z) = \mathsf{SCap}(Z).$$

Suppose that Z is admissible with respect to $Z^\star$. By Item 2, it suffices to show that Theorem 5 holds for $Z^\star$. Fix an arbitrary $\varepsilon > 0$. Recalling Lemma 2, let b be a sufficiently large integer so that

$$\sup_{\tilde{X}_1^b} \frac{I(\tilde{X}_1^b; Z^\star(\tilde{X}_1^b))}{b} \geq \lim_{n\to\infty} \sup_{\tilde{X}_1^n} \frac{I(\tilde{X}_1^n; Z^\star(\tilde{X}_1^n))}{n} - \varepsilon = \mathsf{ICap}(Z^\star) - \varepsilon.$$

Furthermore, let X_1^b be such that

$$\frac{I(X_1^b; Z^\star(X_1^b))}{b} \geq \sup_{\tilde{X}_1^b} \frac{I(\tilde{X}_1^b; Z^\star(\tilde{X}_1^b))}{b} - \varepsilon \geq \mathsf{ICap}(Z^\star) - 2\varepsilon. \tag{1}$$

Let $p_{X_1^b}(\cdot)$ denote the correspoding PMF. Following the approach of [Fei59, LT21], consider the following process $\overline{X}$. First, define the block independent process $\hat{X} = \{\hat{X}_i\}_{i\in\mathbb{N}}$ with blocklength b and probability mass function

$$p_{\hat{X}_1^{tb}}(x_1^{tb}) = \prod_{i=1}^{t} p_{X_1^b}(x_{(i-1)b+1}^{ib}).$$

Write $\hat{X}_i^{[j]} = \hat{X}_{i+j}$. Let V be uniformly distributed over $\{0, 1, \ldots, b-1\}$, and set $\overline{X}_i = \hat{X}_i^{[V]} = \hat{X}_{i+V}$ for every $i \in \mathbb{N}$. Then, it holds that $\overline{X}$ is stationary and ergodic.

By Lemma 3, we know that $I(\overline{X}; Z^\star(\overline{X})) = \lim_{n\to\infty} \frac{I(\overline{X}_1^n; Z^\star(\overline{X}_1^n))}{n}$ since $\overline{X}$ is stationary. We will show that

$$I(\overline{X}; Z^\star(\overline{X})) \geq \frac{I(X_1^b; Z^\star(X_1^b))}{b} - \varepsilon. \tag{2}$$

Combined with Equation (1), this implies that

$$\mathsf{SCap}(Z^\star) \geq I(\overline{X}; Z^\star(\overline{X})) \geq \mathsf{ICap}(Z^\star) - 3\varepsilon.$$

Since ε was arbitrary, it follows that $\mathsf{SCap}(Z^\star) = \mathsf{ICap}(Z^\star)$, and so also $\mathsf{SCap}(Z) = \mathsf{ICap}(Z)$. This yields Theorem 5.

It remains to show Equation (2). We do this by a combination of two lemmas.

Lemma 5. *We have*

$$I(\overline{X}; Z^\star(\overline{X})) = \sum_{j=0}^{b-1} \frac{1}{b} I(\hat{X}^{[j]}; Z^\star(\hat{X}^{[j]})) = I(\hat{X}; Z^\star(\hat{X})),$$

and these limits exist.

Proof. First, we prove that

$$I(\hat{X}^{[j]}; Z^\star(\hat{X}^{[j]})) = I(\hat{X}; Z^\star(\hat{X}))$$

for all $j \in \{0, 1, \ldots, b-1\}$ (in particular, these quantities exist for all $0 \le j < b$, since $\hat{X}$ is stationary). For the sake of exposition we focus on $j = 1$. The argument is analogous for other choices of j. First, note that by Item 4a with $\tau = b - 1$ there exists a sequence γ_m such that $\gamma_m = o(m)$ and $\sum_{m \in \mathbb{N}} \gamma_m/m^2$ converges and a random variable W such that $H(W) \le \gamma_{tb}$ and $Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})$ is completely determined by $Z^\star(\hat{X}_1^{[1]tb-1}), W$. Furthermore, by Item 3, $Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})$ completely determine $Z^\star(\hat{X}_1^{[1]tb-1})$. This means that

$$I(\hat{X}_1^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]tb-1})) \le I(\hat{X}_1^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})) \tag{3}$$

and

$$\begin{aligned} I(\hat{X}_1^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]tb-1})) &\ge I(\hat{X}_1^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})) - H(W) \\ &\ge I(\hat{X}_1^{n+m}; Z^\star(\hat{X}_1^{n}), Z^\star(\hat{X}_{n+1}^{n+m})) - \gamma_{tb}. \end{aligned} \tag{4}$$

Therefore,

$$\begin{aligned} I(\hat{X}_1^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]tb-1})) &\ge I(\hat{X}_1^{[1]b-1}, \hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})) - \gamma_{tb} \\ &= I(\hat{X}_1^{[1]b-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})) \\ &\quad + I(\hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1}) | \hat{X}_1^{[1]b-1}) - \gamma_{tb} \\ &\ge I(\hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1}) | \hat{X}_1^{[1]b-1}) - \gamma_{tb} \\ &= I(\hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_b^{[1]tb-1})) - \gamma_{tb} \\ &= I(\hat{X}_{b+1}^{tb}; Z^\star(\hat{X}_{b+1}^{tb})) - \gamma_{tb} \\ &= I(\hat{X}_1^{(t-1)b}; Z^\star(\hat{X}_1^{(t-1)b})) - \gamma_{tb}. \end{aligned}$$

The first inequality uses Equation (4). The second and fourth equalities use the fact that $\hat{X}$ is block-independent with blocklength b, and so $\hat{X}_b^{[1]tb-1}$ is independent of $\hat{X}_1^{[1]b-1}$ and is identically distributed to $\hat{X}_1^{(t-1)b}$. Similarly,

$$\begin{aligned} I(\hat{X}_1^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]tb-1})) &\le I(\hat{X}_1^{[1]b-1}, \hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})) \\ &= I(\hat{X}_1^{[1]b-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1})) \\ &\quad + I(\hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1}) | \hat{X}_1^{[1]b-1}) \\ &\le I(\hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]b-1}), Z^\star(\hat{X}_b^{[1]tb-1}) | \hat{X}_1^{[1]b-1}) + b \log |\Sigma_{\mathsf{in}}| \\ &= I(\hat{X}_b^{[1]tb-1}; Z^\star(\hat{X}_b^{[1]tb-1})) + b \log |\Sigma_{\mathsf{in}}| \\ &= I(\hat{X}_{b+1}^{tb}; Z^\star(\hat{X}_{b+1}^{tb})) + b \log |\Sigma_{\mathsf{in}}| \\ &= I(\hat{X}_1^{(t-1)b}; Z^\star(\hat{X}_1^{(t-1)b})) + b \log |\Sigma_{\mathsf{in}}|, \end{aligned}$$

where the first inequality uses Equation (3) and the second inequality uses the fact that $H(\hat{X}_1^{[1]b-1}) \le b \log |\Sigma_{\mathsf{in}}|$. As a result, we conclude that (below, we write $a \pm \delta$ for a real number in the interval $[a - \delta, a + \delta]$)

$$I(\hat{X}^{[1]}; Z^\star(\hat{X}^{[1]})) = \lim_{t \to \infty} \frac{I(\hat{X}_1^{[1]tb-1}; Z^\star(\hat{X}_1^{[1]tb-1}))}{tb - 1}$$

$$
\begin{aligned}
&= \lim_{t\to\infty} \frac{I(\hat{X}_1^{(t-1)b}; Z^\star(\hat{X}_1^{(t-1)b})) \pm (b\log|\Sigma_{\mathsf{in}}| + \gamma_{tb})}{tb-1} \\
&= \lim_{t\to\infty} \frac{(t-1)b}{tb-1} \cdot \frac{I(\hat{X}_1^{(t-1)b}; Z^\star(\hat{X}_1^{(t-1)b})) \pm (b\log|\Sigma_{\mathsf{in}}| + \gamma_{tb})}{(t-1)b} \\
&= \lim_{t\to\infty} \frac{I(\hat{X}_1^{(t-1)b}; Z^\star(\hat{X}_1^{(t-1)b}))}{(t-1)b} \\
&= I(\hat{X}; Z(\hat{X})).
\end{aligned}
$$

The fourth equality uses the fact that $|\Sigma_{\mathsf{in}}|$ is a finite constant and $\lim_{t\to\infty} \frac{\gamma_{tb}}{(t-1)b} = 0$, since $\gamma_{tb} = o(tb)$. Furthermore, Lemma 4 guarantees that $I(\hat{X}; Z(\hat{X}))$ exists.

It remains to see the leftmost inequality of the lemma statement. First, note that

$$
I(\overline{X}; Z^\star(\overline{X})) = I(\overline{X}; Z^\star(\overline{X})|V) = \lim_{n\to\infty} \frac{I(\overline{X}_1^n; Z^\star(\overline{X}_1^n)|V)}{n}. \tag{5}
$$

This holds since $H(V) \le \log b$ and b is a fixed constant. Furthermore,

$$
\begin{aligned}
I(\overline{X}; Z^\star(\overline{X})|V) &= \lim_{n\to\infty} \frac{I(\overline{X}_1^n; Z^\star(\overline{X}_1^n)|V)}{n} \\
&= \lim_{n\to\infty} \frac{\frac{1}{b}\sum_{j=0}^{b-1} I(\overline{X}_1^n; Z^\star(\overline{X}_1^n)|V=j)}{n} \\
&= \lim_{n\to\infty} \frac{\frac{1}{b}\sum_{j=0}^{b-1} I(\hat{X}_1^{[j]n}; Z^\star(\hat{X}_1^{[j]n}))}{n}.
\end{aligned} \tag{6}
$$

As we saw above, the limits

$$
I(\hat{X}^{[j]}; Z^\star(\hat{X}^{[j]})) = \lim_{n\to\infty} \frac{I(\hat{X}_1^{[j]n}; Z^\star(\hat{X}_1^{[j]n}))}{n}
$$

exist and equal $I(\hat{X}; Z^\star(\hat{X}))$ for all j. Therefore, since the sum over j is finite, we can swap limit and sum and conclude from Equations (5) and (6) that

$$
\begin{aligned}
I(\overline{X}; Z^\star(\overline{X})) &= I(\overline{X}; Z^\star(\overline{X})|V) \\
&= \frac{1}{b}\sum_{j=0}^{b-1} \lim_{n\to\infty} \frac{I(\hat{X}_1^{[j]n}; Z^\star(\hat{X}_1^{[j]n}))}{n} \\
&= \frac{1}{b}\sum_{j=0}^{b-1} I(\hat{X}^{[j]}; Z^\star(\hat{X}^{[j]})) \\
&= I(\hat{X}; Z^\star(\hat{X})),
\end{aligned}
$$

as desired. □

Lemma 6. *Suppose that $\hat{X}$ is a block independent process with blocklength b. Then, we have*

$$
\left| I(\hat{X}; Z^\star(\hat{X})) - \frac{I(\hat{X}_1^b; Z^\star(\hat{X}_1^b))}{b} \right| \le \alpha_b/b,
$$

where $\lim_{b\to\infty} \frac{\alpha_b}{b} = 0$.

