Electronic Colloquium on Computational Complexity, Report No. 112 (2024)

The Rate of Interactive Codes is Bounded Away from 1

Abstract

Kol and Raz [STOC 2013] showed how to simulate any *alternating* two-party communication protocol designed to work over the noiseless channel, by a protocol that works over a stochastic channel that corrupts each sent symbol with probability $\epsilon > 0$ independently, with only a $1 + \mathcal{O}\left(\sqrt{\mathbb{H}(\epsilon)}\right)$ blowup to the communication. In particular, this implies that the *maximum rate* of such *interactive codes* approaches 1 as ϵ goes to 0, as is also the case for the maximum rate of classical error correcting codes. Over the past decade, followup works have strengthened and generalized this result to other noisy channels, stressing on how fast the rate approaches 1 as ϵ goes to 0, but retaining the assumption that the noiseless protocol is alternating.

In this paper we consider the general case, where the noiseless protocols can have *arbitrary orders of speaking.* In contrast to Kol-Raz and to the followup results in this model, we show that the maximum rate of interactive codes that encode general protocols is upper bounded by a universal constant strictly smaller than 1. To put it differently, we show that there is an inherent blowup in communication when protocols with arbitrary orders of speaking are faced with any constant fraction of errors $\epsilon > 0$. We mention that our result assumes a large alphabet set and resolves the (non-binary variant) of a conjecture by Haeupler [FOCS 2014].

ISSN 1433-8092

Contents

1 Introduction

One of the gems in Shannon's famous 1948 paper introducing information theory [Sha48] is the *channel capacity* formula, that gives the maximum rate possible for an error correcting code over any discrete memoryless channel. Recall that an error correcting code with rate r allows one party to reliably communicate a message consisting of n symbols to a remote second party, with a negligible probability of error, by sending only $n \cdot (1/r + o(1))$ symbols over the channel. Let $\mathsf{C}_{\Gamma,\epsilon}$ be the *symmetric channel* with alphabet set Γ and noise rate ϵ.[1] The channel capacity formula shows that the maximum rate over $\mathsf{C}_{\Gamma,\epsilon}$ approaches 1 as the noise rate ϵ approaches 0. For instance, the maximum rate over $\mathsf{C}_{\{0,1\},\epsilon}$ is $1 - \mathbb{H}(\epsilon)$, where $\mathbb{H}$ is the binary entropy function.

Schulman's groundbreaking work [Sch92] studied error correcting codes in the "two-way" setting, where there are noisy channels between the two communicating parties in both directions. Such error correcting codes are called *interactive codes* and they allow the encoding of interactive protocols, which may consist of many back-and-forth messages, in a noise-resilient way. Following Schulman's question regarding the maximum rate of interactive codes [Sch92], Kol and Raz [KR13] defined the notion of *interactive channel capacity*, which is the analogue of channel capacity in the interactive setting. For every $\epsilon > 0$, they designed an interactive code with rate $r_\epsilon = 1 - \mathcal{O}(\sqrt{\mathbb{H}(\epsilon)})$ over the two-way binary symmetric channel, under the assumption that the protocol being encoded is alternating[2].

It is not hard to see that the interactive coding scheme of Kol and Raz [KR13] also works for the $\mathsf{C}_{\Gamma,\epsilon}$ channel, for every Γ. Their result, stated for such channels, is that for any $\epsilon > 0$, any alphabet set Γ, and any *alternating* protocol Π with alphabet Γ, there exists a protocol Π' that simulates Π over $\mathsf{C}_{\Gamma,\epsilon}$ with negligible error, and has length $|\Pi| \cdot (1/r_\epsilon + o(1))$, where $|\Pi|$ is the length of Π. Observe that, as in the classical setting, the maximum rate approaches 1 as ϵ approaches 0. Following [KR13], the dependence of the maximum rate on ϵ, under the same alternating turns assumption, was further improved by [Hae14] to $1 - \mathcal{O}(\sqrt{\epsilon})$, and was also studied for other two-way channels, including the *adversarial* channel [Hae14, CS19], the (adversarial) *feedback* channel [Pan13, GH14], the adversarial *erasure* channel [GH14], and the adversarial *insertion-deletion* channel [HSV18].

1.1 Our Result

The main result of this paper is Theorem 1.1, showing that in the general case, where the order of speaking in the noiseless communication protocols Π may be *arbitrary*, the maximum achievable rate is bounded away from 1.

Theorem 1.1. *For every $\epsilon > 0$, there exists a set Γ and a deterministic protocol Π with alphabet Γ, such that any randomized protocol Π' that simulates Π over $\mathsf{C}_{\Gamma,\epsilon}$ with probability*

[1] That is, the input and output alphabets of the channel $\mathsf{C}_{\Gamma,\epsilon}$ are Γ, $|\Gamma| \geq 2$. On a sent symbol $z \in \Gamma$, the channel outputs z with probability $1 - \epsilon$, and with probability ϵ, it outputs a random symbol in Γ.

[2] That is, Alice sends a message to Bob in all odd rounds, and vice versa.

0.99 *has length at least* $|\Pi|/(1-\Omega(1))$.

Observe that since Theorem 1.1 holds for the $\mathsf{C}_{\Gamma,\epsilon}$ channel that has stochastic noise and for public-coin protocols Π', it also holds for adversarial noise and private-coin protocols. Furthermore, our proof of Theorem 1.1 actually proves a much stronger claim (see Section 2). For example, it implies that the maximum rate of an interactive code over the *feedback* channel that randomly *erases* a *single* communicated symbol (*i.e.*, one of the sent symbols, selected uniformly at random, is received as '$\perp$' and the sender is notified) is only $1-\Omega(1)$ (*cf.* the results of [KR13, Pan13, Hae14, GH14, HSV18, CS19] for such channels with maximum rate approaching 1).

We mention that our result settles the (non-binary version) of a conjecture by Haeupler (Conjecture 1.1 in [Hae14]), that also appears in Haeupler and Gelles (Question 3 in Section 7 of [GH14]) and in Gelles's excellent survey (Question 2 in Section 5 of [Gel17]). While lower bounds on the maximum rate of various two-way channels (*i.e.*, upper bounds on the overhead of interactive codes) are known, prior to our work, the only non-trivial upper bound was due to [KR13] and is extremely involved (see Section 1.2).

We also mention that Theorem 1.1 uses a *large alphabet* set (specifically, we need $|\Gamma| = \mathrm{poly}(|\Pi|)$), as for such alphabets the single erased symbol cannot be *guessed* by the receiver with high probability (in the binary $|\Gamma| = 2$ case, the erased symbol can be guessed with probability $\frac{1}{2}$). Nevertheless, we believe that Theorem 1.1 still holds for the binary setting (fixing $\Gamma = \{0,1\}$), and proving it is an outstanding question we leave open. Other interesting directions for future work include finding the maximum rate of interactive codes over $\mathsf{C}_{\Gamma,\epsilon}$, say when ϵ approaches 0, and characterizing the "hard" communication orders resulting in maximum rates bounded away from 1.

Finally, we wish to point out a corollary of Theorem 1.1: Many works involving interactive protocols (in the noisy or noiseless settings) assume an alternating order of speaking, as it is often simpler to deal with and only incurs *at most* a factor 2 blowup to the communication. Theorem 1.1 shows that this transformation of general protocols to alternating ones incurs *at least* a factor c blowup, for some $c > 1$: Assume that the blowup is only by a $1 + o(1)$ factor. By converting the hard-to-simulate protocol Π from Theorem 1.1 to a protocol with alternating turns and then applying the [KR13] scheme, we obtain a noise-resilient protocol Π' that simulates Π with only $1+o(1)$ blowup to the communication, in contradiction to Theorem 1.1.

1.1.1 Techniques

The proof of Theorem 1.1 is quite involved and a detailed overview can be found in Section 2. In this section we give some of the highlights of our proof.

Theorem 1.1 is proved by combining Theorem 4.1 and Theorem 4.2. As mentioned above, our result holds even over the very mildly noisy channel that has feedback and only randomly erases a single communicated symbol. Theorem 4.1 considers a pointer chasing protocol with order of speaking σ[3] and shows that it can only be simulated over this mildly noisy channel by a protocol with order of speaking σ', for which σ is a *strong subsequence* of σ'. By a strong subsequence, we mean that for most coordinates i' of σ', σ remains a subsequence of σ' even after coordinate i' is removed. Observe that given Theorem 4.1, to prove Theorem 1.1, all we have to do is exhibit a σ such that any σ' for which σ is a strong subsequence of σ' is a constant factor longer than σ. This is done in Theorem 4.2.

At a high level, Theorem 4.1 is proved by proving a *generalized pointer chasing lower bound*: while prior pointer chasing lower bounds assume that players alternate (*e.g.*, [NW91]), our proof holds for *any* order of speaking. To analyze cases where one of the parties speaks in several consecutive rounds, we use a lower bound for a generalization of the well-known *Index problem*, where the communication is not one-way, but the party holding the index speaks substantially less than what it takes to convey the index. To see why this lower bound is useful, assume for example that in the noiseless protocol Alice speaks three times and then Bob speaks once, *i.e.*, $\sigma = AAAB$. We think of Alice's message in those three rounds as an index i, and of Bob's input as a vector v. When Bob speaks in the fourth round he gives v_i to Alice. Now consider a simulation protocol with order of speaking $\sigma' = BABABA$. Can Bob give v_i to Alice? We show that he cannot. The reason is that Alice only speaks in two instead of three rounds before Bob's final round, thus she can only give partial information about i, which is not enough for Bob to compute v_i.

Theorem 4.2 is a purely combinatorial claim about strong subsequences, and is shown using the probabilistic method. We provide a detailed overview in Section 2.2, but for the high level idea, consider, for any $T > 0$ the pair of strings $(\sigma, \sigma') = \big((AB)^T, (AB)^{T+1}\big)$, and observe that σ is a strong subsequence of σ' and σ' is almost the same length as σ. Roughly speaking, our proof shows that the *only* reason σ is a subsequence of σ' for such a short σ', is that σ is highly "predictable", in the sense that one can "guess" the symbols after coordinate i based on the previous symbols. We formalize this notion and show that a uniformly random σ is not predictable, and use this to show that for most σ, no short σ' will be such that σ is a strong subsequence of σ'.

1.2 Additional Related Work

We next survey the most relevant work on the maximum rates of interactive codes over different channels.

The $\mathsf{C}_{\{0,1\},\epsilon}$ channel. The study of error correcting codes for interactive communication was pioneered by Schulman [Sch92], who showed how to transform any interactive communication protocol over the (noiseless) binary channel to an equivalent noise-resilient protocol that works over the (two-way) $\mathsf{C}_{\{0,1\},\epsilon}$ channel, with only a constant overhead in the communication. This shows that for any $\epsilon < \frac{1}{2}$, the maximum rate of an interactive code over

[3] We think of the order of speaking in a communication protocol as a string $\sigma \in \{A, B\}^*$, where $\sigma_i = A$ means that Alice speaks in round i, and $\sigma_i = B$ means that Bob speaks in round i.

$\mathsf{C}_{\{0,1\},\epsilon}$ is at least some constant strictly greater than 0.

Kol and Raz [KR13] studied the maximum rate r_ϵ achievable by any interactive code over $\mathsf{C}_{\{0,1\},\epsilon}$, but as mentioned above, it is not hard to see that their results hold for every channel $\mathsf{C}_{\Gamma,\epsilon}$. They showed that for alternating noiseless protocols and protocols whose communication order is periodic with a small period, $r_\epsilon = 1 - O\left(\sqrt{\mathbb{H}(\epsilon)}\right)$. The assumption that the noiseless protocol is alternating (or has a small period) is crucial as their coding scheme uses the rewind-if-error mechanism [Sch92], where the parties run the noiseless protocol over the noisy channel, and periodically compare their received transcripts to detect errors. If an error was detected, the parties "rewind" to the last agreed upon point and continue the execution of the noiseless protocol from that point. Since the noiseless protocol is assumed to be alternating, by taking the order of speaking of the simulating protocol to also be alternating, they can ensure that when rewinding, the order of speaking in the simulation matches the assumed order of speaking in the noiseless protocol. Kol and Raz also proved a matching upper bound of $1 - \Omega\left(\sqrt{\mathbb{H}(\epsilon)}\right)$ for some carefully chosen communication orders[4].

We mention that the Kol-Raz result gives the first separation between the maximum rate of classical error correcting codes and that of interactive codes, and observe that Theorem 1.1 gives a substantially stronger separation.

Building on [KR13] and also assuming an alternating order of speaking, [GHK$^+$16] presented a deterministic coding scheme that achieves a rate of r_ϵ (the [KR13] scheme is randomized), and [BKOS21] gave a coding scheme that handles larger ϵ's (observe that the [KR13] scheme is only meaningful for small ϵ's). Specifically, [BKOS21] showed that the maximum rate of interactive codes over $\mathsf{C}_{\{0,1\},\epsilon}$ is at least 0.0302 times the maximum rate of classical error correcting codes over $\mathsf{C}_{\{0,1\},\epsilon}$.