Proof. By Item 4b, there exists a random variable W with $H(W) \leq t \cdot \alpha_b$ such that the sequences $Z^\star(\hat{X}_1^b), Z^\star(\hat{X}_{b+1}^{2b}), \dots, Z^\star(\hat{X}_{(t-1)b+1}^{tb})$ are completely determined by $Z^\star(\hat{X}_1^{tb})$ and W. Therefore,

$$\begin{aligned}
I(\hat{X}_1^{tb}; Z^\star(\hat{X}_1^{tb})) &\geq I(\hat{X}_1^{tb}; Z^\star(\hat{X}_1^b), Z^\star(\hat{X}_{b+1}^{2b}), \dots, Z^\star(\hat{X}_{(t-1)b+1}^{tb})) - t \cdot \alpha_b \\
&= \sum_{i=1}^{t} I(\hat{X}_{(i-1)b+1}^{ib}; Z^\star(\hat{X}_{(i-1)b+1}^{ib})) - t \cdot \alpha_b \\
&= \sum_{i=1}^{t} I(\hat{X}_1^b; Z^\star(\hat{X}_1^b)) - t \cdot \alpha_b \\
&= t(I(\hat{X}_1^b; Z^\star(\hat{X}_1^b)) - \alpha_b).
\end{aligned}$$

Consequently,

$$I(\hat{X}; Z^\star(\hat{X})) = \lim_{t\to\infty} \frac{I(\hat{X}_1^{tb}; Z^\star(\hat{X}_1^{tb}))}{tb} \geq \frac{I(\hat{X}_1^b; Z^\star(\hat{X}_1^b)) - \alpha_b}{b}.$$

On the other hand, by Item 3 we have

$$\frac{I(\hat{X}_1^b; Z^\star(\hat{X}_1^b))}{b} = \frac{I(\hat{X}_1^{tb}; Z^\star(\hat{X}_1^b), Z^\star(\hat{X}_{b+1}^{2b}), \dots, Z^\star(\hat{X}_{(t-1)b+1}^{tb}))}{tb} \geq \frac{I(\hat{X}_1^{tb}; Z^\star(\hat{X}_1^{tb}))}{tb}$$

for all t and b, and so $\frac{I(X_1^b; Z^\star(X_1^b))}{b} \geq I(\hat{X}; Z^\star(\hat{X}))$. □

3.4 Information capacity of admissible channels is achieved by Markov process

We use Theorem 5 to show the following.

Theorem 6. *Suppose that the channel Z is admissible. Then,*

$$\mathsf{ICap}(Z) = \lim_{m\to\infty} \mathsf{SCap}^{(m)}(Z).$$

Proof. Given $\varepsilon > 0$, let X be a stationary ergodic input process and b a large enough integer such that

$$\frac{I(X_1^b; Z^\star(X_1^b))}{b} \geq \mathsf{ICap}(Z) - \varepsilon.$$

Such a process X and integer b are guaranteed to exist by Item 2 and Theorem 5. Let $\hat{X}$ be the stationary $(b-1)$-th order Markov process satisfying

$$p_{\hat{X}_1^b}(x_1^b) = p_{X_1^b}(x_1^b).$$

We will show that

$$I(\hat{X}; Z(\hat{X})) = I(\hat{X}; Z^\star(\hat{X})) \geq \frac{I(X_1^b; Z^\star(X_1^b))}{b} - \frac{\alpha_b}{b}, \tag{7}$$

with $\lim_{b\to\infty} \frac{\alpha_b}{b} = 0$, where the first equality holds by Item 2. Taking b to be large enough, we get that $I(\hat{X}; Z(\hat{X})) \geq \mathsf{ICap}(Z) - 2\varepsilon$, and, since ε was arbitrary, the theorem statement follows.

Since $\hat{X}$ is a Markov process, we have $H(\hat{X}) \geq H(X)$. Therefore, it is enough to prove that

$$H(\hat{X}|Z^\star(\hat{X})) \leq \frac{H(X_1^b|Z^\star(X_1^b))}{b} + \frac{\alpha_b}{b}. \tag{8}$$

We have

$$
\begin{aligned}
H(\hat{X}|Z^\star(\hat{X})) &= \lim_{t\to\infty} \frac{H(\hat{X}_1^{tb}|Z^\star(\hat{X}_1^{tb}))}{tb} \\
&\leq \lim_{t\to\infty} \frac{H(\hat{X}_1^{tb}|Z^\star(\hat{X}_1^{b}),\dots,Z^\star(X_{(t-1)b+1}^{tb})) + t\cdot\alpha_b}{tb} \\
&= \lim_{t\to\infty} \frac{\sum_{i=1}^t H(\hat{X}_{(i-1)b+1}^{ib}|\hat{X}_1^{(i-1)b}, Z^\star(\hat{X}_1^{b}),\dots,Z^\star(X_{(t-1)b+1}^{tb})) + t\cdot\alpha_b}{tb} \\
&\leq \lim_{t\to\infty} \frac{\sum_{i=1}^t H(\hat{X}_{(i-1)b+1}^{ib}|Z^\star(\hat{X}_{(i-1)b+1}^{ib})) + t\cdot\alpha_b}{tb} \\
&= \lim_{t\to\infty} \frac{\sum_{i=1}^t H(\hat{X}_1^{b}|Z^\star(\hat{X}_1^{b})) + t\cdot\alpha_b}{tb} \\
&= \frac{H(X_1^b|Z^\star(X_1^b))}{b} + \frac{\alpha_b}{b},
\end{aligned}
$$

as desired. The first inequality uses Item 4b. The second equality uses the chain rule for conditional entropy. The second inequality holds since further conditioning does not increase entropy. The third equality uses the stationarity of X. □

3.5 Coding capacity equals information capacity, and existence of dense codes from stationary ergodic processes

In this section, we begin by establishing suitable convergence of the information density of block-independent processes to their information rate for admissible channels. We use this result in two ways. First, we use it to show that $\mathsf{CCap}(Z) = \mathsf{ICap}(Z)$. Second, focusing on admissible channels with binary input for simplicity, we apply this result to stationary ergodic processes (which by Theorem 5 achieve information capacity on admissible channels) to conclude that there exist "dense" codes $\mathcal{C}$ that achieve capacity on Z, where "dense" means that every short substring of $c \in \mathcal{C}$ contains a decent fraction of 1s. As discussed in [PLW22], this property is relevant for the construction of efficient capacity-achieving codes for these channels.

3.5.1 Convergence of information density for block-independent process

For two random variables X, Y, we define their information density $i_{X,Y}$ as

$$
i_{X,Y}(x,y) = \log\left(\frac{p_{XY}(x,y)}{p_X(x)\cdot p_Y(y)}\right).
$$

Note that $\mathbb{E}_{(x,y)\sim p_{X,Y}}[i_{X,Y}(x,y)] = I(X;Y)$. We show the following.

Theorem 7. *Let Z be an admissible channel with respect to $Z^\star$. Fix $\varepsilon > 0$ and let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be the sequences guaranteed by Item 4b and Item 5. Let X be a block-independent process with blocklength b such that*

$$
\max(\alpha_b/b, \beta_b/b) \leq \varepsilon^2/3.
$$

Then, there exists a constant t_0 (possibly depending on ε, b, and ε) such that for all $t \geq t_0$ we have

$$
\Pr_{(x,z)\sim X_1^{tb}, Z(X_1^{tb})}\left[\left|\frac{i_{X_1^{tb},Z(X_1^{tb})}(x,z)}{tb} - I(X;Z(X))\right| \leq \varepsilon\right] \geq 1-\varepsilon.
$$

Before we prove Theorem 7 we establish some useful lemmas. We will start by working with the $Z^\star$ channel.

Lemma 7. *Suppose that X is a block-independent process with blocklength b. Let*

$$\eta_t^\star = (Z^\star(X_1^b), Z^\star(X_{b+1}^{2b}), \dots, Z^\star(X_{(t-1)b+1}^{tb})).$$

Then,

$$i_{X_1^{tb},\eta_t^\star}(x,z) = \sum_{i=1}^{t} i_{X_1^b, Z^\star(X_1^b)}(x^{(i)}, z^{(i)}), \tag{9}$$

where $x^{(i)}, z^{(i)}$ denote the i-th blocks of x and z, respectively. In particular, $\frac{1}{tb} i_{X_1^{tb},\eta_t^\star}(X_1^{tb},\eta_t^\star)$ converges almost surely to $\frac{I(X_1^b;Z^\star(X_1^b))}{b}$ as $t\to\infty$.

Proof. Equation (9) follows from the fact that $p_{X_1^{tb},\eta_{t,b}^\star}(x,z) = \prod_{i=1}^t p_{X_1^b,Z^\star(X_1^b)}(x^{(i)},z^{(i)})$ by block-independence of X (with blocklength b). The statement about convergence then follows from the strong law of large numbers, since the random variables $i_{X_1^b,Z^\star(X_1^b)}(X_{(i-1)b+1}^{ib}, Z^\star(X_{(i-1)b+1}^{ib}))$ are i.i.d. for all $i \in [t]$ and their expectation is $I(X_1^b; Z^\star(X_1^b))$. □

For a given input X_1^{tb}, we will couple the $\eta_t^\star$ and $Z^\star(X_1^{tb})$ processes with the help of Item 5. Let Y_t be the random variable such that $\eta_t^\star \to Y_t$ and ϕ the deterministic function such that $Z^\star(X_1^{tb}) = \phi(\eta_t^\star, Y_t)$ and $\max_z \log|\phi^{-1}(z)| \le t\cdot\beta_b$, guaranteed by Item 5.

Lemma 8. *We have*

$$\begin{aligned}
&\mathbb{E}_{(x,z,y)\sim X_1^{tb},\eta_t^\star,Y_t}\left[\left|i_{X_1^{tb},\eta_t^\star}(x,z) - i_{X_1^{tb},Z^\star(X_1^{tb})}(x,\phi(z,y))\right|\right] \\
&= \mathbb{E}_{(x,z,y)\sim X_1^{tb},\eta_t^\star,Y_t}\left[\left|i_{X_1^{tb},\eta_t^\star,Y_t}(x,z,y) - i_{X_1^{tb},Z^\star(X_1^{tb})}(x,\phi(z,y))\right|\right] \\
&\le \max_z \log|\phi^{-1}(z)| \\
&\le t\cdot\beta_b.
\end{aligned}$$

The proof of this lemma relies on the following simple but useful fact about information densities, due to Dobrushin [Dob67].

Lemma 9 ([Dob67])**.** *Let $\phi : \mathcal{A}\to\mathcal{B}$ be a deterministic function with $\mathcal{A}$ and $\mathcal{B}$ finite sets, and write $M_\phi = \max_{z\in\mathcal{B}}|\phi^{-1}(z)|$ for the size of the largest preimage of ϕ. Let A be supported on $\mathcal{A}$, and define $B=\phi(A)$. Then,*

$$\mathbb{E}_{(x,a)\sim X,A}[|i_{X,A}(x,a) - i_{X,B}(x,\phi(a))|] \le \log M_\phi.$$

We are now ready to proceed with the proof of Lemma 8.

Proof of Lemma 8. The first equality follows from the fact that

$$i_{X_1^{tb},\eta_t^\star,Y_t}(x,z,y) = i_{X_1^{tb},\eta_t^\star}(x,z)$$

for all (x,z,y) in the support of $(X_1^{tb},\eta_t^\star,Y_t)$, since $X_1^{tb}\to\eta_t^\star\to\eta_t^\star,Y_t$. The inequality follows from Lemma 9 with $X = X_1^{tb}$ and $A = (\eta_t^\star, Y_t)$. The last equality follows from the hypothesis on ϕ, guaranteed by Item 5. □

We are now ready to prove Theorem 7.

Proof of Theorem 7. Since X is block-independent with blocklength b, by Lemma 6 we have

$$\left| \frac{I(X_1^b; Z^\star(X_1^b))}{b} - I(X; Z^\star(X)) \right| \leq \frac{\alpha_b}{b} \leq \varepsilon/3, \tag{10}$$

where the last inequality uses the hypothesis on b from the theorem statement. Therefore, it is enough to show that

$$\Pr_{(x,z)\sim X_1^{tb}, Z^\star(X_1^{tb})} \left[\left| \frac{i_{X_1^{tb}, Z^\star(X_1^{tb})}(x,z)}{tb} - \frac{I(X_1^b; Z^\star(X_1^b))}{b} \right| \leq 2\varepsilon/3 \right] \geq 1 - \varepsilon. \tag{11}$$

First, by the triangle inequality, we have

$$\begin{aligned}
&\Pr_{(x,z)\sim X_1^{tb}, Z^\star(X_1^{tb})} \left[\left| \frac{i_{X_1^{tb}, Z^\star(X_1^{tb})}(x,z)}{tb} - \frac{I(X_1^b; Z^\star(X_1^b))}{b} \right| \leq 2\varepsilon/3 \right] \\
&\leq \Pr_{(x,z^\star,y)\sim X_1^{tb}, \eta_t^\star, Y_t} \left[\left| \frac{i_{X_1^{tb}, \eta_t^\star}(x,z^\star)}{tb} - \frac{i_{X_1^{tb}, Z^\star(X_1^{tb})}(x,\phi(z^\star,y))}{tb} \right| \leq \varepsilon/3 \right] \\
&+ \Pr_{(x,z^\star)\sim X_1^{tb}, \eta_t^\star} \left[\left| \frac{i_{X_1^{tb}, \eta_t^\star}(x,z^\star)}{tb} - \frac{I(X_1^b; Z^\star(X_1^b))}{b} \right| \leq \varepsilon/3 \right].
\end{aligned} \tag{12}$$

We analyze the two terms in the sum separately. For the first term, combining Lemma 8 with Markov's inequality yields

$$\begin{aligned}
\Pr_{(x,z^\star,w)\sim X_1^{tb}, \eta_t^\star, W_t} \left[\left| \frac{i_{X_1^{tb}, \eta^\star}(x,z^\star)}{tb} - \frac{i_{X_1^{tb}, Z^\star(X_1^{tb})}(x,\phi(z^\star,w))}{tb} \right| > \varepsilon/3 \right] &\leq \frac{\max_z \log|\phi^{-1}(z)|}{\varepsilon tb} \\
&\leq \frac{\beta_b}{\varepsilon b} \\
&\leq \varepsilon/2,
\end{aligned} \tag{13}$$

where the last inequality holds by the hypothesis on b from the theorem statement. For the second term, by Lemma 7, for all $t \geq t_0$ with t_0 a sufficiently large constant depending on ε, b, and X, we have

$$\Pr_{(x,z^\star)\sim X_1^{tb}, \eta_t^\star} \left[\left| \frac{i_{X_1^{tb}, \eta_t^\star}(x,z^\star)}{tb} - \frac{I(X_1^b; Z^\star(X_1^b))}{b} \right| > \varepsilon/3 \right] \leq \varepsilon/2. \tag{14}$$

Combining Equation (12) with Equations (13) and (14) yields Equation (11), as desired. □

3.5.2 Capacity-achieving codes for admissible channels

Theorem 7 implies, via standard methods, that the coding capacity and information capacity of admissible channels coincide. For completeness, we discuss this in detail. Later in Section 3.5.3 we combine Theorems 5 and 7 to show the existence of "dense" capacity-achieving codes suitable for bootstrapping efficient constructions.

We begin by relying on the following well-known theorem that formalizes the guarantees of MAP decoding.

Lemma 10 ([PW24, Theorem 18.5], adapted)**.** *Fix an input random variable X supported on Σ_{in}^n and a channel Z with input alphabet Σ_{in}. Then, for any $\tau > 0$ there exists a code $\mathcal{C}$ with blocklength n, size M, and average decoding error probability ε satisfying*

$$\varepsilon \leq \Pr_{(x,z)\sim X,Z(X)}[i_{X,Z(X)}(x,z) \leq \log M + \tau] + 2^{-\tau}.$$

The code guaranteed by Lemma 10 is obtained by sampling M codewords i.i.d. according to X. We briefly discuss how Lemma 10 can be combined with Theorem 7 to obtain codes with arbitrarily small decoding error probability and the desired rate for an arbitrary admissible channel.

Corollary 1. *Let Z be an admissible channel. Fix an arbitrary $\varepsilon > 0$. Then, there exists a constant $b(\varepsilon)$ such that for any $b \geq b(\varepsilon)$ and any block-independent input process X with blocklength b there exists a constant $t(\varepsilon, b, X)$ such that for any $t \geq t(\varepsilon, b, X)$ there exists an $(n = tb, R, \varepsilon)$-code for Z with $R \geq I(X; Z(X)) - \varepsilon$.*

Proof. Fix $\varepsilon > 0$ and let $b(\varepsilon), t(\varepsilon)$ be the constants guaranteed by Theorem 7. Consider any block-independent input process X with blocklength $b \geq b(\varepsilon)$ and number of blocks $t \geq t(\varepsilon)$. Set $n = tb$, $M = 2^{n(I(X;Z(X))-3\varepsilon)}$, and $\tau = \varepsilon n$. By Lemma 10, there exists an $(n = tb, R, \lambda)$-code for Z with

$$\begin{aligned}
\lambda &\leq \Pr_{(x,z)\sim X_1^n,Z(X_1^n)}[i_{X_1^n,Z(X_1^n)}(x,z) \leq \log M + \tau] + 2^{-\tau} \\
&= \Pr_{(x,z)\sim X_1^n,Z(X_1^n)}\left[\frac{i_{X_1^n,Z(X_1^n)}(x,z)}{n} \leq I(X;Z(X)) - 2\varepsilon\right] + 2^{-\varepsilon n} \\
&\leq \varepsilon + 2^{-\varepsilon n},
\end{aligned}$$

where the last inequality follows from Theorem 7. Now, we may set $b(\varepsilon)$ and $t(\varepsilon)$ large enough as a function of ε so that $2^{-\varepsilon n} \leq \varepsilon$, and so $\lambda \leq 2\varepsilon$. □

Extension to all blocklengths. We now argue how Corollary 1 can be extended to all blocklengths, which shows that for any admissible channel Z the coding capacity equals the information capacity.

Theorem 8. *Suppose that Z is an admissible channel. Then, $\mathsf{ICap}(Z) = \mathsf{CCap}(Z)$.*

Proof. We begin by considering the following "trimming" version of Z, which we denote by Z'. On input x, this channel first sends x through Z to obtain output $Z(x)$, and then trims the maximal substring of 0s at the end of $Z(x)$. Since Z' is a degraded version of Z, for any input process X we have $I(X; Z'(X)) \leq I(X; Z(X))$. Also, $Z(X_1^n)$ is completely determined by $Z'(X_1^n)$ and the length L of the run of 0s trimmed from the end of $Z(X_1^n)$ to obtain $Z'(X_1^n)$. Since $H(L) \leq \log n$, this means that $I(X_1^n; Z'(X_1^n)) \geq I(X_1^n; Z(X_1^n)) - \frac{\log n}{n} \to I(X; Z(X))$ as $n \to \infty$. We conclude that Z' is also admissible with respect to $Z^\star$ and $\mathsf{ICap}(Z') = \mathsf{ICap}(Z)$.