Other (non-adaptive) channels. The maximum rates of interactive codes over other two-way channels, that are well studied in the context of classical codes, were also considered with the alternating communication order assumption. Pankratov [Pan13] studied the rate of interactive codes over channels with random errors and *feedback*, and gave a scheme with rate $1 - O(\sqrt{\epsilon})$. Haeupler and Gelles [GH14] improved his result and gave a scheme with rate $1 - \Theta(\mathbb{H}(\epsilon))$ that works for the adversarial feedback channel. A scheme with the same rate was also given by [GH14] for the adversarial *erasure* channel. The *adversarial* channel with corruption errors (bit flips) was considered in [Sch93, BR11, Hae14, CS19], and an interactive code with rate r_ϵ for this model was presented in [CS19]. The adversarial binary *insertion-deletion* channel was considered by [HSV18], who demonstrated an interactive coding scheme with rate r_ϵ for it.

[4] Specifically, the upper and lower bounds match when the communication order is periodic with a period k that satisfies $\epsilon = \Theta\left(\frac{\log k}{k^2}\right)$. Indeed, Hauepler and Velingker [HV17] showed that if the parties alternate in sending $k = \Omega(\text{poly}(1/\epsilon))$ consecutive symbols, then the maximum rate is $1 - \Theta(\mathbb{H}(\epsilon))$, violating the upper bound of [KR13].

Adaptive channels. In this paper as well as in most of the prior works on interactive coding, including all the paper surveyed so far, the assumption is that the protocols Π and Π' have a *non-adaptive* (*a.k.a*, *oblivious* or *static*) communication order, meaning that the order of communication in the protocol is fixed in advance. Haeupler [Hae14] considered the *adaptive* setting, where at any round, each party decides whether to send a bit or listen for one based on its input and received transcript (which, in turn, depends on the channel's noise). Observe, however, that protocols in these models may have several parties attempting to send a symbol in the same round, or even no senders at all, and the received bits in these cases need to be specified.

Haeupler [Hae14] constructed interactive codes that encode non-adaptive protocols Π (with any communication order) by adaptive protocols over the $\mathsf{C}_{\{0,1\},\epsilon}$ channel with rate $1-\mathcal{O}(\sqrt{\epsilon})$, bypassing the upper bound of [KR13]. Put together, [KR13] and [Hae14] imply a separation between the maximum rates obtained via adaptive and non-adaptive encodings. [Hae14] conjectured that this separation can be strengthened, even for a single erasure error, and our Theorem 1.1 proves his conjecture. We mention that other adaptive models were studied in the literature, see *e.g.*, [AGS16, EHK20].

Acknowledgements

We thank Sepehr Assadi for Lemma A.7 (and a clean proof of it).

2 Overview

Our main result (Theorem 1.1) says that regardless of how small the noise parameter ϵ is, the overhead required to simulate a noiseless channel over a noisy channel that corrupts each symbol with probability ϵ independently, is a constant strictly larger than 1. As mentioned in Section 1.1, we will actually prove a much stronger version of this, showing that it holds even if the channel corrupts exactly one (uniformly chosen) round, and the parties know in advance which round it is (and therefore, can change the simulation protocol they use arbitrarily based on this round, as long as this change does not affect the order in which the parties speak in the other rounds).

Showing a lower bound for a simulation protocol in a noise model that allows the protocol to change in response to the noise essentially means that we have to show a lower bound for a *noiseless* protocol, where all we know about the noiseless protocol is that its order of speaking is the same regardless of which round is corrupted by noise. Thus, a big part of our proof (Section 5) is, given two orders of speaking σ, σ', understanding when can noiseless protocols with order of speaking σ' simulate noiseless protocols with order of speaking σ. This part subsumes and generalizes famous "pointer-chasing" and "round-complexity" lower bounds in communication complexity [NW91, *e.g.*] and is overviewed in Section 2.1. The answer turns out to be quite elegant: σ' can simulate σ if and only if σ is a subsequence of σ'.

We now look back at our original problem of designing a noiseless protocol Π that cannot be simulated by any (short) protocol over a noisy channel, even when the noise corrupts only one random symbol in the simulation protocol that is known to the parties as soon as they fix the order of speaking in the simulation protocol. Having shown that an order σ' can simulate σ if and only if σ is a subsequence of σ', this means that we have to construct an order of speaking σ (which will be the order in which the parties speak in Π) such that any short order of speaking σ' satisfies the property that σ is not a *"strong" subsequence* of σ'. By that we mean that removing one uniformly chosen coordinate from σ' ensures that, with high probability, σ is not a subsequence of σ' with that coordinate removed (see Definition 3.7). This is the second main part of our proof and is described in Section 6 and overviewed in Section 2.2.

2.1 Lower Bounds on Noiseless Simulations

Recall that the order of speaking for a protocol Π of length T is a string $\sigma \in \{A, B\}^T$ that captures the order in which Alice and Bob speak in Π, in the sense that, for all $i \in [T]$, party σ_i is the party speaking in round i of Π. The goal of this section is to show that, given any two orders of speaking σ and σ', all (noiseless) protocols with order of speaking σ can be simulated by (noiseless) protocols with order of speaking σ' if and only if σ is a subsequence of σ'. The "if" direction is straightforward and we shall focus on showing the "only if" direction.

We argue this in the contrapositive. Suppose that two orders σ and σ' are given such that σ is not a subsequence of σ'. We first note that if σ is alternating, *i.e.*, σ is of the form $ABABA\ldots$, then the desired result follows from (an easy extension of) the pointer chasing lower bounds in [NW91] and subsequent work. However, a lower bound only for alternating σ is not good enough for us, as we want the lower bound for a string σ such that any short σ' satisfies the property that σ is not a strong subsequence of σ'. This is provably not the case for alternating σ as for any alternating σ, the string $\sigma' = AB||\sigma$ satisfies the property that σ is a strong subsequence of σ', where $||$ denotes concatenation.

However, as our lower bound must subsume these lower bounds, it is important to understand them. For this, consider the case when $\sigma = AB$ so that σ' (as σ cannot be a subsequence of σ') is of the form $BB\ldots AAAA\ldots$, say $\sigma' = BBBAAA$. Consider now the well-known Index problem, where Bob has a large array and Alice has an index for the array, and the goal of the parties is to output the element at Alice's index in Bob's array. There is a simple protocol with order of speaking σ that solves this problem, where Alice first sends her index and then Bob sends the element at that index. However, if the order of speaking is restricted to be σ' there is no way for Bob to send the right element to Alice, as all his messages are before he acquires any knowledge of Alice's index (unless of course, he sends to Alice the entire array, but this is impossible if Alice's alphabet is large enough).

For our more general result, we first extend the above lower bound to a more general class of σ that has many Alice messages before the last Bob message, say, $\sigma = AAAB$. The hard protocol for these σ is also the protocol for the Index problem except that this time, Alice's index is so large that it will not fit in one message (and requires three messages). For all σ' where all Bob's messages precede all Alice's messages, the argument is the same as before, but this time there are additional σ' that are not of the form above and satisfy that σ is not a subsequence of σ', for example $\sigma' = BABABA$.

When $\sigma' = BABABA$, as a protocol with order of speaking σ' proceeds, Bob does get some messages from Alice (in rounds 2 and 4) but these messages are not long enough to contain her entire index. Thus, to show a lower bound for such σ', we need to extend the aforementioned lower bound for Index to work for protocols where Bob has partial information about Alice's index. This is exactly what we do, showing that such partial information from Alice cannot help Bob in guessing the right index a whole lot, and he still cannot send her the right index without sending a huge portion of his array. However, Bob cannot send a huge portion without having high communication, which is impossible if σ' is not much longer than σ.

To extend this argument to general σ and σ' such that σ is not a subsequence of σ', we break the string σ into "intervals", where an interval is defined a set of consecutive rounds where the same party is speaking, *e.g.*, the first three Alice rounds in $\sigma = AAAB$. For each such interval starting from the first, we treat it like the Index problem above, and show that the interval cannot be simulated unless the party speaking in that interval has spoken enough times in the simulation. Once the party has spoken enough times, we remove the interval from σ and the corresponding rounds from σ' and arrive at a smaller problem with a fewer number of intervals. As σ is not a subsequence of σ', we will run out of rounds in σ' before we run out of intervals in σ, giving us a trivial protocol for a non-trivial task, a contradiction.

2.2 Analysis of Strong Subsequences

In this part, our goal is to show that there exists an order of speaking $\sigma \in \{A, B\}^*$, such that for any $\sigma' \in \{A, B\}^*$ for which σ is a strong subsequence of σ', it holds that σ' is a constant factor longer than σ. Recall that σ is a strong subsequence of σ' if, for most coordinates i of σ', it holds that σ is a subsequence of σ' with coordinate i removed. Throughout this section, we will disregard the connection of σ and σ' to communication protocols, and look at them simply as strings in $\{A, B\}^*$. Also, we let T be the length of σ and assume that the length of σ' is $T' = (1 + \delta)T$, where $\delta > 0$ is a small constant.

Patterns. We will show this using the probabilistic method, categorizing the relevant pairs (σ, σ') into various "patterns", where for each pattern ρ and each σ, there is exactly one σ' such that the pair (σ, σ') is in the pattern ρ. We then show that, for every fixed pattern, and a randomly chosen pair (σ, σ') in the pattern, the probability that σ is a strong subsequence of σ' is extremely small, small enough to union bound over all the patterns, and our result follows.

Specifically, a pattern for us will be defined by a string $\rho \in \{A, B, \bullet\}^{T'}$ such that the number of coordinates of ρ that are equal to the "bullet" symbol $\bullet$ is T. We say that a pair (σ, σ') is in the pattern ρ if it holds that upon "inserting" the string σ in the bullet coordinates of ρ, we get the string σ'. Note that we can indeed restrict attention to the pairs (σ, σ') that are in some pattern, as if a pair is not in any pattern, then it must be the case that σ is not a subsequence of σ', and therefore, it cannot be a strong subsequence of σ' either. Moreover, the number of patterns ρ is at most $\binom{(1+\delta)T}{T} \cdot 2^{\delta T} \leq 2^{\mathcal{O}\left(\delta \log \frac{1}{\delta}\right) \cdot T}$ and we will ensure that, assuming δ is a small enough constant, the relevant probabilities are small enough for a union bound over all the patterns.

Analyzing a toy pattern. Following the above framework, we now fix a pattern ρ and show that for a random pair (σ, σ') in ρ, we have that σ is not a strong subsequence of σ' with high probability. The high level idea here is best understood by taking $\rho = \bullet^T$, the T-length string each of whose coordinates are $\bullet$, even though this is not a valid pattern according to the definition above. However, picking $\rho = \bullet^T$ also means that the only way a pair (σ, σ') can be in this pattern is if $\sigma = \sigma'$, implying that σ is trivially not a strong subsequence of $\sigma' = \sigma$. Thus, to make the argument non-trivial, we will show that, for a uniformly random $\sigma \in \{A, B\}^T$, even a prefix of σ of length $0.9T$ is not a strong subsequence of σ.

For this, consider what happens if we erase the first coordinate of σ to get a string σ_{-1}, and try to estimate the length of the longest prefix of σ that is a subsequence of σ_{-1} (the case when a different coordinate is erased is similar). To estimate the length of the longest prefix, we consider the greedy algorithm "matching" the string σ to the string σ_{-1}: Namely, match each coordinate of σ to the earliest coordinate possible[5] in σ_{-1}. To analyze this algorithm, for all $i \in [T]$, define lag_i to be the difference between i and the coordinate in σ corresponding to the coordinate in σ_{-1} that i is matched to. For example, if coordinate 1 of σ is matched to the first coordinate in σ_{-1} (equivalently, the second coordinate in σ), then, $\mathsf{lag}_1 = 1$.

As σ is uniformly random in $\{A, B\}^T$, each coordinate of σ is uniformly and independently random in $\{A, B\}$, and thus each coordinate in σ will take (in expectation) two coordinates of σ_{-1} to find a match. This means that, in expectation[6], we have $\mathsf{lag}_i \geq \mathsf{lag}_{i-1} + 1$. Using concentration bounds, we can conclude that, except with probability exponentially small in T, at most a 0.9 fraction of the coordinates will end up being matched, implying that the length of the longest prefix of σ that is a subsequence of the resulting string σ_{-1} is at most $0.9T$, as desired.

Towards actual patterns. The argument above does not extend to actual patterns $\rho \in \{A, B, \bullet\}^{T'}$, but for a very specific reason: To understand the reason, note that the argument

[5] For example, to match $AABA$ in the string $ABABAB$, the matching will look like the following (matched characters underlined) $\underline{A}B\underline{ABA}B$.

[6] Note that if a match is found after l coordinates, the lag increases by $l - 1$.

above crucially relied on the fact that lag is non-zero throughout (in fact, it starts from 1 and never decreases). This means that we are always trying to compare a coordinate in σ to another "fresh" coordinate, which is independently and uniformly random, and this allowed us to say that lag increases by 1 in expectation. In fact, if the lag were to have been 0, then we would be comparing every coordinate in σ to itself, which means that it will aways match, and we will therefore be able to match all of σ.

Now, observe that the presence of non-bullet coordinates in ρ can actually decrease the lag and make it 0, ruining our argument above. For an example, consider the case $T' = T+2$, and the pattern $\rho = \bullet, A, B, \bullet^{T-1}$. Specifically, consider the case where the first coordinate is erased, creating a lag of 1. However, as the two coordinates A, B immediately follow the erased coordinate, one can always match the first coordinate of σ to one of these coordinates, bringing the lag back down to 0, and allowing the rest of σ to be matched as is.