Fix an arbitrary $\varepsilon > 0$ and a block-independent process X with block length $b = b(\varepsilon)$ such that $I(X; Z'(X)) \geq \mathsf{ICap}(Z') - \varepsilon$ (the existence of such a process follows, e.g., from the discussion after Equation (1)). By Corollary 1 applied to Z' and X, there exists a family of codes $\{\mathcal{C}_{tb}\}_{t\in\mathbb{N}}$ such that $\mathcal{C}_{tb}$ is a $(tb, R_{tb}, \varepsilon_{tb})$-code with $R_{tb} \geq I(X; Z'(X)) - \varepsilon$ and $\varepsilon_{tb} \leq \varepsilon$ for all large enough $t \geq t(\varepsilon, X)$. We extend this family to all blocklengths n as follows. For $n < t(\varepsilon)b(\varepsilon)$, we set $\mathcal{C}_n = \Sigma_{\mathsf{in}}^n$. For $tb < n < (t+1)b$ with $t \geq t(\varepsilon, X)$, we construct an (n, R_n, ε_n)-code $\mathcal{C}_n$ by appending a run of $n - tb < b$ 0s to codewords of $\mathcal{C}_{tb}$. First, note that since Z' trims all 0s at the end of a codeword, the decoding error probability of $\mathcal{C}_n$ equals that of $\mathcal{C}_{tb}$, and so $\varepsilon_n = \varepsilon_{tb} \leq \varepsilon$ for all large enough n

depending only on ε. Second, $R_n = R_{tb} \cdot \frac{tb}{(t+1)b} \geq R_{tb} - \varepsilon$ for all large enough n (again, depending on ε and X).

Now, for each $k \in \mathbb{N}$ define $\varepsilon^{(k)} = 1/k$ and the associated family $\{\mathcal{C}_n^{(k)}\}_{n\in\mathbb{N}}$ of $(n, R_n^{(k)}, \varepsilon_n^{(k)})$-codes guaranteed by the last paragraph, where $R_n^{(k)} \geq \mathsf{ICap}(Z') - 2\varepsilon^{(k)}$ and $\varepsilon_n^{(k)} \leq \varepsilon^{(k)}$ for all large enough $n \geq n(k)$ (note that the choice of input process X is fixed for each $k \in \mathbb{N}$, so $n(k)$ really only depends on k). Taking codes from the k-th family for every blocklength n between $n(k)$ and $n(k+1)$, we get a family of codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ where each $\mathcal{C}_n$ is an (n, R_n, ε_n)-code for Z' with $R_n \to I(X; Z'(X))$ and $\varepsilon_n \to 0$ as $n \to \infty$. Furthermore, since Z' is a degraded version of Z, we also have that each $\mathcal{C}_n$ is an (n, R_n, ε_n)-code for Z. Since $\mathsf{ICap}(Z) = \mathsf{ICap}(Z')$, we conclude that $\mathsf{CCap}(Z) = \mathsf{CCap}(Z') = \mathsf{ICap}(Z') = \mathsf{ICap}(Z)$. □

3.5.3 Dense capacity-achieving codes for admissible channels

In the previous section we showed that the coding capacity and information capacity of an admissible channel are the same. However, this alone is not sufficient if we wish to obtain *efficiently encodable and decodable* capacity-achieving codes for an admissible channel. In this section, following ideas of [PLW22], we combine the fact that capacity is achieved by stationary ergodic processes (Theorem 5) with Theorem 7 to show the existence of capacity-achieving codes with density properties for admissible channels.[2] useful for constructing efficient capacity-achieving codes. For simplicity we will focus on admissible channels with binary input alphabet, although our discussion generalizes further.

The following lemma was proved in [PLW22].

Lemma 11 ([PLW22, Proposition 3.4], adapted)**.** *Let X be a stationary ergodic process supported on $\{0,1\}$ with $\Pr[X_1 = 1] \in (0,1)$. Then, for any $\zeta > 0$ there exists $\gamma \in (0, 1/2)$ and an integer $n_0 > 0$ such that the following holds for all $n \geq n_0$. With probability at least 0.99 over the sampling of $x \sim X_1^n$, we have $\gamma\zeta n \leq w(x_i^{i+\zeta n}) \leq (1-\gamma)\zeta n$ for all $i \in [(1-\zeta)n]$, where $w(\cdot)$ denotes the Hamming weight.*

Let Z be an arbitrary admissible channel with binary input alphabet. Let X be a stationary ergodic process such that $I(X; Z(X)) > 0$. Then, it must be the case that $\Pr[X_1 = 1] \in (0,1)$, and so Lemma 11 applies to X with some constants $\zeta > 0$, $\gamma \in (0, 1/2)$, and n_0.

It will be slightly easier to work with a block-independent process. Fix $\varepsilon > 0$. From the proof of Theorem 5 in Section 3.3, we know that there is a stationary ergodic process $\overline{X}$ such that $I(\overline{X}; Z(\overline{X})) \geq \mathsf{ICap}(Z) - \varepsilon$, and moreover $\overline{X}$ is created by choosing an appropriate block-independent process $\hat{X}$ with blocklength $b(\varepsilon)$, then choosing a uniformly random starting point in the first block of $\hat{X}$, and starting $\hat{X}$ from that point. Since $\overline{X}$ is obtained from $\hat{X}$ by trimming at most b bits from the beginning of $\hat{X}$, we conclude that $\hat{X}$ also satisfies the properties laid out in Lemma 11 with possibly a slightly smaller γ and slightly larger n_0. Recalling that the code $\mathcal{C}$ guaranteed by Corollary 1 applied to Z and $\hat{X}$ is obtained by sampling codewords i.i.d. according to $\hat{X}_1^n$ yields Theorem 4. This is because with high probability more than, say, a 0.9-fraction of codewords $c \in \mathcal{C}$ will satisfy $\gamma\zeta n \leq w(c_i^{i+\zeta n}) \leq (1-\gamma)\zeta n$ for all $i \in [(1-\zeta)n]$, and throwing away all codewords of $\mathcal{C}$ that do not satisfy this property will not affect the asymptotic rate.

[2] Pernice, Li, and Wootters [PLW22] focused on channels with i.i.d. deletions and replications. The analog of Theorem 5 for these channels was already shown in [Dob67].

4 Some special cases of our capacity theorems

4.1 The Mao-Diggavi-Kannan ISI model

Consider the *ℓ-ISI-synchronization channel* Z from [MDK18], for an arbitrary fixed integer $\ell \geq 0$. This channel replaces the i-th input bit x_i by a string $y_i \in \{0,1\}^*$ with probability

$$p(y_i|x_i, x_{i-1}, \dots, x_{i-\ell}).$$

For simplicity, we focus on the case where $p(\cdot|x_i, x_{i-1}, \dots, x_{i-\ell})$ is supported on $\{0,1\}^{\leq a} = \bigcup_{j=0}^{a}\{0,1\}^j$ for some integer $a \geq 1$ and any choice of $x_i, x_{i-1}, \dots, x_{i-\ell}$, although our argument below generalizes further. We show that our capacity theorems apply to this channel, and so they generalize the corresponding results of [MDK18].

Consider the special channel $Z^\star$ that behaves like Z, except that it does not corrupt the first ℓ input bits and separately outputs the last ℓ input bits. We show that Z is admissible with respect to $Z^\star$.

- **Item 1:** Since the output associated to the i-th input bit has length at most a fixed constant a, we have $H(Z^\star(X_1^n)) \leq n \cdot (a+1)$.

- **Item 2:** Fix an arbitrary input process X. Note that

  $$X_1^n \to Z^\star(X_1^n) \to Z(X_1^n).$$

  Let W denote, for each $j \in [\ell]$, the string v_j that X_j was replaced by in $Z(X_1^n)$. Then,

  $$X_1^n \to Z(X_1^n), W, X_{n-\ell+1}^n \to Z^\star(X_1^n).$$

  Note that there are at most 2^{a+1} choices for each v_j. Therefore, $H(W, X_{n-\ell+1}^n) \leq \ell + \ell \cdot (a+1)$, and so

  $$I(X_1^n; Z^\star(X_1^n)) - \ell(a+2) \leq I(X_1^n; Z(X_1^n)) \leq I(X_1^n; Z^\star(X_1^n)).$$

  As a result,

  $$\lim_{n\to\infty} \frac{|I(X_1^n; Z(X_1^n)) - I(X_1^n; Z^\star(X_1^n))|}{n} \leq \lim_{n\to\infty} \frac{\ell(a+2)}{n} = 0.$$

- **Item 3:** Note that $Z^\star(X_1^m)$ does not corrupt the first ℓ bits of X_1^m, and furthermore it outputs $X_{m-\ell+1}^m$. Otherwise, it behaves exactly like Z. Analogously, $Z^\star(X_{m+1}^n)$ does not corrupt the first ℓ bits of X_{m+1}^n and it outputs $X_{n-\ell+1}^n$. Therefore, from $Z^\star(X_1^m), Z^\star(X_{m+1}^n)$ we have the necessary information to apply the correct errors to the first ℓ bits X_{m+1}^n, and we also know the last ℓ input bits $X_{n-\ell+1}^n$. This means that we fully determine $Z^\star(X_1^n)$.

- **Item 4:**

  1. Fix integers τ and $n \geq \tau$, and an input process X. Let N_1 denote the number of output bits corresponding to X_1^τ in $Z^\star(X_1^n)$. Also, let W_{pre} include for each $j \in \{\tau+1, \dots, \tau+\ell\}$ the string v_j that X_j was replaced by in $Z^\star(X_1^n)$. Then, $Z^\star(X_1^n), N, W_{\mathsf{pre}}, X_{\tau-\ell+1}^\tau, X_{\tau+1}^{\tau+\ell}$ completely determine $Z^\star(X_1^\tau), Z^\star(X_{\tau+1}^n)$, and

     $$H(N, W_{\mathsf{pre}}, X_{\tau+1}^{\tau+\ell}, X_{n-\ell+1}^n) \leq \log(\tau \cdot 2^{a+1}) + \ell \cdot (a+1) + 2\ell.$$

     Therefore, the prefix-partitioning half of Item 4a holds with $\gamma_m = \log\tau + (a+1) + \ell(a+3)$. An analogous argument establishes the suffix-partitioning property with the same γ_m.

2. Fix a blocklength b, number of blocks t, and an input process X. For each $i \in [t]$, let N_i denote the number of output bits corresponding to $X_{(i-1)b+1}^{ib}$. Also, let W_i include, for each $j \in \{(i-1)b+1, (i-1)b+\ell\}$ the string v_j that X_j is replaced by. Let

$$W = (N_i, W_i, X_{ib-\ell+1}^{ib})_{i\in[t]}.$$

Then, $Z^\star(X_1^{tb}), W$ completely determines the sequence $Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb})$, and

$$H(W) \le \sum_{i=1}^{t} (\log(b \cdot 2^{a+1}) + \ell \cdot (a+1) + \ell) = t \cdot (\log(b \cdot 2^{a+1}) + \ell \cdot (a+1) + \ell).$$

Therefore, Item 4b holds with $\alpha_m = \log(m \cdot 2^{a+1}) + \ell \cdot (a+1) + \ell = o(m)$.

- **Item 5:** Fix a blocklength b and a number of blocks t. Consider the random variable $Y = (Y_i)_{i\in[t]}$, where each Y_i includes, for each $j \in \{(i-1)b+1, \dots, (i-1)b+\ell\}$, the string v_j that X_j should be replaced by in $Z^\star(X_1^n)$. Then $X_1^{tb} \to Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb}) \to Y$, since the distribution of Y_i is completely determined by $X_{(i-1)b-\ell+1}^{(i-1)b}$ and $X_{(i-1)b+1}^{(i-1)b+\ell}$, where the former is revealed by the $(i-1)$-th output block $Z^\star(X_{(i-2)b+1}^{(i-1)b})$ and the latter is revealed by the i-th output block $Z^\star(X_{(i-1)b+1}^{ib})$. Moreover, we have $Z^\star(X_1^{tb}) = \phi(Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb}), Y)$ for the deterministic function ϕ that discards the final t bits $X_{ib-\ell+1}^{ib}$ from each output block $Z^\star(X_{(i-1)b+1}^{ib})$ with $i < t$, applies the corruptions dictated by Y to $Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb})$, and then concatenates these blocks.

  It remains to upper bound $\max_z \log|\phi^{-1}(z)|$ appropriately. First, note that there are at most $\binom{t(b2^{a+1}+1)}{t}$ ways of splitting z into t blocks of length at most $b2^{a+1}$ each. Second, for each block there are at most $2^{(a+1)\cdot\ell}$ choices for the first t input bits and the corresponding strings in $\{0,1\}^{\le a}$ that they were replaced by at the output, and there are at most 2^ℓ possibilities for the last t input bits. Putting these observations together implies that

$$\max_z |\phi^{-1}(z)| \le \binom{t(b2^{a+1}+1)}{t} \cdot 2^{(a+1)\cdot\ell} \cdot 2^\ell \le 2^{t(b2^{a+1}+1)h\left(\frac{1}{b2^{a+1}+1}\right)} \cdot 2^{(a+1)\ell},$$

  where we recall that h is the binary entropy function. Therefore,

$$\max_z \log|\phi^{-1}(z)| \le t(b2^{a+1}+1)h\left(\frac{1}{b2^{a+1}+1}\right) + (a+1)\ell,$$

  and so Item 5 holds with $\gamma_m = (m2^{a+1}+1)h\left(\frac{1}{m2^{a+1}+1}\right) + (a+1)\ell = o(m)$.

4.2 Multi-trace channels with input-correlated synchronization errors

We argue how our capacity theorems above apply to a wide class of *multi-trace* input-correlated synchronization channels. Fix an integer $\ell \ge 1$ (the number of traces), and consider the multi-trace channel Z given by

$$Z(X_1^n) = (Z_1(X_1^n), \dots, Z_\ell(X_1^n)),$$

where the Z_i's are possibly distinct input-correlated synchronization channels, and each $Z_i(X_1^n)$ is an independent trace of X_1^n (i.e., the $Z_i(X_1^n)$'s are conditionally independent given X_1^n).

Theorem 9. *Suppose that each Z_i is admissible with respect to $Z_i^\star$, and that*

$$I(X;Z_1(X),\dots,Z_\ell(X)) = I(X;Z_1^\star(X),\dots,Z_\ell^\star(X)) \tag{15}$$

for all input processes X. Then, the ℓ-trace channel Z is admissible with respect to the ℓ-trace channel $Z^\star$ given by

$$Z^\star(X_1^n) = (Z_1^\star(X_1^n),\dots,Z_\ell^\star(X_1^n)).$$

In particular, we have $\mathsf{ICap}(Z) = \mathsf{SCap}(Z) = \mathsf{CCap}(Z)$.

The assumption in Equation (15) appears stronger than simply requiring that each Z_i satisfy Item 2 with respect to $Z_i^\star$. Nevertheless, it still seems reasonable. Concretely, it is natural (as in all of our applications, for example) that $Z_i^\star(X_1^n)$ is completely determined by $Z_i(X_1^n)$ and some additional side information W_i satisfying $H(W_i) = o(n)$. In this case, we get that $Z^\star(X_1^n)$ is completely determined by $Z(X_1^n)$ and the side information $W = (W_1,\dots,W_\ell)$, which satisfies $H(W) \leq \sum_{i=1}^\ell H(W_i) = \ell \cdot o(n) = o(n)$, since the number of traces ℓ is constant. This means that

$$\frac{|I(X_1^n;Z(X_1^n)) - I(X_1^n;Z^\star(X_1^n))|}{n} \leq \frac{H(W)}{n} \to 0$$

as $n\to\infty$.

Proof of Theorem 9. We verify that Z is admissible with respect to $Z^\star$. Item 2 is already guaranteed by Equation (15), and so we focus on showing the other properties.

- **Item 1:** Fix an arbitrary input process X and integer $n \geq 1$. For each i we know that $H(Z_i^\star(X_1^n)) \leq c_i n$ for some constant $c_i > 0$. Let $c^\star = \max_{i\in[\ell]} c_i$. Then, $H(Z^\star(X_1^n)) = H((Z_i^\star(X_1^n)_{i\in[\ell]}) \leq \sum_{i=1}^\ell H(Z_i^\star(X_1^n)) \leq \ell\cdot c^\star n$, and so Item 1 holds with constant $c = \ell\cdot c^\star$.
- **Item 3:** For an arbitrary input process X, indices $1 \leq m \leq n$ and each $i\in[\ell]$, we have that $X_1^n \to Z_i^\star(X_1^m), Z_i^\star(X_{m+1}^n) \to Z_i^\star(X_1^n)$. In particular, this means that

  $$X_1^n \to (Z_i^\star(X_1^m))_{i\in[\ell]}, (Z_i^\star(X_{m+1}^n)_{i\in[\ell]} = Z^\star(X_1^m), Z^\star(X_{m+1}^n) \to (Z_i^\star(X_1^n))_{i\in[\ell]} = Z^\star(X_1^n).$$

- **Item 4:**
  - Item 4a: Fix any integers $\tau \geq 1$ and $n \geq \tau$. For each $i\in[\ell]$ let $\gamma_m^{(i)} = o(m)$ and $W_{\mathsf{pre},i}$ and $W_{\mathsf{suf},i}$ be the sequence and random variables guaranteed by Item 4a applied to Z_i. Consider $\gamma_m = \sum_{i=1}^\ell \gamma_m^{(i)}$ and $W_{\mathsf{pre}} = (W_{\mathsf{pre},i})_{i\in[\ell]}$ and $W_{\mathsf{suf}} = (W_{\mathsf{suf},i})_{i\in[\ell]}$. Then, $\gamma_m = o(m)$ and $\sum_{m\in\mathbb{N}} \gamma_m/m^2$ converges, and $H(W_{\mathsf{pre}}), H(W_{\mathsf{suf}}) \leq \gamma_n$. Furthermore, $(Z^\star(X_1^n), W_{\mathsf{pre}}) = ((Z_i^\star(X_1^n))_{i\in[\ell]}, W_{\mathsf{pre}})$ completely determines $(Z^\star(X_1^\tau), Z^\star(X_{\tau+1}^n)) = (Z_i^\star(X_1^\tau), Z_i^\star(X_{\tau+1}^n))$. The reasoning for W_{suf} is analogous.
  - Item 4b: For each $i\in[\ell]$ let $\alpha_m^{(i)} = o(m)$ and W_i be the sequence and random variable, respectively, guaranteed by Item 4b applied to Z_i. Consider $W = (W_1,\dots,W_\ell)$ and $\alpha_m = \sum_{i=1}^\ell \alpha_m^{(i)}$. Then, $\alpha_m = o(m)$ and $H(W) \leq \sum_{i=1}^\ell H(W_i) \leq t\alpha_b$, and $(Z^\star(X_1^{tb}), W) = ((Z_i^\star(X_1^{tb}), W_i))_{i\in[\ell]}$ completely determines $Z^\star(X_1^b),\dots,Z^\star(X_{(t-1)b+1}^{tb}) = (Z_i^\star(X_1^b),\dots,Z_i^\star(X_{(t-1)b+1}^{tb}))_{i\in[\ell]}$.