To get around this, we use the observation that any non-bullet symbol in ρ can decrease the lag by at most 1. Thus, as the number of non-bullet symbols in σ is δT, if we could somehow magically start with $\mathsf{lag} = \delta T$, then, lag will never vanish for any fixed pattern ρ and we can apply exactly the same analysis as in the toy pattern above to get that σ will not be a subsequence except with probability exponentially small in T. This probability is small enough for us to union bound over all possible ρ and get that except with probability exponentially small in T, a uniformly random σ will not have any σ' such that σ is a strong subsequence of σ'.

Starting with a small lag. All we need to do now is to apply the argument above for large initial lag to the case at hand where the initial lag is 1. For this, observe from our example above (and also in the toy example $\rho = \bullet^T$) that lag will actually increase as i increases, in the sense that the final lag is at least $\Omega(T)$ more than the initial lag, except with exponentially small probability in T. This holds despite the fact that we have a small number ($= \delta T$) of non-bullet symbols, as each such symbol can decrease lag by at most 1, but the much larger number ($= T$) of bullet symbols, each increasing lag by 1 in expectation (as the small number of non-bullet symbols are not enough to make lag vanish), will eventually override the effect of the non-bullet symbols.

The fact that the lag increases can be used as follows: Suppose we are currently considering an i such that lag_i is some value $L > 0$. Consider the pattern starting from the coordinate where i is matched and look at the segment consisting of the next, say, $L/\sqrt{\delta}$ coordinates. As only a δ fraction of the coordinates are non-bullet, this segment is expected to have $\sqrt{\delta} \cdot L < L$ of non-bullet symbols. As the number of non-bullet symbols is smaller than the initial lag, we can conclude that (even after union bounding over all possible ways to place the non-bullet coordinates) except for a "bad" event that happens with probability exponentially small in L, this segment is expected to increase the lag to $L' = \Omega\left(L/\sqrt{\delta}\right)$.

Now, we can consider the next segment of length $L'/\sqrt{\delta}$, and show that except with probability exponentially small in L', this segment is expected to increase lag even more. The increasing length of these segments allows us to show that the sum of the bad probabilities converges to a constant, implying that for any erased coordinate i, one of the following holds: (1) The fraction of non-bullet symbols in one of the segments that are generated is much larger than δ. (2) There exists a segment generated from i for which the bad event occurs. (3) When i is erased, the final lag increases to be $\Omega(T)$.

By Markov's inequality, both Items 1 and 2 will happen for at most a small constant fraction of i. Thus, there exists a σ such that for most i, Item 3 will occur, implying that, for any σ' such that σ is a subsequence of σ', we have that σ is not a subsequence of σ'_{-i}. It follows that there exists a σ such that for no σ' is it the case that σ is a strong subsequence of σ', as desired. In terms of organization, the definition of i in Item 2 is formalized in Definition 6.6 and the proof that there is a small number of such i is in Lemmas 6.7 and 6.10. Analogous statements about Item 1 can be found in Lemma 6.12 while the proof of Item 3 can be found in Section 6.6 (specifically, Lemma 6.15).

3 Model and Preliminaries

3.1 Notation

For $n > 0$, we use $[n]$ to denote the set $\{1, 2, \dots, n\}$. For $a, b > 0$, we use $[a, b]$ to denote the set $\{a, a + 1, \dots, b\}$. Additionally, we use $(a, b]$ to denote the set $\{a + 1, \dots, b\}$. The notations $[a, b)$ and (a, b) are defined analogously.

Let Σ be an alphabet set and $n > 0$ be an integer. For a string $s \in \Sigma^n$ and a set $I \subset [n]$, we use s_I to denote the $|I|$-length string obtained by taking only those coordinates of s that are in I, *e.g.*, we have $(ABAAB)_{\{1,3,4\}} = AAA$. For $i \in [n]$, we sometimes abbreviate $s_{\{i\}}$ to s_i, $s_{[i]}$ to $s_{\leq i}$, and $s_{[n]\setminus\{i\}}$ to s_{-i}. We also use the notations $s_{<i}$, $s_{>i}$ and $s_{\geq i}$ that are defined analogously. Whenever we have $C \in \{A, B\}$, we use $\overline{C}$ to denote the unique element of $\{A, B\}$ not equal to C.

Throughout this paper, we use sans-serif letters to denote random variables.

3.2 Embedding Strings

Let $\sigma, \sigma' \in \{A, B\}^*$ and $T = |\sigma|$ and $T' = |\sigma'|$. For $i \in \{0\} \cup [T]$, we define the function $\mathsf{Emb}(\sigma, \sigma', i)$ inductively as follows:

$$\mathsf{Emb}(\sigma, \sigma', i) = \begin{cases} 0, & \text{if } i = 0 \\ \min(\{\mathsf{Emb}(\sigma, \sigma', i-1) < i' \leq T' \mid \sigma'_{i'} = \sigma_i\} \cup \{T' + i\}), & \text{if } i > 0 \end{cases}. \tag{1}$$

Note that the min above is taken over a finite non-empty set, and is therefore well-defined. We also define the set:

$$\mathsf{E}(\sigma, \sigma') = \{i' \in [T'] \mid \exists i \in [T] : \mathsf{Emb}(\sigma, \sigma', i) = i'\}. \tag{2}$$

We say that σ is a *subsequence* of σ' if $|\mathsf{E}(\sigma, \sigma')| = T$.

Observation 3.1. *Let $\sigma, \sigma' \in \{A, B\}^*$ and $T = |\sigma|$. For all $i_1 \leq i_2 \in \{0\} \cup [T]$, we have:*

$$\mathsf{Emb}(\sigma, \sigma', i_1) + i_2 - i_1 \leq \mathsf{Emb}(\sigma, \sigma', i_2).$$

Lemma 3.2. *Let $\sigma, \sigma' \in \{A, B\}^*$ be given. Additionally, let $i \in [|\sigma|]$, $i' \in \{0\} \cup [|\sigma'|]$ be such that $\mathsf{Emb}(\sigma, \sigma', i - 1) \leq i'$. For all $i'' \geq i' \in [|\sigma'|]$, we have:*

$$\mathsf{Emb}(\sigma, \sigma', i - 1 + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|) \leq i'' - |(i', i''] \setminus \mathsf{E}(\sigma, \sigma')|.$$

Proof. We have by Observation 3.1 that:

$$\begin{aligned}\mathsf{Emb}(\sigma, \sigma', i - 1 + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|) &\leq \mathsf{Emb}(\sigma, \sigma', i - 1) + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']| \\ &\leq i' + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']| \\ &\leq i'' - |(i', i''] \setminus \mathsf{E}(\sigma, \sigma')|.\end{aligned}$$

□

Lemma 3.3. *Let $\sigma, \sigma' \in \{A, B\}^*$ be given. Additionally, let $i \in [|\sigma|]$, $i' \in \{0\} \cup [|\sigma'|]$ be such that $\mathsf{Emb}(\sigma, \sigma', i) > i'$. For all $i'' \geq i' \in [|\sigma'|]$, we have:*

$$\mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|) > i''.$$

Proof. Proof by induction on i''. The base case $i' = i''$ is trivial. We show the result for $i'' > i'$ assuming it holds for $i'' - 1$. By the induction hypothesis, we have:

$$\mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'')|) > i'' - 1.$$

Assume first that $i'' \notin \mathsf{E}(\sigma, \sigma')$. In this case, by Eq. (2), $\mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'')|) = i''$ is not possible so we must have that $\mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'')|) > i''$. We get:

$$\mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|) \geq \mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'')|) > i''.$$

Now, assume that $i'' \in \mathsf{E}(\sigma, \sigma')$. In this case, we have by Observation 3.1 that:

$$\begin{aligned}\mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|) &= \mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'')| + 1) \\ &\geq \mathsf{Emb}(\sigma, \sigma', i + |\mathsf{E}(\sigma, \sigma') \cap (i', i'')|) + 1 \\ &> i''.\end{aligned}$$

□

Lemma 3.4. *Let $\sigma, \tau, \sigma', \tau' \in \{A, B\}^*$ be given. For all $i' \in \{0\} \cup [\min(|\sigma'|, |\tau'|)]$ such that*

$\sigma'_{\leq i'} = \tau'_{\leq i'}$ *and all* $i \in \{0\} \cup [\min(|\sigma|, |\tau|)]$ *such that* $\sigma_{\leq i} = \tau_{\leq i}$*, we have:*

$$\mathsf{Emb}(\sigma, \sigma', i) \leq i' \implies \mathsf{Emb}(\tau, \tau', i) = \mathsf{Emb}(\sigma, \sigma', i).$$

Moreover, we also have:

$$\mathsf{Emb}(\sigma, \sigma', i) > i' \implies \mathsf{Emb}(\tau, \tau', i) > i'.$$

Proof. Fix $\sigma, \tau, \sigma', \tau'$ and i' as in the lemma statement. For convenience, define $S = |\sigma|$, $T = |\tau|$, $S' = |\sigma'|$, and $T' = |\tau'|$. As one can simply switch the roles of σ, τ and σ', τ', the lemma follows if we show that, for all $i \in \{0\} \cup [\min(S, T)]$ such that $\sigma_{\leq i} = \tau_{\leq i}$, we have:

$$\mathsf{Emb}(\sigma, \sigma', i) \leq i' \implies \mathsf{Emb}(\tau, \tau', i) \leq \mathsf{Emb}(\sigma, \sigma', i).$$

We prove this by induction on i. The base case $i = 0$ is straightforward. We prove the result for $i > 0$ assuming it holds for $i - 1$. Assume that $\mathsf{Emb}(\sigma, \sigma', i) \leq i'$ implying that $\mathsf{Emb}(\sigma, \sigma', i - 1) \leq i'$ by Observation 3.1. We have:

$$\begin{aligned}
\mathsf{Emb}(\tau, \tau', i) &= \min(\{\mathsf{Emb}(\tau, \tau', i - 1) < i'' \leq T' \mid \tau'_{i''} = \tau_i\} \cup \{T' + i\}) && \text{(Eq. (1))} \\
&\leq \min\{\mathsf{Emb}(\tau, \tau', i - 1) < i'' \leq i' \mid \tau'_{i''} = \tau_i\} \\
&\leq \min\{\mathsf{Emb}(\tau, \tau', i - 1) < i'' \leq i' \mid \tau'_{i''} = \sigma_i\} && (\text{As } \sigma_{\leq i} = \tau_{\leq i}) \\
&\leq \min\{\mathsf{Emb}(\tau, \tau', i - 1) < i'' \leq i' \mid \sigma'_{i''} = \sigma_i\}. && (\text{As } \sigma'_{\leq i'} = \tau'_{\leq i'})
\end{aligned}$$

To finish the proof, we show that $\mathsf{Emb}(\sigma, \sigma', i) \in \{\mathsf{Emb}(\tau, \tau', i - 1) < i'' \leq i' \mid \sigma'_{i''} = \sigma_i\}$. Indeed, we have $\mathsf{Emb}(\tau, \tau', i - 1) \leq \mathsf{Emb}(\sigma, \sigma', i - 1)$ by the induction hypothesis implying that $\mathsf{Emb}(\tau, \tau', i - 1) \leq \mathsf{Emb}(\sigma, \sigma', i - 1) < \mathsf{Emb}(\sigma, \sigma', i)$ by Observation 3.1. By assumption, we also have $\mathsf{Emb}(\sigma, \sigma', i) \leq i'$ which together with $i' \leq S'$ and Eq. (1) gives $\sigma'_{\mathsf{Emb}(\sigma, \sigma', i)} = \sigma_i$ as desired. □

Lemma 3.5. *Let* $\sigma, \sigma' \in \{A, B\}^*$ *be given. Additionally, let* $i \in [|\sigma|]$, $i' \in \{0\} \cup [|\sigma'|]$ *be such that* $\mathsf{Emb}(\sigma, \sigma', i - 1) \leq i' < \mathsf{Emb}(\sigma, \sigma', i)$*. For all* $i'' \geq i' \in [|\sigma'|]$ *and all* $0 \leq b \leq |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|$*, we have:*

$$\max(i', \mathsf{Emb}(\sigma, \sigma', b + i - 1)) = \mathsf{Emb}\big(\sigma_{\geq i}, \sigma'_{(i', i'']}, b\big) + i' \leq i''.$$

Proof. Proof by induction on b. The base case $b = 0$ is trivial. We show the lemma for $b > 0$ assuming it holds for $b - 1$. For this, note first that Lemma 3.2 and Observation 3.1 implies that

$$i' < \mathsf{Emb}(\sigma, \sigma', i) \leq \mathsf{Emb}(\sigma, \sigma', b + i - 1) \leq \mathsf{Emb}(\sigma, \sigma', i - 1 + |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|) \leq i'' \leq |\sigma'|.$$

Using this, we derive:

$$\begin{aligned}
&\max(i', \mathsf{Emb}(\sigma, \sigma', b+i-1)) \\
&\qquad = \mathsf{Emb}(\sigma, \sigma', b+i-1) \\
&\qquad = \min\{\mathsf{Emb}(\sigma, \sigma', b+i-2) < j' \le |\sigma'| \mid \sigma'_{j'} = \sigma_{b+i-1}\} \\
&\qquad\qquad\qquad\qquad (\text{Eq. (1) and } \mathsf{Emb}(\sigma, \sigma', b+i-1) \le |\sigma'|) \\
&\qquad = \min\{\max(i', \mathsf{Emb}(\sigma, \sigma', b+i-2)) < j' \le i'' \mid \sigma'_{j'} = \sigma_{b+i-1}\} \\
&\qquad\qquad\qquad\qquad (\text{As } i' < \mathsf{Emb}(\sigma, \sigma', b+i-1) \le i'') \\
&\qquad = \min\{\mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b-1\big) + i' < j' \le i'' \mid \sigma'_{j'} = \sigma_{b+i-1}\} \\
&\qquad\qquad\qquad\qquad (\text{Induction hypothesis}) \\
&\qquad = \min\{\mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b-1\big) < j' \le i'' - i' \mid \sigma'_{i'+j'} = \sigma_{b+i-1}\} + i' \\
&\qquad = \min\Big\{\mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b-1\big) < j' \le i'' - i' \mid \big(\sigma'_{(i',i'']}\big)_{j'} = (\sigma_{\ge i})_b\Big\} + i' \\
&\qquad = \mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b\big) + i'. \qquad (\text{Eq. (1) and } i' < \mathsf{Emb}(\sigma, \sigma', b+i-1) \le i'')
\end{aligned}$$

□

Lemma 3.6. *Let $\sigma, \sigma' \in \{A, B\}^*$ be given. Additionally, let $i \in [|\sigma|]$, $i' \in \{0\} \cup [|\sigma'|]$ be such that $\mathsf{Emb}(\sigma, \sigma', i-1) \le i' < \mathsf{Emb}(\sigma, \sigma', i)$. For all $i'' \ge i' \in [|\sigma'|]$ and all $0 \le a \le b \le |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|$, we have:*

$$\mathsf{Emb}\big(\sigma_{\ge i+a}, \sigma'_{(i',i'']}, b-a\big) \le \mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b\big) \le i'' - i'.$$

Proof. The second inequality follows from Lemma 3.5. We prove the first by induction on $b-a$. The base case $b = a$ is trivial. We show the lemma for $b > a$ assuming it holds for $b-1$. For this, we derive:

$$\begin{aligned}
&\mathsf{Emb}\big(\sigma_{\ge i+a}, \sigma'_{(i',i'']}, b-a\big) \\
&\quad = \min\big(\{\mathsf{Emb}\big(\sigma_{\ge i+a}, \sigma'_{(i',i'']}, b-a-1\big) < j' \le i'' - i' \mid \sigma'_{i'+j'} = \sigma_{b+i-1}\} \cup \{i'' - i' + b - a\}\big) \\
&\qquad\qquad\qquad\qquad (\text{Eq. (1)}) \\
&\quad \le \min\big(\{\mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b-1\big) < j' \le i'' - i' \mid \sigma'_{i'+j'} = \sigma_{b+i-1}\} \cup \{i'' - i' + b - a\}\big) \\
&\qquad\qquad\qquad\qquad (\text{Induction hypothesis}) \\
&\quad \le \min\{\mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b-1\big) < j' \le i'' - i' \mid \sigma'_{i'+j'} = \sigma_{b+i-1}\} \\
&\quad = \mathsf{Emb}\big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b\big). \qquad \Big(\text{As } \mathsf{Emb}\Big(\sigma_{\ge i}, \sigma'_{(i',i'']}, b\Big) \le i'' - i' \text{ by Lemma 3.5}\Big)
\end{aligned}$$

□

Definition 3.7 (Strong Subsequences). *For $\sigma, \sigma' \in \{A, B\}^*$, we say that σ is a strong subsequence of σ' if there exists a set $I \subseteq [|\sigma'|]$ such that $|I| \ge \frac{|\sigma'|}{10}$ and for all $i \in I$ we have that σ is a subsequence of σ'_{-i}.*

Observation 3.8. *For any strings $\sigma, \sigma' \in \{A, B\}^*$, if σ is a strong subsequence of σ', then σ is a subsequence of σ'.*

3.3 Our Noisy Channel

Let Γ be a set with $|\Gamma| \geq 2$. A (deterministic) protocol with the alphabet set Γ is defined by a tuple:

$$\Pi = \left(T, \sigma, \mathcal{X}^A, \mathcal{X}^B, \mathcal{Y}, M_1, \dots, M_T, \mathsf{out}^A, \mathsf{out}^B\right), \tag{3}$$

where: (1) $T > 0$ is a parameter denoting the length of the protocol, (2) $\sigma \in \{A, B\}^T$ is a string that determines which party speaks when (*i.e.*, for all $i \in [T]$, party σ_i is the unique party speaking in round j), (3) $\mathcal{X}^C$ for $C \in \{A, B\}$ is the input set of party C, (4) $\mathcal{Y}$ is the set of possible outputs of the protocol, (5) For all $i \in [T]$, $M_i : \mathcal{X}^{\sigma_i} \times \Gamma^{i-1} \to \Gamma$ is a function that computes the message sent in round i based on the input of the party σ_i speaking in round i and the transcript $\in \Gamma^{i-1}$ received by party σ_i in the first $i-1$ rounds, (6) $\mathsf{out}^C : \Gamma^T \to \mathcal{Y}$ for $C \in \{A, B\}$ are functions that each player uses to compute the output from the transcript of the protocol. We suppress items on the right hand side of Eq. (3) when they are clear from context. We use the notation $\mathsf{spkrs}(\Pi) = \sigma$ and $|\Pi| = T$. We define a randomized protocol $\mathsf{\Pi}$ to be a distribution over deterministic protocols Π that all have the same value of $\left(T, \sigma, \mathcal{X}^A, \mathcal{X}^B, \mathcal{Y}\right)$. We define $\mathsf{spkrs}(\mathsf{\Pi})$ and $|\mathsf{\Pi}|$ to be the common value of $\mathsf{spkrs}(\Pi)$ and $|\Pi|$ respectively.

Execution of a protocol. Let Π be a protocol as above and $\epsilon \geq 0$. We now describe how Π is executed over the channel $\mathsf{C}_{\Gamma,\epsilon}$ that corrupts each sent symbol (independently) to a uniformly random symbol in Γ with probability ϵ. To describe this execution, we let $\star$ be a special symbol not in Γ indicating "no noise" and $N \in (\Gamma \cup \{\star\})^T$ be a noise vector such that for all $i \in [T]$, the symbol $N_i = \star$ with probability $1 - \epsilon$ and a uniformly random symbol from Γ with probability ϵ (independently for all i), so that N captures the noise inserted by the aforementioned channel. We shall abuse notation and use $\mathsf{C}_{\Gamma,\epsilon}$ to denote both the channel and the above distribution over noise vectors.

The execution begins with both parties $C \in \{A, B\}$ having input $x^C \in \mathcal{X}^C$ and proceeds in T rounds, maintaining the invariant that before round $i \in [T]$, both parties $C \in \{A, B\}$ have a partial transcript $\Pi^C_{<i} \in \Gamma^{i-1}$. In round i, party σ_i computes the symbol $\gamma_i = M_i(x^{\sigma_i}, \Pi^{\sigma_i}_{<i})$, appends it to its own partial transcript, and sends it over the channel to the other party $\overline{\sigma}_i$.

The noise N then acts on the symbol as follows: If $N_i = \star$, then the symbol is sent uncorrupted and party $\overline{\sigma}_i$ receives the symbol γ_i. Otherwise, we have $N_i \in \Gamma$ and party $\overline{\sigma}_i$ receives the symbol N_i. In either case, party $\overline{\sigma}_i$ appends the received symbol to its partial transcript and the execution proceeds to the next round.

After T rounds are over, each party $C \in \{A, B\}$ outputs $\mathsf{out}^C\left(\Pi^C_{\leq T}\right) \in \mathcal{Y}$. Note that this execution is entirely determined by the triple $\left(x^A, x^B, N\right)$, which we shall often write as (X, N) using X to denote the pair of inputs $\left(x^A, x^B\right)$. This fact allows us to write $\Pi_i^C(X, N)$, $\Pi_{\leq i}^C(X, N)$, *etc.* to denote the corresponding value in the execution of Π in the presence of noise N when the inputs are X. For $C \in \{A, B\}$, we also define the notation $\mathsf{res}_\Pi^C(X, N)$ to denote the output of party C in the above execution and $\mathsf{res}_\Pi(X, N) = \left(\mathsf{res}_\Pi^A(X, N), \mathsf{res}_\Pi^B(X, N)\right)$. We omit N from the above notations when $\epsilon = 0$ and the execution is noiseless, as in this case, N is always the vector with all coordinates equal to $\star$. Note that, in this case, the transcripts for Alice and Bob are the same and we can omit the superscript C in the notation.

Simulations and hole simulations. Let Γ be an alphabet set as above. Let $\mathsf{\Pi}$ and $\mathsf{\Pi}'$ be two randomized protocols with alphabet Γ and with the same input sets $\mathcal{X}^A, \mathcal{X}^B$ for Alice and Bob. For $p \in [0, 1]$ and $\epsilon \geq 0$, we say the protocol Π' simulates the protocol Π over the channel $\mathsf{C}_{\Gamma,\epsilon}$ with probability p if for all $x^A \in \mathcal{X}^A, x^B \in \mathcal{X}^B$, it holds that

$$\Pr_{N \sim \mathsf{C}_{\Gamma,\epsilon}, \Pi \sim \mathsf{\Pi}, \Pi' \sim \mathsf{\Pi}'}\left(\mathsf{res}_{\Pi'}(X, N) = \mathsf{res}_\Pi(X)\right) \geq p.$$

Throughout this text, the protocol Π being simulated will be deterministic and we shall omit it from the subscript above. As our main result in a lower bound, the fact that Π is deterministic only makes our result stronger. As Γ is determined by Π, we shall sometimes omit writing "over the channel $\mathsf{C}_{\Gamma,\epsilon}$" when $\epsilon = 0$.

For our proof of Theorem 1.1, we actually work with a different (and weaker[7]) notion of simulation that we call "hole simulation" and is defined as follows: Let $\sigma' \in \{A, B\}^*$ and Π be a protocol as above. For $p \in [0, 1]$, we say that σ' hole-simulates Π with probability p if there exists a set $I' \subseteq [|\sigma'|]$, $|I'| \geq \frac{|\sigma'|}{10}$ such that for all $i' \in I'$, there exists a randomized protocol $\Pi'_{i'}$ with alphabet Γ and the same input sets as Π that simulates the protocol Π (over the channel $\mathsf{C}_{\Gamma,0}$) with probability p and satisfies $\mathsf{spkrs}(\Pi'_{i'}) = \sigma'_{-i'}$.

3.4 Pointer Chasing

Let $m, T \in \mathbb{N}$. We inductively define the function $\mathsf{PC}_{m,T}$ that takes as input functions $(f_i)_{i\in[T]}$ where $f_i : [m]^{i-1} \to [m]$ for all $i \in [T]$ and outputs a value in $[m]^T$ as follows: For the case $T = 1$, we simply define $\mathsf{PC}_{m,1}(f_1) = f_1$. For $T > 1$ and functions $(f_i)_{i\in[T]}$, let $z = \mathsf{PC}_{m,T-1}((f_i)_{i<T})$ be the value defined by the induction hypothesis, and define $\mathsf{PC}_{m,T}((f_i)_{i\in[T]}) = z||f_T(z)$. We omit the parameters m, T when they are clear from context. It is clear from the above definition that for all $T' \in [T]$, the value of $\mathsf{PC}((f_i)_{i\in[T]})$ is independent of $(f_i)_{i\in[T']}$ as long as $\mathsf{PC}((f_i)_{i\in[T']})$ is the same. Correspondingly, for $z \in [m]^{T'}$, we sometimes write $\mathsf{PC}(z, (f_i)_{i\in(T',T]})$ to denote the value of $\mathsf{PC}((f_i)_{i\in[T]})$ when $\mathsf{PC}((f_i)_{i\in[T']}) = z$.

[7] The fact that we work with a weaker notion makes are proof stronger, and in particular, would also work for the erasure channel. To see the formal sense in which this is weaker, see Section 4.

Pointer chasing protocols. Let $T \in \mathbb{N}$ and $\sigma \in \{A, B\}^T$. Define $m = (200T)^{200}$ and the protocol $\mathtt{PC}_\sigma$ to be the T-round communication protocol with alphabet $[m]$ where Alice's input are functions $(f_i : [m]^{i-1} \to [m])_{i:\sigma_i = A}$ and Bob's input are functions $(f_i : [m]^{i-1} \to [m])_{i:\sigma_i = B}$, and the message sent in round t, for $t \in [T]$ is coordinate t of $\mathtt{PC}_{m,T}((f_i)_{i\in[T]})$. After T rounds, the parties output all of $\mathtt{PC}_{m,T}((f_i)_{i\in[T]})$.

4 Proof of Theorem 1.1

The goal is section is to prove Theorem 1.1 assuming two other theorems that we shall prove in the following sections. We start by stating these two theorems.