- **Item 5:** For each $i \in [\ell]$, let $\beta_m^{(i)} = o(m)$, Y_i, and ϕ_i be the sequence, random variable, and deterministic function guaranteed by Item 5 applied to Z_i. Set $\beta_m = \sum_{i=1}^{\ell} \beta_m^{(i)} = o(m)$, $Y = (Y_i)_{i\in[\ell]}$, and

$$\phi(Z^\star(X_1^b), \ldots, Z^\star(X_{(t-1)b+1}^{tb}), Y) = (\phi_i(Z_i^\star(X_1^b), \ldots, Z_i^\star(X_{(t-1)b+1}^{tb}), Y_i))_{i\in[\ell]}.$$

Then,

$$X_1^{tb} \to Z^\star(X_1^b), \ldots, Z^\star(X_{(t-1)b+1}^{tb}) = (Z_i^\star(X_1^b), \ldots, Z_i^\star(X_{(t-1)b+1}^{tb}))_{i\in[\ell]} \to (Y_i)_{i\in[\ell]} = Y$$

and

$$\max_z \log|\phi^{-1}(z)| \le \max_z \log\left(\prod_{i=1}^{\ell} |\phi_i^{-1}(z_i)|\right) \le \sum_{i=1}^{\ell} \max_{z_i} \log|\phi_i^{-1}(z_i)| \le t\sum_{i=1}^{\ell} \beta_b^{(i)} = t\beta_b. \qquad \square$$

4.3 Capacity theorems for trimming synchronization channels

In this section, we argue that our capacity theorems are robust to additional "trimming" of channel outputs. This property is relevant for the marker-based construction of efficient capacity-achieving codes based on our capacity theorems.

For concreteness, let Z be an arbitrary channel with binary input and output alphabets. We consider the "0-trimming" version of Z, denoted by Z_0, which on input x first sends it through Z to obtain output $Z(x)$, and then trims the runs of 0s at the beginning and end of $Z(x)$.

Lemma 12. *Let Z be admissible with respect to $Z^\star$. Then, Z_0 is also admissible with respect to $Z^\star$. In particular, the capacities of Z and Z_0 are all equal.*

Proof. It suffices to show that $I(X; Z_0(X)) = I(X; Z(X))$ for any input process X. Fix an arbitrary integer $n > 0$. First, since $X_1^n \to Z(X_1^n) \to Z_0(X_1^n)$, we have that $I(X_1^n; Z_0(X_1^n)) \le I(X_1^n; Z(X_1^n))$. On the other hand, if L_0 and L_1 denote the number of 0s trimmed by Z_0 from the beginning and end of $Z(X_1^n)$, we have that $X_1^n \to Z_0(X_1^n), L_0, L_1 \to Z(X_1^n)$, and $H(L_0, L_1) \le 2\log(n+1)$. Therefore,

$$\lim_{n\to\infty} \frac{|I(X_1^n; Z_0(X_1^n)) - I(X_1^n; Z(X_1^n))|}{n} \le \lim_{n\to\infty} \frac{2\log(n+1)}{n} = 0,$$

which implies the desired result. $\square$

4.4 Channels with runlength-dependent deletions

In this section, we apply our framework to binary channels Z with runlength-dependent deletions, in the sense that Z deletes each bit in a run of length ℓ independently with some probability $d(\ell)$ (and so we may see d as a function $d : \mathbb{N} \to [0,1]$). Consider the special $Z^\star$ that on input x behaves exactly like Z except that it does not apply deletions to the first and last runs of x, and additionally reveals the lengths of these runs. We show that Z is admissible with respect to $Z^\star$.

- **Item 1:** Fix an arbitrary input process X and integer $n \ge 1$. Since $Z^\star$ only applies deletions, we have $H(Z^\star(X_1^n)) \le n + 2\log n \le 3n$.

- **Item 2:** Fix an arbitrary input process X. Note that

$$Z^\star(X_1^n) \to Z(X_1^n).$$

Let L_1, B_1 (resp. L_2, B_2) denote the number of bits deleted by Z from the first (resp. last) input run and the bit value of this run. Let also V denote whether the first and last input runs are distinct. Then,

$$Z(X_1^n), L_1, B_1, L_2, B_2, V \to Z^\star(X_1^n).$$

Therefore, since $H(L_1, B_1, L_2, B_2, V) \leq 1 + 2(\log(n+1) + 1)$, we have

$$I(X_1^n; Z^\star(X_1^n)) - (1 + 2(\log(n+1) + 1)) \leq I(X_1^n; Z(X_1^n)) \leq I(X_1^n; Z^\star(X_1^n)),$$

and so

$$\lim_{n\to\infty} \frac{|I(X_1^n; Z(X_1^n)) - I(X_1^n; Z^\star(X_1^n))|}{n} \leq \lim_{n\to\infty} \frac{1 + 2(\log(n+1) + 1)}{n} = 0.$$

This implies that $I(X; Z(X)) = I(X; Z^\star(X))$.

- **Item 3:** Fix arbitrary integers $1 \leq m \leq n$ and an input process X. Note that $Z^\star(X_1^m)$ does not apply deletions to the first and last runs of X_1^m and also reveals the lengths of these runs, and likewise for $Z^\star(X_{m+1}^n)$. To all other runs of X_1^m and X_{n+1}^m these channels apply the same deletion rate as $Z^\star(X_1^n)$, because these runs are not broken up by the partitioning of X_1^n into X_1^m and X_{m+1}^n. Furthermore, $Z^\star(X_1^m), Z^\star(X_{m+1}^n)$ reveal the lengths of the first and last runs of X_1^n and do not apply deletions to these runs. Therefore, it is enough to argue that knowing $Z^\star(X_1^m), Z^\star(X_{m+1}^n)$ allows us to apply the correct deletion rates to the last run of X_1^m and first run of X_{m+1}^n, which may actually be part of the same run of X_1^n.

  In the special case where the last run of X_1^m is also its first run (i.e., when $X_1^m = b^m$ for some $b \in \{0,1\}$) then we apply no deletions to it, nor to the first run of X_{m+1}^n in case it matches the bit value of X_1^m. In this case, the length of the first run of X_1^n, which is part of the output of $Z^\star(X_1^n)$ can be obtained from the lengths of the first runs of X_1^m and X_{m+1}^n, which are revealed by $Z^\star(X_1^m), Z^\star(X_{m+1}^n)$. An analogous argument holds for the special case where the last run of X_{m+1}^n is also its first run (i.e., when $X_{m+1}^n = b^{n-m}$ for some $b \in \{0,1\}$).

  In all other cases, since $Z^\star(X_1^m)$ reveals the length of the last run of X_1^m and $Z^\star(X_{m+1}^n)$ reveals the length of the first run of X_{m+1}^n, we know the length of the corresponding run(s) of X_1^n and so we also know the deletion rate that must be applied. Moreover, since no deletions were applied to the last run of X_1^m and the first run of X_{m+1}^n by $Z^\star(X_1^m)$ and $Z^\star(X_{m+1}^n)$, respectively, we can perfectly emulate the behavior of $Z^\star(X_1^n)$ on the corresponding runs.

- **Item 4:**

  1. Fix integers τ and $n \geq \tau$ and an input process X. Let N denote the number of bits deleted by $Z^\star$ from X_1^τ. Furthermore, let L_1, B_1 (resp. L_2, B_2) denote the length of the last run of X_1^τ (resp. first run of $X_{\tau+1}^n$) and the bit value of this run, respectively, and let N_1 (resp. N_2) denote the number of bit deleted from the last run of X_1^τ (resp. first run of $X_{\tau+1}^n$). Set $W_{\mathsf{pre}} = (N, L_1, B_1, N_1, L_2, B_2, N_2)$. Then, from $Z^\star(X_1^n), W_{\mathsf{pre}}$ we can exactly locate the output bits corresponding to the last run of X_1^τ and to the first run of $X_{\tau+1}^n$, and restore them to their original lengths. Therefore, $Z^\star(X_1^n), W_{\mathsf{pre}}$ completely determine $Z^\star(X_1^\tau), Z^\star(X_{\tau+1}^n)$, and $H(W_{\mathsf{pre}}) \leq \log(\tau+1) + 4\log(n+1) + 2$, and so Item 4a holds with $\gamma_m = \log(\tau+1) + 4\log(m+1) + 2$, which satisfies the required properties.

     An analogous argument establishes the suffix-partitioning property with the same γ_m.

2. Fix a blocklength b, number of blocks t, and an input process X. For each $i \in [t]$, let N_i denote the number of bits deleted from $X_{(i-1)b+1}^{ib}$ by $Z^\star(X_1^n)$, $L_{1,i}, N_{1,i}, B_{1,i}$ (resp. $L_{2,i}, N_{2,i}, B_{2,i}$) denote the length of the first (resp. last) run of $X_{(i-1)b+1}^{ib}$, the number of bits deleted from this run by $Z^\star(X_1^n)$, and the bit value of this run, respectively, and V_i denote whether the first and last runs of $X_{(i-1)b+1}^{ib}$ are distinct runs. Note that

$$H(L_{1,i}, N_{1,i}, B_{1,i}, L_{2,i}, N_{2,i}, B_{2,i}, V_i) \leq 4\log(b+1) + 3$$

for every $i \in [t]$. Let $W = (L_{1,i}, N_{1,i}, B_{1,i}, L_{2,i}, N_{2,i}, B_{2,i}, V_i)_{i\in[t]}$. Then, using a similar argument to previous items, we see that $Z^\star(X_1^{tb}), W$ completely determines the sequence $Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb})$, and

$$H(W) \leq \sum_{i=1}^{t}(4\log(b+1)+3) = t \cdot (4\log(b+1)+3).$$

Therefore, Item 4b holds with $\alpha_m = 4\log(m+1) + 3 = o(m)$.

- **Item 5:** Fix a blocklength b and a number of blocks t. Consider the random variable $Y = (L_{1,i}, L_{2,i})_{i\in[t]}$ where $L_{1,i}, L_{2,i}$ denote the number of bits to be deleted from the first and last runs of $Z^\star(X_{(i-1)b+1}^{ib})$ so that we can transform $Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb})$ into $Z^\star(X_1^{tb})$. Then $X_1^{tb} \to Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb}) \to Y$, since the distribution of Y is completely determined by the lengths of the first and last runs of each block $Z^\star(X_{(i-1)b+1}^{ib})$. Moreover, we obtain $Z^\star(X_1^{tb}) = \phi(Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb}), Y)$ for the deterministic function ϕ that applies the deletions dictated by Y to the first and last runs of $Z^\star(X_1^b), \dots, Z^\star(X_{(t-1)b+1}^{tb})$, concatenates these blocks, and uses the lengths of the first and last runs of each block $X_{(i-1)b+1}^{ib}$ revealed by $Z^\star(X_{(i-1)b+1}^{ib})$ to compute the lengths of the first and last runs of X_1^n. Together, these form $Z^\star(X_1^{tb})$.