Theorem 4.1. *Let $\sigma, \sigma' \in \{A, B\}^*$ be given. Assume that $|\sigma'| \leq 5 \cdot |\sigma|$. If σ' hole simulates $\mathtt{PC}_\sigma$ with probability $\frac{1}{5}$, then σ is a strong subsequence of σ'.*

Theorem 4.2. *For all $T > 0$, there exists $\sigma \in \{A, B\}^T$ such that for all $\sigma' \in \{A, B\}^*$ such that σ is a strong subsequence of σ', we have $|\sigma'| \geq (1 + 10^{-100}) \cdot |\sigma|$.*

We are now ready to prove Theorem 1.1 (assuming Theorems 4.1 and 4.2).

Proof of Theorem 1.1. Fix $\epsilon > 0$ and assume that $\epsilon < 0.001$ without loss of generality. Define $T = \frac{1}{\epsilon^2}$ and let σ be as promised by Theorem 4.2. Define $\Gamma = \left[(200T)^{200}\right]$ and $\Pi = \mathtt{PC}_\sigma$. Let $\mathsf{\Pi}'$ be a randomized protocol that simulates Π with over the channel $\mathsf{C}_{\Gamma,\epsilon}$ with probability 0.99 and $\sigma' = \mathsf{spkrs}(\Pi)$. As the proof is trivial otherwise, assume that $|\sigma'| = |\mathsf{\Pi}'| \leq 5T$. We claim that σ' hole simulates $\mathtt{PC}_\sigma$ with probability 0.5. This finishes the proof as it implies using Theorem 4.1 that σ is a strong subsequence of σ' which using Theorem 4.2 means that $|\sigma'| \geq (1 + 10^{-100}) \cdot T$, as desired.

It remains to show the claim. To this end, let $T' = |\sigma'|$ and for $i' \in [T']$, define the protocol $\mathsf{\Pi}'_{i'}$ with $\mathsf{spkrs}(\mathsf{\Pi}'_{i'}) = \sigma'_{-i'}$ as follows: For $t \in \{0\} \cup [T']$, let $p_t = \binom{T'}{t} \cdot \epsilon^t (1 - \epsilon)^{T'-t}$ be the probability that the channel $\mathsf{C}_{\Gamma,\epsilon}$ corrupts exactly t symbols in $\mathsf{\Pi}'$. This means that $\frac{p_t}{1-p_0}$ is the probability the channel $\mathsf{C}_{\Gamma,\epsilon}$ corrupts exactly t symbols in $\mathsf{\Pi}'$ conditioned on it corrupting at least one symbol. Then, the protocol $\mathsf{\Pi}'_{i'}$ is exactly the same as $\mathsf{\Pi}'$ except that it (1) It does not have round i' and the party supposed to receive in this round assumes it got a uniformly random symbol in Γ. Observe that this can equivalently be seen as the channel always corrupting round i' in $\mathsf{\Pi}'$. (2) Samples $t' \in [T']$ with probability $\frac{p_{t'}}{1-p_0}$, and then artificially corrupts $t' - 1$ rounds, ignoring the bit actually received in these rounds and using a uniformly random symbol in Γ instead.

Observe that picking $i' \in [T']$ uniformly at random and running $\mathsf{\Pi}'_{i'}$ over the noiseless channel $\mathsf{C}_{\Gamma,0}$ is the same as running $\mathsf{\Pi}'$ over $\mathsf{C}_{\Gamma,\epsilon}$ and conditioning on the fact that the channel corrupts at least one symbol. As $\mathsf{\Pi}'$ simulates Π with over the channel $\mathsf{C}_{\Gamma,\epsilon}$ with probability 0.99, we get:

$$\frac{1}{T'} \cdot \sum_{i'=1}^{T'} \Pr_{\Pi'_{i'} \sim \mathsf{\Pi}'_{i'}} \Big(\mathsf{res}_{\Pi'_{i'}}(X) = \mathsf{res}_{\Pi}(X) \Big) = \Pr_{N \sim \mathsf{C}_{\Gamma,\epsilon}, \Pi' \sim \mathsf{\Pi}'} \big(\mathsf{res}_{\Pi'}(X, N) = \mathsf{res}_{\Pi}(X) \mid N \neq \star^T \big)$$

$$
\begin{aligned}
&\geq \Pr_{N\sim\mathsf{C}_{\Gamma,\epsilon},\Pi'\sim\mathsf{\Pi}'}(\mathsf{res}_{\Pi'}(X,N)=\mathsf{res}_{\Pi}(X)) \\
&\qquad\qquad - \Pr_{N\sim\mathsf{C}_{\Gamma,\epsilon}}\left(N=\star^{T}\right) \\
&\geq 0.9.
\end{aligned}
$$

It follows that $\Pr_{\Pi'_{i'}\sim\mathsf{\Pi}'_{i'}}\left(\mathsf{res}_{\Pi'_{i'}}(X)=\mathsf{res}_{\Pi}(X)\right)\geq 0.5$ for at least $\frac{T'}{10}$ values of i', as desired.

□

5 Proof of Theorem 4.1

The goal of this section is to show Theorem 4.1. Owing to the definition of hole simulations and Definition 3.7, it suffices to show the following lemma:

Lemma 5.1. *Let $\tau\in\{A,B\}^*$, $m=(200\cdot|\tau|)^{200}$, and $\mathsf{\Pi}$ be a randomized protocol with alphabet $[m]$. Assume that $|\mathsf{\Pi}|\leq 5\cdot|\tau|$. If $\mathsf{\Pi}$ simulates $\mathtt{PC}_\tau$ with probability $\frac{1}{|\tau|}$, then τ is a subsequence of $\mathsf{spkrs}(\mathsf{\Pi})$.*

We prove Lemma 5.1 in the rest of this section. Fix $\tau,\mathsf{\Pi}$ and define $n=|\tau|$, $T=|\mathsf{\Pi}|$, and $\sigma=\mathsf{spkrs}(\mathsf{\Pi})$. We shall show the lemma in the contrapositive, assuming that τ is a not subsequence of σ and showing that $\mathsf{\Pi}$ does not simulate $\mathtt{PC}_\tau$ with probability $\frac{1}{n}$. As τ,σ are fixed we shall often omit them from our notation and write $\mathsf{Emb}(\cdot)$ instead of $\mathsf{Emb}(\tau,\sigma,\cdot)$ and E instead of $\mathsf{E}(\tau,\sigma)$.

Let $\mathcal{X}^A$ and $\mathcal{X}^B$ be input sets of Alice and Bob respectively in $\mathtt{PC}_\tau$ (and therefore also in $\mathsf{\Pi}$). Recall from Section 3.3 that we have to show that there exist $x^A\in\mathcal{X}^A,x^B\in\mathcal{X}^B$ such that:

$$\Pr_{\Pi\sim\mathsf{\Pi}}(\mathsf{res}_{\Pi}(X)=\mathsf{res}_{\mathtt{PC}_\tau}(X))<\frac{1}{n}.$$

Let $\mathcal{F}$ be the uniform distribution over all inputs of $\mathtt{PC}_\tau$, defined as in Section 3.4. To show the foregoing equation, we fix an arbitrary deterministic protocol Π in the support of $\mathsf{\Pi}$ and show that (noting that $\mathsf{res}_{\mathtt{PC}_\tau}(F)=\mathtt{PC}(F)$):

$$\Pr_{F\sim\mathcal{F}}(\mathsf{res}_{\Pi}(F)=\mathtt{PC}(F))<\frac{1}{n}. \tag{4}$$

5.1 Notation

For a finite non-empty set S, we shall use $\mathcal{U}(S)$ to denote the uniform distribution over S. We omit S from the notation when it is clear from the context. All probabilities and random variables will be defined over the randomness in $\mathcal{F}$, and we will often abbreviate $\Pr_{F\sim\mathcal{F}}$ to $\Pr$ for brevity of notation. Throughout, if X is a random variable and x is a value that X can take, we sometimes abbreviate the event $\mathsf{X}=x$ as simply x when it is clear from context. Thus, we may write $\Pr(x)$ instead of $\Pr(\mathsf{X}=x)$ and $\Pr(\cdot\mid x)$ instead of $\Pr(\cdot\mid\mathsf{X}=x)$. We use $\mathsf{dist}(\mathsf{X})$ to denote the distribution of a random variable X.

We will use F to denote the random variable corresponding to a sample from $\mathcal{F}$ and F to denote a given value of F. Observe that F is an n-tuple $(f_1, f_2, \cdots, f_n)$. For a set $S \subseteq [n]$, we define $f_S = (f_i)_{i \in S}$. For $i \in [n]$, we may write $f_{\leq i}$ instead of $f_{[i]}$ and $f_{<i}$ instead of $f_{[i-1]}$. We also define $f^A = f_{\{i \in [n] \mid \tau_i = A\}}$ and $f^B = f_{\{i \in [n] \mid \tau_i = B\}}$. We may combine these notations and use $f^A_{\leq i} = f_{\{i' \in [i] \mid \tau_{i'} = A\}}$, *etc.*. We will use $\mathsf{f}_{\leq i}$ to denote the random variable corresponding to $f_{\leq i}$. The notations f_S, $\mathsf{f}_{<i}$, f^A, *etc.* are defined similarly.

Recall the functions $\Pi_t(\cdot)$ and $\Pi_{\leq t}(\cdot)$ from Section 3.3. In this section, we extend this notation to sets $S \subseteq [T]$ by defining $\Pi_S(\cdot) = (\Pi_t(\cdot))_{t \in S}$. For $t \in [T]$, we will use $\mathsf{\Pi}_t = \Pi_t(\mathsf{F})$ to denote the random variable obtained by sampling F and outputting $\Pi_t(\mathsf{F})$, and use Π_t to denote a value $\mathsf{\Pi}_t$ can take. The notations $\mathsf{\Pi}_{\leq t}$, $\mathsf{\Pi}_S$ are defined analogously.

5.2 Definitions

Recall that we fixed $\tau \in \{A, B\}^n$ as the order in which the parties speak in the protocol $\mathtt{PC}_\tau$ being simulated. We also fixed a deterministic protocol Π and defined $T = |\Pi|$ and $\sigma = \mathsf{spkrs}(\Pi) \in \{A, B\}^T$.

The set $\mathcal{L}$. We consider the set of indices $i \in [n]$ where the value of τ_i is different from τ_{i+1}. Define the set:

$$\mathcal{L} = \{n\} \cup \{i \in [n-1] \mid \tau_i \neq \tau_{i+1}\}. \tag{5}$$

Informally, $\mathcal{L}$ is the rounds where τ "switches" from A to B or B to A. Equivalently, we partition τ into consecutive *intervals* consisting of the same player and $\mathcal{L}$ is the set of endpoints of these intervals. The element n is added to $\mathcal{L}$ for convenience. For $i \in [n]$, we define $\ell^{\geq}_i$ to be smallest value $\ell \in \mathcal{L}$ satisfying $\ell \geq i$. This is well defined as $n \in \mathcal{L}$ is one such value. Similarly, define $\ell^{<}_i$ to be largest value $\ell \in \{0\} \cup \mathcal{L}$ satisfying $\ell < i$. Observe that for all $i \in [n]$, we have $\tau_i = \tau_{\ell^{\geq}_i}$ and $\ell^{<}_i > 0$ implies $\tau_i \neq \tau_{\ell^{<}_i}$ (so $\tau_i = \overline{\tau}_{\ell^{<}_i}$).

Defining Good and Rem. For $t \in \{0\} \cup [T]$, define the set

$$\mathsf{Good}(t) = \left\{ i \in [n] : \mathsf{Emb}(\ell^{<}_i) \leq t < \mathsf{Emb}\left(\ell^{\geq}_i\right) \right\}. \tag{6}$$

Informally, Good gets a round t of the protocol Π and outputs the first interval of τ that we do not expect to have fully simulated after round t. Observe that $\mathsf{Good}(t) \neq \emptyset$ for all $t \in \{0\} \cup [T]$. Let $t \in \{0\} \cup [T]$ and $i \in \mathsf{Good}(t)$. Define:

$$\mathsf{Rem}^i(t) = \begin{cases} (i - \ell^{<}_i - 0.1) \cdot \log m, & \text{if } \mathsf{Emb}(\ell^{<}_i) = t \\ (i - \ell^{<}_i - 0.3 - |\mathsf{E} \cap (\mathsf{Emb}(\ell^{<}_i), t]|) \cdot \log m, & \text{if } \mathsf{Emb}(\ell^{<}_i) < t \end{cases}. \tag{7}$$

Roughly speaking, $\mathsf{Rem}^i(t)$ is the amount of (min-)entropy remaining in the random variable $\mathtt{PC}_{>\ell^{<}_i}(\mathsf{f}_{\leq i})$ before round t of the protocol.

“Revealing” information. To make our analysis cleaner, we reveal some information to the players at various points in the protocol. More precisely, let $t' \in [3T]$ be given and F in the support[8] of F be arbitrary. Let $t \in [T]$ be the unique value satisfying $3(t-1) < t' \le 3t$. We shall define values $\Phi_{t'}(F)$ inductively. If $t' = 3t - 2$, define:

$$\Phi_{t'}(F) = \Pi_t(F). \tag{8}$$

If $t' = 3t - 1$, define:

$$\Phi_{t'}(F) = \begin{cases} \mathsf{PC}_{>\ell_i^<}\left(f_{\le \ell_i^\ge}\right), & \text{if } \exists i \in [n] : t = \mathsf{Emb}\left(\ell_i^\ge\right) \\ 0, & \text{otherwise} \end{cases}. \tag{9}$$

Informally, this definition amounts to revealing the correct transcript for any interval at the end of the interval. Finally, $t' = 3t$, define:

$$\Phi_{t'}(F) = \mathbb{1}\Big(\exists i \in [n] : t = \mathsf{Emb}(\ell_i^<), \Pr\Big(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) = \mathsf{PC}_{>\ell_i^<}(f_{\le i}) \mid \Phi_{<t'}(F)\Big) > 2^{-1-\mathsf{Rem}^i(t)}\Big). \tag{10}$$

Informally, this definition amounts to revealing, at the end of an interval, whether the right answer for the next interval can be guessed with probability much better than what Rem would indicate. We will later show that, this answer is no ($= 0$) with high probability. Henceforth, we treat Φ the same way as Π in our notation, *i.e.*, we let $\mathbf{\Phi}_{t'} = \Phi_{t'}(\mathsf{F})$ denote the random variable obtained by sampling F and outputting $\Phi_{t'}(\mathsf{F})$, use $\Phi_{t'}$ to denote a value $\mathbf{\Phi}_{t'}$ can take, and define $\mathbf{\Phi}_{\le t'}$, $\mathbf{\Phi}_S$ *etc.* analogously to $\mathbf{\Pi}_{\le t}$, $\mathbf{\Pi}_S$, *etc.*

We claim that Φ provides all the necessary information in order to reconstruct $\mathsf{PC}(\mathsf{f})$, as claimed below.

Claim 5.2. *For any $\ell \in \mathcal{L}$ satisfying $\mathsf{Emb}(\ell) \le T$, the value of $\Phi_{\le 3\mathsf{Emb}(\ell)-1}$ fixes the value of $\mathsf{PC}(\mathsf{f}_{\le \ell})$.*

Proof. Let $u = \mathsf{Emb}(\ell)$. We prove this by an inductive argument on elements of $\mathcal{L}$.

To begin with, let ℓ be the smallest element of $\mathcal{L}$. Note that $\ell_\ell^< = 0$. By Eq. (9), fixing $\Phi_{\le 3u-1}$ fixes Φ_{3u-1}, which fixes thus fixes

$$\mathsf{PC}_{>\ell_\ell^<}\left(\mathsf{f}_{\le \ell_\ell^\ge}\right) = \mathsf{PC}(\mathsf{f}_{\le \ell}).$$

Otherwise, suppose that this claim holds for all $\ell' \in \mathcal{L}$ such that $\ell' < \ell$. Therefore, in particular, it holds for $\ell' = \ell_\ell^<$, the largest element of $\mathcal{L}$ smaller than ℓ. Note that $\ell' < \ell$, so $\mathsf{Emb}(\ell') < \mathsf{Emb}(\ell)$ by Observation 3.1. Thus, fixing $\Phi_{\le 3u-1} = \Phi_{\le 3\mathsf{Emb}(\ell)-1}$ also fixes $\Phi_{\le 3\mathsf{Emb}(\ell')-1}$. Thus, by our inductive hypothesis, we get that $\mathsf{PC}(\mathsf{f}_{\le \ell'}) = \mathsf{PC}_{\le \ell'}(\mathsf{f}_{\le \ell})$ is fixed.

Now, note that as $\ell = \ell_\ell^\ge$, that means that fixing $\Phi_{\le 3u-1}$ in particular fixes Φ_{3u-1}, which in turn fixes $\mathsf{PC}_{>\ell_\ell^<}(\mathsf{f}_{\le \ell}) = \mathsf{PC}_{>\ell'}(\mathsf{f}_{\le \ell})$ by Eq. (9). Combining this with the above thus fixes

[8] Henceforth, we omit writing “in the support of” when the random variable is clear from context.

all of $\mathtt{PC}(\mathsf{f}_{\leq \ell})$. □

We also claim that Φ is a transcript of a protocol. By this, we mean that each coordinate of Φ can be computed fully by just one player using only their input and the transcript Φ so far. Formally:

Lemma 5.3. *For $t' \in [3T]$, there exists $\sigma'_{t'} \in \{A, B\}$ and a function $M'_{t'}$ such that for any F,*

$$\Phi_{t'}(F) = M'_{t'}\Big(f^{\sigma'_{t'}}, \Phi_{<t'}(F)\Big).$$

Furthermore, for $t \in [T]$, we have that:

$$\begin{aligned} \sigma'_{3t-2} &= \sigma_t \\ \sigma'_{3t-1} &= \sigma_t \\ \sigma'_{3t} &= \overline{\sigma}_t. \end{aligned}$$

Proof. When $t' = 3t - 2$, as $\Phi_{t'}(F) = \Pi_t(F)$, the result follows directly from the definition of $\Pi_t(F)$, with $\sigma'_{t'} = \sigma_t$.

When $t' = 3t - 1$, if there does not exist $i \in [n]$ such that $t = \mathsf{Emb}\big(\ell_i^{\geq}\big)$, then $\Phi_{t'}(F) = 0$ is constant and independent of both f^A and f^B, so $\sigma'_{t'}$ can be arbitrary and $M'_{t'}$ can be the constant 0 function. Otherwise, suppose that there exists such an i, and let $\sigma'_{t'} = \sigma_t$. Note that $\Phi_{t'}(F) = \mathtt{PC}_{>\ell_i^<}\Big(f_{\leq \ell_i^{\geq}}\Big) = \mathtt{PC}_{>\ell_i^<}\Big(\mathtt{PC}\Big(f_{\leq \ell_i^<}\Big), (f_{i'})_{i' \in (\ell_i^<, \ell_i^{\geq}]}\Big)$. By Claim 5.2, $\mathtt{PC}\Big(f_{\leq \ell_i^<}\Big)$ is determined by $\Phi_{<t'}(F)$. Furthermore, for all $i' \in (\ell_i^<, \ell_i^{\geq}]$, $\tau_{i'} = \tau_{\ell_i^{\geq}} = \sigma_t = \sigma'_{t'}$, so $(f_{i'})_{i' \in (\ell_i^<, \ell_i^{\geq}]}$ is determined by the input $f^{\sigma'_{t'}}$. The result follows.

When $t' = 3t$, if there does not exist $i \in [n]$ such that $t = \mathsf{Emb}(\ell_i^<)$, then $\Phi_{t'}(F) = 0$ is constant and independent of both f^A and f^B, so $\sigma'_{t'}$ can be arbitrary and $M'_{t'}$ can be the constant 0 function. Otherwise, suppose that there exists such an i and let $\sigma'_{t'} = \overline{\sigma}_t$. Note that $\Phi_{t'}(F)$ can be computed from $\Phi_{<t'}(F)$ and $\mathtt{PC}_{>\ell_i^<}(f_{\leq i})$, so it suffices to show that the latter can be computed from $\Phi_{<t'}(F)$ and $f^{\sigma'_{t'}}$. Note that $\mathtt{PC}_{>\ell_i^<}(f_{\leq i}) = \mathtt{PC}_{>\ell_i^<}\Big(\mathtt{PC}\Big(f_{\leq \ell_i^<}\Big), (f_{i'})_{i' \in (\ell_i^<, i]}\Big)$. By Claim 5.2, $\mathtt{PC}\Big(f_{\leq \ell_i^<}\Big)$ is determined by $\Phi_{<t'}(F)$. Furthermore, for all $i \in [n]$ such that $t = \mathsf{Emb}(\ell_i^<)$, $\tau_i = \overline{\tau}_{\ell_i^<} = \overline{\sigma}_t = \sigma'_{t'}$, so $(f_{i'})_{i' \in (\ell_i^<, i]}$ is determinted by the input $f^{\sigma'_{t'}}$. □

The sets Guess and Info. We are now ready to define the sets Guess and Info, the primary focus of our analysis. For $t \in \{0\} \cup [T]$ and $i \in \mathsf{Good}(t)$, we define:

$$\mathsf{Guess}^i(t) = \Big\{\Phi_{\leq 3t} \mid \mathbb{H}_\infty\Big(\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i}) \mid \Phi_{\leq 3t}\Big) < \mathsf{Rem}^i(t)\Big\}. \tag{11}$$

Informally, this is the set of transcripts that allow us to guess the edges in the current interval (until i) with probability better than that indicated by Rem. For $t \in \{0\} \cup [T]$ and

$C \in \{A, B\}$, define:

$$\mathsf{Info}^C(t) = \left\{\Phi_{\leq 3t} \mid \mathbb{D}\big(\mathsf{dist}\big(\mathsf{f}^C \mid \Phi_{\leq 3t}\big) \mid\mid \mathcal{U}\big) > m^{0.01}\right\}. \tag{12}$$

Informally, this is the set of transcripts that give a lot of information about party C's input.

5.3 Properties of Info

This section is dedicated to proving Lemma 5.4, which will be key to proving our main result. Roughly, Lemma 5.4 says that transcripts are unlikely to be informative enough to be in the set Info as that requires information $\geq m^{0.01}$ which is much more than the communication (as m is larger than the communication to the power of 200).

Lemma 5.4. *For all $t \in \{0\} \cup [T]$ and $C \in \{A, B\}$,*

$$\Pr\big(\Phi_{\leq 3t} \in \mathsf{Info}^C(t)\big) \leq \frac{1}{m^{0.5}}.$$

For all $t' \in [3T]$, let $\sigma'_{t'}$ and $M'_{t'}$ be as in Lemma 5.3. We define for all $C \in \{A, B\}$, for all $t' \in \{0\} \cup [3T]$, for all $\Phi_{\leq t'}$, the set:

$$\mathsf{Rec}^C(\Phi_{\leq t'}) = \left\{f^C \mid \ \forall t'' \in [t'] \text{ s.t. } \sigma'_{t''} = C, \text{ we have } \Phi_{t''} = M'_{t''}\big(f^C, \Phi_{<t''}\big)\right\}. \tag{13}$$

Roughly, our definition of Φ ensures that the pairs of inputs that lead to the transcript $\Phi_{\leq t'}$ form a combinatorial rectangle (Rec denotes rectangle), and Rec^C denotes the projection of this rectangle on party C's inputs. In other words, Rec^C is the set of all inputs of party C that may lead to the transcript $\Phi_{\leq t'}$.

Observation 5.5. *For all $t' \in [3T]$, for all $\Phi_{\leq t'}$, for all $C \in \{A, B\}$, if $\sigma'_{t'} \neq C$, $\mathsf{Rec}^C(\Phi_{\leq t'}) = \mathsf{Rec}^C(\Phi_{<t'})$.*

We now show several properties of Rec^C.

Lemma 5.6. *For all $t' \in \{0\} \cup [3T]$, for all $\Phi_{\leq t'}$, the event $\big(\forall C \in \{A, B\} : \mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{\leq t'})\big)$ and the event $\Phi_{\leq t'}$ are equivalent.*

Proof. We prove this by induction on t'. The result is obvious for $t' = 0$. As such, we will show that it holds for some $t' > 0$ given that it holds for $t' - 1$. We have that the following events are equivalent:

$$\begin{aligned}
\Phi_{\leq t'} &\equiv (\Phi_{<t'}, \Phi_{t'}) \\
&\equiv \left(\Phi_{<t'}, M'_{t'}\left(\mathsf{f}^{\sigma'_{t'}}, \Phi_{<t'}\right) = \Phi_{t'}\right) && \text{(Lemma 5.3)} \\
&\equiv \left(\forall C \in \{A, B\} : \mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{<t'}), M'_{t'}\left(\mathsf{f}^{\sigma'_{t'}}, \Phi_{<t'}\right) = \Phi_{t'}\right) && \text{(Induction Hypothesis)} \\
&\equiv \big(\forall C \in \{A, B\} : \mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{\leq t'})\big). && \text{(Eq. (13) and Observation 5.5)}
\end{aligned}$$

□

Lemma 5.7. *For all $t' \in \{0\} \cup [3T]$, for all $\Phi_{\le t'}$,*

$$\Pr(\Phi_{t'} \mid \Phi_{<t'}) = \frac{\left|\mathsf{Rec}^{\sigma'_{t'}}(\Phi_{\le t'})\right|}{\left|\mathsf{Rec}^{\sigma'_{t'}}(\Phi_{< t'})\right|}.$$