  It remains to upper bound $\max_z \log|\phi^{-1}(z)|$ appropriately. First, note that there are at most $\binom{t(b+1)}{t}$ ways of splitting z into t blocks of length at most b each. Second, for each of the t blocks of length at most b, there are at most $(b+1)^2$ possibilities for the number of bits that were deleted from the first and last input runs of each block, $(b+1)^2$ possibilities for the lengths of these runs, 2^2 possibilities for the bit values of these runs, and 2 possibilities for whether these two runs are distinct runs or not. Putting these observations together implies that

$$\max_z |\phi^{-1}(z)| \leq \binom{t(b+1)}{t} \cdot (b+1)^{4t} \cdot 2^{3t} \leq 2^{t(b+1)h\left(\frac{1}{b+1}\right)} \cdot (b+1)^{2t} \cdot 2^{3t},$$

  where we have used the standard inequality $\binom{n}{k} \leq 2^{nh(k/n)}$, with $h(p) = -p\log p - (1-p)\log(1-p)$ the binary entropy function. Therefore,

$$\begin{aligned}\max_z \log|\phi^{-1}(z)| &\leq t(b+1)h\left(\frac{1}{b+1}\right) + 4t\log(b+1) + 3t \\ &= t \cdot \left((b+1)h\left(\frac{1}{b+1}\right) + 4\log(b+1) + 3\right),\end{aligned}$$

  and so Item 5 holds with $\beta_m = (m+1)h\left(\frac{1}{m+1}\right) + 4\log(m+1) + 3 = o(m)$.

Combined with Lemma 12, the above implies that Theorem 4 applies to the 0-trimming version of any runlength-dependent deletion channel. Formally, we get the following corollary.

Corollary 2. *Let Z be a 0-trimming runlength-dependent deletion channel. Then, for any $\varepsilon, \zeta > 0$ and any $b \geq b(\varepsilon, \zeta)$ and $t \geq t(b, \varepsilon, \zeta)$, where $b(\varepsilon, \zeta)$ depends only on ε and ζ and $t(b, \varepsilon, \zeta)$ depends only on b, ε, and ζ, there exists a code $\mathcal{C}$ with blocklength $n = tb$, rate $R \geq \mathsf{ICap}(Z) - \varepsilon$, decoding error probability at most ε, and a constant $\gamma \in (0, 1/2)$ such that for all codewords $c \in \mathcal{C}$ we have $\gamma\zeta n \leq w(c_i^{i+\zeta n}) \leq (1-\gamma)\zeta n$ for all $i \in [(1-\zeta)n]$.*

5 Efficient capacity-achieving codes for channels with runlength-dependent deletions

In Corollary 2, we showed that there is a code that achieves capacity on the 0-trimming runlength-dependent channel. Furthermore, we showed that this code has a large density of 1s in every "not too short" interval.

In this section, we will closely follow the arguments of Pernice, Li, and Wootters [PLW22] who showed how one can transform any code for the binary deletion channel (and, in fact more i.i.d. channels) into an explicit and efficient code with a negligible loss in the rate. With some needed modifications, we will show how to transform (non-explicit and non-efficient) capacity-achieving codes for a smaller class of runlength-dependent channels into explicit and efficient capacity achieving codes.

We start by defining the class of runlength-dependent channels for which we aim to construct efficient codes.

Definition 11. *Let* $\text{BDC-R-L-Bounded}(d, \mu, M)$ *be a runlength-dependent channel defined with a monotonically increasing function $d(\ell) : \mathbb{N} \to [0, 1]$, number $\mu \in (0, 1)$, and an integer M such that $d(\ell) = d(M) < 1 - \mu$ for all $\ell \geq M$.*[3]

The theorem we will prove in this section is as follows.

Theorem 10. *Let* $\text{BDC-R-L-Bounded}(d, \mu, M)$ *be a runlength-dependent channel that complies with Definition 11 where μ and M are constants.*[4] *For every $\varepsilon > 0$, there exists a family of binary codes $\{C_i\}_{i=1}^{\infty}$ for the* $\text{BDC-R-L-Bounded}(d, \mu, M)$ *where the block length C_i goes to infinity as $i \to \infty$ and*

1. *C_i is encodable in linear time and decodable in quasi-linear time.*
2. *The decoding failure probability is $\exp(-\Omega(n))$.*
3. *The rate of the C_i is $R > \text{Cap}(\text{BDC-R-L-Bounded}(d, \mu, M)) - \varepsilon$.*

Outer and inner codes. The coding scheme follows from code concatenation. For the outer code, we use the code by Haeupler and Shahrasbi [HS21] that can correct from adversarial insertions and deletions (insdel errors, for short).

Theorem 11 ([HS21])**.** *For every $\epsilon_{\text{out}}, \delta_{\text{out}} \in (0, 1)$ there exists a family of codes $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$, where $\mathcal{C}_n$ has blocklength n, of rate $\mathcal{R}_{\text{out}} = 1 - \delta - \epsilon_{\text{out}}$ over an alphabet Σ of size $|\Sigma| = O_{\epsilon_{\text{out}}}(1)$ that can correct δn adversarial insdel errors. The codes $\mathcal{C}_n$ support linear time encoding and quasi-linear time decoding.*

[3] The monotonicity assumption is for simplicity and also since it makes sense to assume that longer runs are more likely to suffer from higher deletion rate.

[4] By constants, we mean with respect to n, the length of the code.

We will denote by C_{out} the outer code given by Theorem 11. By C_{in}, we denote the inner code that achieves capacity on the 0-trimming BDC-R-L-Bounded(d, μ, M) channel, given by Corollary 2. Formally, our encoding process is as follows:

Encoding. Given as input a message $x \in \Sigma^{\mathcal{R}_{\text{out}} n}$, we encode it with the code given in Theorem 11 to obtain an outer codeword $c^{(\text{out})} = (\sigma_1, \ldots, \sigma_n) \in C_{\text{out}} \subset \Sigma^n$. Then, every symbol in $c^{(\text{out})}$, $\sigma_i \in \Sigma = \{0,1\}^{m \cdot \mathcal{R}_{\text{in}}}$, is encoded using the inner code to a codeword that we denote $c^{(\text{in})}_{\sigma_i}$. We thus get a codeword in the concatenated code

$$\left(c^{(\text{in})}_{\sigma_1}, \ldots, c^{(\text{in})}_{\sigma_n}\right) \in C_{\text{out}} \circ C_{\text{in}} \, .$$

Now that we have a codeword in the concatenated code, we add an additional layer of encoding that is crucial for preserving synchronization. Every two adjacent inner codewords are separated by a buffer of zeros of length $\lceil \nu \cdot m/(1-d(M)) \rceil$ where recall that ν is a small constant such that $\nu \leq \mu \cdot \epsilon_{\text{out}}$ (recall that by Definition 11, μ is such that $d(M) \leq 1-\mu$ and that $d(M)$ is the largest deletion probability the channel can impose).

Remark 3. *We note that in order for the concatenation to make sense, we must have $\mathcal{R}_{\text{in}} m = \lceil \log_2 |\Sigma| \rceil = \lceil \log_2(O_{\epsilon_{\text{out}}}(1)) \rceil$. We shall choose a small enough ϵ_{out} so that the length of the buffer will be $\geq M$, which would imply that the channel will apply deletions to it with probability $d(M)$.*

Decoding. The decoding consists of the following step:

1. Decoding buffers: Identify all runs of the symbol 0 in the received string that are of length at least $\frac{\nu m}{2}$.

   Denote by $\tilde{c}_1, \ldots, \tilde{c}_s$ the strings obtained in this step.

2. Decode each $\tilde{c}_i$ (brute force) into an outer code symbol $\tilde{\sigma}_i$.

3. Run the outer code decoding algorithm on $\tilde{\sigma}^{(\mathsf{out})} = (\tilde{\sigma}_1, \ldots, \tilde{\sigma}_s)$ to obtain $\tilde{x}$ and return $\tilde{x}$.

Proof of Theorem 10. Our goal is to show that with probability $1 - \exp(-\Omega(n))$, it holds that $\mathrm{ED}(\sigma^{(\text{out})}, \tilde{\sigma}^{(\text{out})}) < \delta_{\text{out}} n$. This would imply that Step 3 in our decoding algorithm succeeded and the correct message is returned.

There are three types of error that can increase the edit distance between $\sigma^{(\mathsf{out})}$ and $\tilde{\sigma}^{(\mathsf{out})}$:

1. *Deleted buffer:* This happens when the channel deletes too many bits from a buffer so that less than $\nu m/2$ bits of the transmitted buffer survived the channel, and we did not identify this buffer when we identify buffers in Step 1.

   We claim that this happens with probability $\exp(-\Omega(m))$. Indeed, recall that the length of the buffer $\lceil \nu m/(1-d(M)) \rceil$ is greater than M. Now, let Z be the random variable that corresponds to the number of bits from that buffer that survived the transmission through the BDC-Thr$(\tau, d_{\mathsf{short}}, d_{\mathsf{long}})$. Clearly, $Z \sim \mathrm{Bin}(\lceil \nu m/(1-d(M)) \rceil, 1-d(M))$ and then, By Chernoff's bound

$$\Pr\left[Z < \frac{\nu m}{2}\right] \leq \Pr\left[Z < \left(1 - \frac{1}{2}\right)\nu m\right] < \exp\left(-\frac{1}{8}\nu m\right) .$$

2. *Spurious buffer:* In this case the algorithm mistakenly identifies a buffer inside an inner codeword. This might happen if there are many consecutive runs of the symbol 1 that were deleted by the channel. As a result, a long run of the symbol 0 is created and the algorithm will mistakenly identify it as a buffer in Step 1.