Proof. We have that:

$$\begin{aligned}
\Pr(\Phi_{t'} \mid \Phi_{<t'}) &= \frac{\Pr(\Phi_{\le t'})}{\Pr(\Phi_{<t'})} \\
&= \frac{\Pr\big(\forall C \in \{A,B\}, \mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{\le t'})\big)}{\Pr\big(\forall C \in \{A,B\}, \mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{< t'})\big)} && \text{(Lemma 5.6)} \\
&= \frac{\prod_{C\in\{A,B\}} \Pr\big(\mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{\le t'})\big)}{\prod_{C\in\{A,B\}} \Pr\big(\mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{< t'})\big)} && \text{(Independence of } \mathsf{f}^A \text{ and } \mathsf{f}^B\text{)} \\
&= \frac{\Pr\Big(\mathsf{f}^{\sigma'_{t'}} \in \mathsf{Rec}^{\sigma'_{t'}}(\Phi_{\le t'})\Big)}{\Pr\Big(\mathsf{f}^{\sigma'_{t'}} \in \mathsf{Rec}^{\sigma'_{t'}}(\Phi_{< t'})\Big)} \\
&= \frac{\left|\mathsf{Rec}^{\sigma'_{t'}}(\Phi_{\le t'})\right|}{\left|\mathsf{Rec}^{\sigma'_{t'}}(\Phi_{< t'})\right|}. && (\mathsf{f}^{\sigma'_{t'}} \text{ is uniform})
\end{aligned}$$

□

Lemma 5.8. *For all $t' \in \{0\} \cup [3T], C \in \{A, B\}$,*

$$\Pr\left(\frac{\left|\mathsf{Rec}^C(\Phi_{\le t'})\right|}{|\mathsf{supp}(\mathsf{f}^C)|} < \frac{1}{m^{2t'}}\right) \le \frac{t'}{m^{0.6}}.$$

Proof. We will prove this by proving the following stronger bound. Let

$$\mathsf{P}(t') = \left\{ i \in \mathsf{Good}\big(\left\lceil \tfrac{t'-1}{3} \right\rceil\big) \;\middle|\; \mathsf{Emb}(i) \le \left\lceil \tfrac{t'-1}{3} \right\rceil \right\}.$$

Then, we claim that:

$$\Pr\left(\frac{\left|\mathsf{Rec}^C(\Phi_{\le t'})\right|}{|\mathsf{supp}(\mathsf{f}^C)|} < \frac{1}{m^{2(t'-|\mathsf{P}(t')|)}}\right) \le \frac{t'}{m^{0.6}}.$$

We will prove this by induction over t'. For $t' = 0$, this holds since $\mathsf{Rec}^C(\Phi_{\le t'}) = \mathsf{supp}\big(\mathsf{f}^C\big)$

by Eq. (13). We show the result holds for $t' > 0$, assuming it holds for $t' - 1$. We have that:

$$\begin{aligned}
&\Pr\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{|\mathsf{supp}(\mathsf{f}^C)|} < \frac{1}{m^{2(t' - |\mathsf{P}(t')|)}}\right) \\
&\le \Pr\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|}{|\mathsf{supp}(\mathsf{f}^C)|} < \frac{1}{m^{2(t'-1-|\mathsf{P}(t'-1)|)}}\right) + \Pr\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|} < \frac{1}{m^{2(1+|\mathsf{P}(t'-1)|-|\mathsf{P}(t')|)}}\right) \\
&\hfill \text{(Union Bound, as } ab < \alpha\beta \text{ implies } a < \alpha \text{ or } b < \beta \text{ for positive values)} \\
&\le \frac{(t'-1)}{m^{0.6}} + \Pr\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|} < \frac{1}{m^{2(1+|\mathsf{P}(t'-1)|-|\mathsf{P}(t')|)}}\right). \qquad \text{(Inductive Hypothesis)}
\end{aligned}$$

We now have to consider several cases. First, suppose that $t' \neq 3t - 1$ for any $t \in [T]$. Then, note that $\mathsf{P}(t'-1) = \mathsf{P}(t')$, so it suffices to show the bound $\Pr\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|} < \frac{1}{m^2}\right) \le \frac{1}{m^{0.6}}$. If $C \neq \sigma'_{t'}$, then the result holds by Observation 5.5. As such, it just suffices to show the case where $C = \sigma'_{t'}$. We will show this conditioned on an arbitrary $\Phi_{<t'}$. We have:

$$\begin{aligned}
\Pr\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|} < \frac{1}{m^2} \mid \Phi_{<t'}\right) &= \sum_{\Phi_{t'}} \Pr(\Phi_{t'} \mid \Phi_{<t'}) \cdot \mathbb{1}\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|} < \frac{1}{m^2}\right) \\
&= \sum_{\Phi_{t'}} \Pr(\Phi_{t'} \mid \Phi_{<t'}) \cdot \mathbb{1}\left(\Pr(\Phi_{t'} \mid \Phi_{<t'}) < \frac{1}{m^2}\right) \\
&\hfill \text{(Lemma 5.7)} \\
&\le \frac{1}{m^{0.6}}.
\end{aligned}$$

Now, let us consider the case where $t' = 3t - 1$ for some $t \in [T]$. This means that $t = \left\lceil \frac{t'-1}{3} \right\rceil$ and $t - 1 = \left\lceil \frac{t'-2}{3} \right\rceil$. Once again, if $C \neq \sigma'_{t'}$, then the result holds by Observation 5.5. Furthermore, if $t \neq \mathsf{Emb}(\ell_i^{\ge})$ for any $i \in [n]$, then $\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'}) = \mathsf{Rec}^C(\mathsf{\Phi}_{< t'})$ by Eqs. (9) and (13), and the result holds. As such, it just suffices to show the case where $C = \sigma'_{t'}$ and $t = \mathsf{Emb}(\ell_i^{\ge})$ for some $i \in [n]$. Note that in this case, for all $i' \in \mathsf{Good}(t)$, $\mathsf{Emb}(i') > t$, so $\mathsf{P}(t') = \emptyset$. Furthermore, $\mathsf{P}(t'-1) = \{i' \in \mathsf{Good}(t-1) \mid \mathsf{Emb}(i') \le t - 1\} = \left\{\ell_i^{<} + 1, \ldots, \ell_i^{\ge} - 1\right\}$. Thus, $|\mathsf{P}(t'-1)| = \ell_i^{\ge} - \ell_i^{<} - 1$.

We will show that the result holds even when conditioned on an arbitrary $\Phi_{<t'}$. We have:

$$\begin{aligned}
\Pr\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|} < \frac{1}{m^{2(1+|\mathsf{P}(t'-1)|)}} \mid \Phi_{<t'}\right) &= \sum_{\Phi'_t} \Pr(\Phi_{t'} \mid \Phi_{<t'}) \mathbb{1}\left(\frac{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{\le t'})\right|}{\left|\mathsf{Rec}^C(\mathsf{\Phi}_{< t'})\right|} < \frac{1}{m^{2\left(\ell_i^{\ge} - \ell_i^{<}\right)}}\right) \\
&= \sum_{\Phi'_t} \Pr(\Phi_{t'} \mid \Phi_{<t'}) \mathbb{1}\left(\Pr(\Phi_{t'} \mid \Phi_{<t'}) < \frac{1}{m^{2\left(\ell_i^{\ge} - \ell_i^{<}\right)}}\right) \\
&\hfill \text{(Lemma 5.7)}
\end{aligned}$$

$$\begin{aligned}&\le \frac{m^{\ell_i^{\geq}-\ell_i^{<}}}{m^{2\left(\ell_i^{\geq}-\ell_i^{<}\right)}}\\ &\le \frac{1}{m^{0.6}}.\end{aligned}$$

□

We now have the tools necessary to finish the proof of Lemma 5.4.

Proof of Lemma 5.4. For all $\Phi_{\leq 3t} \in \mathsf{Info}^{C}(t)$ we have:

$$\begin{aligned}
&\mathbb{D}\big(\mathsf{dist}(\mathsf{f}^{C} \mid \Phi_{\leq 3t}) \mid\mid \mathcal{U}\big) > m^{0.01} && \text{(Eq. (12))}\\
&\implies \mathbb{D}\Big(\mathsf{dist}\Big(\mathsf{f}^{C} \mid \forall C' \in \{A,B\} : \mathsf{f}^{C'} \in \mathsf{Rec}^{C'}(\Phi_{\leq 3t})\Big) \mid\mid \mathcal{U}\Big) > m^{0.01} && \text{(Lemma 5.6)}\\
&\implies \mathbb{D}\big(\mathsf{dist}\big(\mathsf{f}^{C} \mid \mathsf{f}^{C} \in \mathsf{Rec}^{C}(\Phi_{\leq 3t})\big) \mid\mid \mathcal{U}\big) > m^{0.01} && \text{(Independence of } \mathsf{f}^{A} \text{ and } \mathsf{f}^{B}\text{)}\\
&\implies \frac{\left|\mathsf{Rec}^{C}(\Phi_{\leq 3t})\right|}{|\mathsf{supp}(\mathsf{f}^{C})|} < \frac{1}{2^{m^{0.01}}}. && \text{(Lemma A.9)}\\
&\implies \frac{\left|\mathsf{Rec}^{C}(\Phi_{\leq 3t})\right|}{|\mathsf{supp}(\mathsf{f}^{C})|} < \frac{1}{m^{6t}}.
\end{aligned}$$

Thus, we get that $\Pr\big(\Phi_{\leq 3t} \in \mathsf{Info}^{C}(t)\big) \leq \Pr\left(\frac{\left|\mathsf{Rec}^{C}(\Phi_{\leq 3t})\right|}{|\mathsf{supp}(\mathsf{f}^{C})|} < \frac{1}{m^{6t}}\right)$. The result follows from Lemma 5.8. □

5.4 Key Lemma

We now show our key lemma.

Lemma 5.9. *Let $t \in \{0\} \cup [T]$ and $i \in \mathsf{Good}(t)$. We have:*

$$\Pr\big(\Phi_{\leq 3t} \in \mathsf{Guess}^{i}(t)\big) \leq \frac{t}{m^{0.1}}.$$

Proof. Proof by induction on t. The base $t = 0$ is straightforward as we get $\mathsf{Emb}(\ell_i^{<}) = 0$ which means that $\ell_i^{<} = 0$ implying that $\mathsf{Guess}^{i}(t) = \emptyset$. We show the result for $t > 0$ assuming it holds for smaller values of t. We consider the following cases:

- **When $\mathsf{Emb}(\ell_i^{<}) < t$:** At a high level, this case amounts to analyzing a variant of the well-known Index problem, where the party holding the index can communicate a small number of bits but not enough to send the entire index. Let $u = \mathsf{Emb}(\ell_i^{<})$ for convenience. Note that $i \in \mathsf{Good}(u)$. Applying the induction hypothesis on u, we get:

$$\Pr\big(\Phi_{\leq 3u} \in \mathsf{Guess}^{i}(u)\big) \leq \frac{u}{m^{0.1}}.$$

It is therefore sufficient to show that

$$\Pr\big(\Phi_{\leq 3t} \in \mathsf{Guess}^i(t) \mid \Phi_{\leq 3u} \notin \mathsf{Guess}^i(u)\big) \leq \frac{1}{m^{0.1}}.$$

We assume $\tau_i = A$ as the argument for $\tau_i = B$ is analogous. For this, we shall fix an arbitrary $\Phi_{\leq 3u} \notin \mathsf{Guess}^i(u)$, and an arbitrary f^B and show that

$$\Pr\big(\Phi_{\leq 3t} \in \mathsf{Guess}^i(t) \mid \Phi_{\leq 3u}, f^B\big) \leq \frac{1}{m^{0.1}}.$$

By Eq. (11), it suffices to show that:

$$\Pr\left(\Phi_{(3u,3t]} \in \left\{\Phi_{(3u,3t]} \mid \mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i}) \mid \Phi_{\leq 3t}\right) \leq \mathsf{Rem}^i(t)\right\} \mid \Phi_{\leq 3u}, f^B\right) \leq \frac{1}{m^{0.1}}. \quad (14)$$

We now focus on showing Eq. (14). Define $z = (0.2 + |\mathsf{E} \cap (u,t]|) \cdot \log m$ for convenience. For this, we will apply Lemma A.7 with $\mathsf{X} = \mathsf{f}^A$, $f(\mathsf{X}) = \mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i})$, $g(\mathsf{X}) = \Phi_{(3u,3t]}$, $E = \big(\Phi_{\leq 3u}, f^B\big)$, and $t = z$. Note that as conditioning on $\Phi_{\leq 3u}$ fixes the value of $\mathsf{PC}\left(\mathsf{f}_{\leq \ell_i^<}\right)$ (Claim 5.2), $\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i})$ is indeed a function of f^A. Similarly, as we condition on f^B, $\Phi_{(3u,3t]}$ is indeed a function of f^A. Finally, observe from Eqs. (9) and (10) that for all $u' \in (u,t]$, we have $\Phi_{3u'-1} = \Phi_{3u'} = 0$ and thus $g(\cdot)$ takes at most $m^{|\mathsf{E} \cap (u,t]|}$ many values.[9]

From Lemma A.7, we get:

$$\Pr\big(\Phi_{(3u,3t]} \in G^* \mid \Phi_{\leq 3u}, f^B\big) \leq \frac{1}{m^{0.1}}, \quad (15)$$

where:

$$G^* = \left\{\Phi_{(3u,3t]} \mid \mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i}) \mid \Phi_{\leq 3t}, f^B\right) \leq \mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i}) \mid \Phi_{\leq 3u}, f^B\right) - z\right\}.$$