   We claim that this event also happens with probability $\exp(-\Omega(m))$. Clearly, in order for the channel to create a run of zeros of length $\nu m/2$, it must be that the channel deleted all 1s from an interval of length at least $\nu m/2$. Fix an interval of length $\nu m/3$ in an inner codeword. By Corollary 2, this interval contains at least $\gamma\nu m/3$ ones. Let $a_i, i \in [M-1]$ be the number of ones in this interval that belong to runs of length i and denote by a_M, the number of bits that belong to runs of length $\geq M$ in this interval. Deleting all these ones happens with probability

$$d(1)^{a_1} d(2)^{a_2} \cdots d(M)^{a_M} \leq \left(\frac{3 \cdot \left(\sum_{i=1}^{M} a_1 d(i) \right)}{\gamma\nu m} \right)^{\gamma\nu m/3} \leq (1-\mu)^{\gamma\nu m/3} = \exp(-\Omega(m)) \ ,$$

   where the first inequality follows from the weighted AM-GM inequality and the second inequality is due to the assumption that $d(1) \leq d(2) \leq \cdots \leq d(M) \leq 1-\mu$ and that $\sum_{i=1}^{M} a_i = \gamma\nu m/3$. Union bounding over all such intervals, we get also that the probability of the existence of a spurious buffer inside an inner codeword is at most $m \cdot \exp(-\Omega(m)) \cdot \exp(-\Omega(m))$.

   Note that the above argument bounds the probability that a spurious buffer is identified inside an inner codeword. However, it can be that the decoder identifies several spurious buffer. The maximal number of spurious buffers the decoder can identify inside an inner codeword is at most $2/(\gamma\nu) = O(1)$.

3. *Wrong inner decoding:* Assuming that the buffers before and after were identified correctly and there were no spurious buffers in between, the decoding of the inner code returns a different inner codeword.

   Here, by Corollary 2, the decoding failure probability is ε which can be made as small as we wish.

We analyze the contribution of each of the error types (deleted buffer, spurious buffer and wrong inner decoding) on $\mathrm{ED}(\sigma^{(\mathrm{out})}, \tilde{\sigma}^{(\mathrm{out})})$. A deleted buffer causes two inner codewords to merge and thus be decoded incorrectly by the inner code's decoding algorithm. This introduces two deletions and one insertion in the outer code's level. Similarly, one can verify that a single spurious buffer inside an inner codeword introduces one deletion and two insertions. Furthermore, ℓ spurious buffers inside an inner codeword introduce at most $\ell+1$ insertions and one deletion. Finally, a wrong inner decoding causes a substitution which is equivalent to one deletion followed by one insertion.

As mention above, the outer decoding algorithm fails if $\mathrm{ED}(\sigma^{(\mathrm{out})}, \tilde{\sigma}^{(\mathrm{out})})$. For this to happen, at least one of the following must occur:

1. $\delta_{\mathrm{out}} n/9$ deleted buffers.

2. $\delta_{\mathrm{out}} n/9$ spurious buffers.

3. $\delta_{\mathrm{out}} n/9$ wrong inner decodings.

Observe that for large enough m and small enough ε, we get that the probability for a deleted buffer or a wrong inner decoding is $< \delta_{\mathrm{out}}/10$ and thus the probability that there are $\delta_{\mathrm{out}} n/9$ deleted buffer or wrong inner decoding is at most $\exp(-\Omega(n))$. Furthermore, since the expected number of spurious buffers between two adjacent real buffers is $\exp(-\Omega(m)) < \delta_{\text{out}}/10$ for large enough m and the maximal number of such spurious buffers is $O(1)$, we can apply the **??** and get that the total number of all spurious buffers is $< \delta_{\text{out}} n/9$ with probability $1 - \exp(-\Omega(n))$. The claim about the decoding failure follows.

We now justify the time complexity of our encoding and decoding algorithms. The complexity of the encoder is as follows. The encoder of the outer code runs in linear time. Then, the encoding of each outer code symbol using the inner code is performed in constant time and thus the encoding all the n symbols is done in $O(n)$. Finally, adding the buffers also takes linear time. The decoding complexity is dominated by the outer code's decoding time which is quasi-linear. Indeed, identifying the buffers and the decoding of all the corrupted inner codewords takes $O(n)$.

Finally, we show that the rate of this scheme is as claimed.

$$\begin{aligned} R &= \frac{\log_2 |\Sigma|^{\mathcal{R}_{\text{out}} n}}{mn + (n-1)\lceil \nu m/(1-d(M)\rceil} \\ &\geq \frac{\mathcal{R}_{\text{in}}\mathcal{R}_{\text{out}}}{1 + \nu/(1-d(M)) + 1/m} \\ &\geq \frac{(1-\delta_{\text{out}} - \epsilon_{\text{out}})\mathcal{R}_{\text{in}}}{1 + \nu/\mu + 1/m} \end{aligned}$$

where the last inequality follows since $1 - d(M) > \mu$.

Now, observe that for any $\epsilon > 0$, one can choose small enough $\epsilon_{\text{out}}, \delta_{\text{out}}$, and μ so that the rate is at least $\mathcal{R}_{\text{in}} - \epsilon$ so that the final rate is indeed $\text{Cap}(\text{BDC-R-L-Bounded}(d, \mu, M)) - \varepsilon$. The theorem follows. $\square$

Remark 4. *We emphasize the order in which we choose the parameters of the scheme. We first choose the constant ε which is the gap to capacity we want to achieve. Then, we choose small enough $\delta_{\text{out}}, \epsilon_{\text{out}}$, and ν which gives this gap to capacity and ensures that all the failure probabilities computed in the proof are indeed smaller than $\delta_{\text{out}}/10$ (recall that deleted buffer and spurious buffer happen with probability $\exp(-\Omega(m)) = \exp(-\Omega(\log_2(|\Sigma|/\mathcal{R}_{\text{in}}))) = (O_{\epsilon_{\text{out}}}(1))^{-\Omega(1)}$).*

6 Acknowledgments

The authors thank Elena Grigorescu for many fruitful discussions on this problem.

References

[AT23] Dar Arava and Ido Tal. Stronger polarization for the deletion channel. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1711–1716. IEEE, 2023.

[BLS20] Joshua Brakensiek, Ray Li, and Bruce Spang. Coded trace reconstruction in a constant number of traces. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 482–493, 2020.

[CGMR20] Mahdi Cheraghchi, Ryan Gabrys, Olgica Milenkovic, and João Ribeiro. Coded trace reconstruction. *IEEE Transactions on Information Theory*, 66(10):6084–6103, 2020.

[CR21] Mahdi Cheraghchi and João Ribeiro. An overview of capacity results for synchronization channels. *IEEE Transactions on Information Theory*, 67(6):3207–3232, 2021.

[CS22] Roni Con and Amir Shpilka. Improved constructions of coding schemes for the binary deletion channel and the Poisson repeat channel. *IEEE Transactions on Information Theory*, 68(5):2920–2940, 2022.

[dBE52] Nicolaas Govert de Bruijn and Paul Erdős. Some linear and some quadratic recursion formulas. ii. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen: Series A: Mathematical Sciences*, 14:152–163, 1952.

[Dob67] Roland L. Dobrushin. Shannon's theorems for channels with synchronization errors. *Problemy Peredachi Informatsii*, 3(4):18–36, 1967.

[Fei59] Amiel Feinstein. On the coding theorem and its converse for finite-memory channels. *Information and Control*, 2(1):25–44, 1959.

[FR20] Zoltán Füredi and Imre Z Ruzsa. Nearly subadditive sequences. *Acta Mathematica Hungarica*, 161(2):401–411, 2020.

[GL19] Venkatesan Guruswami and Ray Li. Polynomial time decodable codes for the binary deletion channel. *IEEE Transactions on Information Theory*, 65(4):2171–2178, 2019.

[HMG19] Reinhard Heckel, Gediminas Mikutis, and Robert N Grass. A characterization of the DNA data storage channel. *Scientific reports*, 9(1):9663, 2019.

[HS21] Bernhard Haeupler and Amirbehshad Shahrasbi. Synchronization strings: Codes for insertions and deletions approaching the Singleton bound. *J. ACM*, 68(5), September 2021.

[LT21] Yonglong Li and Vincent Y. F. Tan. On the capacity of channels with deletions and states. *IEEE Transactions on Information Theory*, 67(5):2663–2679, 2021.

[MBT10] Hugues Mercier, Vijay K. Bhargava, and Vahid Tarokh. A survey of error-correcting codes for channels with symbol synchronization errors. *IEEE Communications Surveys Tutorials*, 12(1):87–96, First Quarter 2010.

[MD24] Ruslan Morozov and Tolga M. Duman. On the capacity of channels with Markov insertions, deletions and substitutions. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 3444–3449, 2024.

[MDK18] Wei Mao, Suhas N. Diggavi, and Sreeram Kannan. Models and information-theoretic bounds for nanopore sequencing. *IEEE Transactions on Information Theory*, 64(4):3216–3236, 2018.

[Mit09] Michael Mitzenmacher. A survey of results for deletion channels and related synchronization channels. *Probability Surveys*, 6:1–33, 2009.

[MSV24] Brendon McBain, James Saunderson, and Emanuele Viterbo. On noisy duplication channels with Markov sources. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 3438–3443, 2024.

[MVS24] Brendon McBain, Emanuele Viterbo, and James Saunderson. Information rates of the noisy nanopore channel. *IEEE Transactions on Information Theory*, 70(8):5640–5652, 2024.

[PLW22] Francisco Pernice, Ray Li, and Mary Wootters. Efficient capacity-achieving codes for general repeat channels. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 3097–3102, 2022.

[PT21] Henry D. Pfister and Ido Tal. Polar codes for channels with insertions, deletions, and substitutions. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 2554–2559, 2021.

[PW24] Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2024. Available at https://people.lids.mit.edu/yp/homepage/data/itbook-export.pdf.

[RRC+13] Michael G Ross, Carsten Russ, Maura Costello, Andrew Hollinger, Niall J Lennon, Ryan Hegarty, Chad Nusbaum, and David B Jaffe. Characterizing and measuring bias in sequence data. *Genome biology*, 14:1–20, 2013.

[Rub22] Ittai Rubinstein. Explicit and efficient construction of nearly optimal rate codes for the binary deletion channel and the Poisson repeat channel. In *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, pages 105:1–105:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

[TPFV22] Ido Tal, Henry D. Pfister, Arman Fazeli, and Alexander Vardy. Polar codes for the deletion channel: Weak and strong polarization. *IEEE Transactions on Information Theory*, 68(4):2239–2265, 2022.