Next, we claim that we can "drop" the conditioning on f^B. For this, recall from Lemma 5.6 that the events $\Phi_{\leq 3t}$ and $\big(\forall C \in \{A,B\} : \mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{\leq 3t})\big)$ are equivalent. As the latter event is a combinatorial rectangle (it is of the form $\big(\mathsf{f}^A \in \mathcal{A}\big) \wedge \big(\mathsf{f}^B \in \mathcal{B}\big)$ for some sets $\mathcal{A}, \mathcal{B}$) and the random variables f^A and f^B are independent, we get that the random variables f^A and f^B are also independent conditioned on $\Phi_{\leq 3t}$. Next, recall that $\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i})$ is a function of f^A conditioned on $\Phi_{\leq 3u}$, and conclude that $\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i})$ and f^B are also independent conditioned on $\Phi_{\leq 3t}$ allowing us to drop f^B. A similar

[9] For $t' \in (u,t]$ where $\tau_{t'} = B$, conditioning on f^B makes each message $\Phi_{3t'-2}$ a deterministic function of the transcript so far. As such, there are at most $m^{|\{t' \in (u,t] : \tau_{t'} = A\}|}$ possible transcripts, as there are m possible values of $\Phi_{3t'-2}$ for each $t' \in (u,t]$ where $\tau_{t'} = A$. Finally, by the definition of $\ell_i^{\geq}$ and Eqs. (1) and (2), we get that $\{t \in (u,t] : \tau_{t'} = A\} = \mathsf{E} \cap (u,t]$.

argument allows us to drop f^B from the other min-entropy term and we get:

$$G^* = \left\{ \Phi_{(3u,3t]} \mid \mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) \mid \Phi_{\le 3t}\right) \le \mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) \mid \Phi_{\le 3u}\right) - z \right\}.$$

By Eq. (11) and our choice of $\Phi_{\le 3u} \notin \mathsf{Guess}^i(u)$, we have:

$$G^* \supseteq \left\{ \Phi_{(3u,3t]} \mid \mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) \mid \Phi_{\le 3t}\right) \le \mathsf{Rem}^i(u) - z \right\}.$$

Using Eq. (7) and the definition of z, we get:

$$G^* \supseteq \left\{ \Phi_{(3u,3t]} \mid \mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) \mid \Phi_{\le 3t}\right) \le \mathsf{Rem}^i(t) \right\}.$$

This together with Eq. (15) shows Eq. (14).

- **When $\mathsf{Emb}(\ell_i^<) = t$:** At a high level, the analysis in this case follows the popular pointer chasing lower bound of [NW91]. We assume $\tau_i = A$ as the argument for $\tau_i = B$ is analogous. As $t > 0$, we have $\ell_i^< > 0$ and we get $\tau_{\ell_i^<} = B$. It follows that $\sigma_t = B$. Let $u = t - 1$. Applying the induction hypothesis on u and $\ell_i^<$, we get:

$$\Pr\left(\Phi_{\le 3u} \in \mathsf{Guess}^{\ell_i^<}(u)\right) \le \frac{u}{m^{0.1}}.$$

By Lemma 5.4, we also have:

$$\Pr\left(\Phi_{\le 3u} \in \mathsf{Info}^A(u)\right) \le \frac{1}{m^{0.5}}.$$

Thus, it suffices to show that:

$$\Pr\left(\Phi_{\le 3t} \in \mathsf{Guess}^i(t) \mid \Phi_{\le 3u} \notin \mathsf{Guess}^{\ell_i^<}(u) \cup \mathsf{Info}^A(u)\right) \le \frac{1}{m^{0.15}}.$$

For this, we shall fix an arbitrary $\Phi_{\le 3u} \notin \mathsf{Guess}^{\ell_i^<}(u) \cup \mathsf{Info}^A(u)$ and show that:

$$\Pr\left(\Phi_{\le 3t} \in \mathsf{Guess}^i(t) \mid \Phi_{\le 3u}\right) \le \frac{1}{m^{0.15}}. \tag{16}$$

For all $i' \in \left(\ell_i^<, \ell_i^\ge\right]$, define the set:

$$S_{i'} = \left\{ z \in [m]^{\ell_i^<} \mid \mathbb{D}\left(\mathsf{dist}\left(\mathsf{PC}_{>\ell_i^<}\left(z, \mathsf{f}_{(\ell_i^<, i']}\right) \mid \Phi_{\le 3u}\right) \mid\mid \mathcal{U}\right) \ge \frac{1}{m^{0.42}} \right\}. \tag{17}$$

Also, define $S = \bigcup_{i' \in \left(\ell_i^<, \ell_i^\ge\right]} S_{i'}$. Roughly speaking, S is the set of prefixes z that allow the parties to guess the transcript in the next interval. Recall that $\mathsf{Emb}(\ell_i^<) = t$ means that we are currently at the end of an interval. We now show that the probability of landing in S is small, as formalized in Eq. (18) below. Next, use the fact that $\tau_i = A$ and the definition of $\ell_i^<$ and $\ell_i^{\geq}$ to conclude that $\tau_{i'} = A$ for all $i' \in \left(\ell_i^<, \ell_i^{\geq}\right]$. It follows that for all $i' \in \left(\ell_i^<, \ell_i^{\geq}\right]$, f^A determines $\mathsf{f}_{\left(\ell_i^<, i'\right]}$. This, together with Lemma A.11 and the fact that $\mathbb{D}\big(\mathsf{dist}\big(\mathsf{f}^A \mid \Phi_{\leq 3u}\big) \mid\mid \mathcal{U}\big) \leq m^{0.01}$ (which follows as $\Phi_{\leq 3u} \notin \mathsf{Info}^A(u)$) implies that $\mathbb{D}\Big(\mathsf{dist}\Big(\mathsf{f}_{\left(\ell_i^<, i'\right]} \mid \Phi_{\leq 3u}\Big) \mid\mid \mathcal{U}\Big) \leq m^{0.01}$. We get, for all $i' \in \left(\ell_i^<, \ell_i^{\geq}\right]$, that:

$$\begin{aligned}
m^{0.01} &\geq \mathbb{D}\Big(\mathsf{dist}\Big(\mathsf{f}_{\left(\ell_i^<, i'\right]} \mid \Phi_{\leq 3u}\Big) \mid\mid \mathcal{U}\Big) \\
&\geq \sum_{z \in [m]^{\ell_i^<}} \mathbb{D}\Big(\mathsf{dist}\Big(\mathsf{PC}_{>\ell_i^<}\Big(z, \mathsf{f}_{\left(\ell_i^<, i'\right]}\Big) \mid \Phi_{\leq 3u}\Big) \mid\mid \mathcal{U}\Big) && \text{(Lemma A.11)} \\
&\geq \sum_{z \in S_{i'}} \mathbb{D}\Big(\mathsf{dist}\Big(\mathsf{PC}_{>\ell_i^<}\Big(z, \mathsf{f}_{\left(\ell_i^<, i'\right]}\Big) \mid \Phi_{\leq 3u}\Big) \mid\mid \mathcal{U}\Big) \\
&\geq |S_{i'}| \cdot \frac{1}{m^{0.42}}. && \text{(Eq. (17))}
\end{aligned}$$

As such, we get that for all $i' \in \left(\ell_i^<, \ell_i^{\geq}\right]$, we have $|S_{i'}| \leq m^{0.44}$ implying that $|S| \leq m^{0.45}$. Next, note that as $\Phi_{\leq 3u} \notin \mathsf{Guess}^{\ell_i^<}(u)$, we also have by Eq. (11) that $\mathbb{H}_\infty(\mathsf{PC}_{>\ell^<_{\ell_i^<}}(\mathsf{f}_{\leq \ell_i^<}) \mid \Phi_{\leq 3u}) \geq \mathsf{Rem}^{\ell_i^<}(u)$. Observe from Claim 5.2 that conditioning on $\Phi_{\leq 3u}$ fixes the value of $\mathsf{PC}(\mathsf{f}_{\leq \ell^<_{\ell_i^<}}) = \mathsf{PC}_{\leq \ell^<_{\ell_i^<}}(\mathsf{f}_{\leq \ell_i^<})$. Thus, we get $\mathbb{H}_\infty(\mathsf{PC}(\mathsf{f}_{\leq \ell_i^<}) \mid \Phi_{\leq 3u}) \geq \mathsf{Rem}^{\ell_i^<}(u)$. It follows from Eq. (7) and Definition A.5 that

$$\Pr\Big(\mathsf{PC}\Big(\mathsf{f}_{\leq \ell_i^<}\Big) \in S \mid \Phi_{\leq 3u}\Big) \leq m^{0.45} \cdot 2^{-\mathsf{Rem}^{\ell_i^<}(u)} \leq \frac{1}{m^{0.25}}. \tag{18}$$

As a consequence, Eq. (16) follows if we show that:

$$\Pr\Big(\Phi_{\leq 3t} \in \mathsf{Guess}^i(t) \mid \Phi_{\leq 3u}, \mathsf{PC}\Big(\mathsf{f}_{\leq \ell_i^<}\Big) \notin S\Big) \leq \frac{1}{m^{0.2}}.$$

Next, note from Claim 5.2 that the value of $\Phi_{<3t}$ fixes the value of $\mathsf{PC}(\mathsf{f}_{\leq \ell_i^<})$ (and also of $\Phi_{\leq 3u}$). Thus, it suffices to fix an arbitrary $\Phi_{<3t}$ that agrees with $\Phi_{\leq 3u}$ and for which the corresponding value of $\mathsf{PC}(f_{\leq \ell_i^<}) \notin S$, and show that

$$\Pr\big(\Phi_{\leq 3t} \in \mathsf{Guess}^i(t) \mid \Phi_{<3t}\big) \leq \frac{1}{m^{0.2}}.$$

By Eq. (11), this is the same as:

$$\Pr\Big(\Phi_{3t} \in \Big\{\Phi_{3t} \mid \mathbb{H}_\infty\Big(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i}) \mid \Phi_{\leq 3t}\Big) < \mathsf{Rem}^i(t)\Big\} \mid \Phi_{<3t}\Big) \leq \frac{1}{m^{0.2}}.$$

Fixing such a $\Phi_{<3t}$, this is because of the following two claims, that we show later.

Claim 5.10. *It holds that:*

$$\Pr(\Phi_{3t} = 1 \mid \Phi_{<3t}) \leq \frac{1}{m^{0.2}}.$$

Claim 5.11. *It holds that:*

$$\mathbb{H}_\infty\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i}) \mid \Phi_{<3t}, \Phi_{3t} = 0\right) \geq \mathsf{Rem}^i(t).$$

□

We now show Claims 5.10 and 5.11.

Proof of Claim 5.10. For all $i' \in \left(\ell_i^<, \ell_i^\geq\right]$, define the set:

$$W_{i'} = \left\{w \in [m]^{i'-\ell_i^<} : \Pr\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i'}) = w \mid \Phi_{<3t}\right) > 2^{-1-\mathsf{Rem}^{i'}(t)}\right\}. \tag{19}$$

With this definition and a union bound, we get:

$$\begin{aligned}
\Pr(\Phi_{3t} = 1 \mid \Phi_{<3t}) &\leq \sum_{i' \in \left(\ell_i^<, \ell_i^\geq\right]} \Pr\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i'}) \in W_{i'} \mid \Phi_{<3t}\right) && (\text{Eq. (10)}) \\
&\leq \sum_{i' \in \left(\ell_i^<, \ell_i^\geq\right]} \sum_{w \in W_{i'}} \Pr\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i'}) = w \mid \Phi_{<3t}\right) \\
&\leq \sqrt{2} \cdot \sum_{i' \in \left(\ell_i^<, \ell_i^\geq\right]} \sum_{w \in W_{i'}} \left(\Pr\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i'}) = w \mid \Phi_{<3t}\right) - 2^{-\left(i'-\ell_i^<\right)\cdot\log m}\right) \\
&\qquad (\text{Eqs. (7) and (19) imply that } \Pr\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i'}) = w \mid \Phi_{<3t}\right) > 10 \cdot 2^{-\left(i'-\ell_i^<\right)\cdot\log m}) \\
&\leq \sqrt{2} \cdot \sum_{i' \in \left(\ell_i^<, \ell_i^\geq\right]} \left\|\mathsf{dist}\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i'}) \mid \Phi_{<3t}\right) - \mathcal{U}\right\|_{\mathrm{TV}} && (\text{Definition A.12}) \\
&\leq \sum_{i' \in \left(\ell_i^<, \ell_i^\geq\right]} \sqrt{\mathbb{D}\left(\mathsf{dist}\left(\mathsf{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i'}) \mid \Phi_{<3t}\right) \,||\, \mathcal{U}\right)} && (\text{Fact A.13}) \\
&= \sum_{i' \in \left(\ell_i^<, \ell_i^\geq\right]} \sqrt{\mathbb{D}\left(\mathsf{dist}\left(\mathsf{PC}_{>\ell_i^<}\left(\mathsf{PC}\left(\mathsf{f}_{\leq \ell_i^<}\right), \mathsf{f}_{\left(\ell_i^<, i'\right]}\right) \mid \Phi_{<3t}\right) \,||\, \mathcal{U}\right)} \\
&= \sum_{i' \in \left(\ell_i^<, \ell_i^\geq\right]} \sqrt{\mathbb{D}\left(\mathsf{dist}\left(\mathsf{PC}_{>\ell_i^<}\left(\mathsf{PC}\left(f_{\leq \ell_i^<}\right), \mathsf{f}_{\left(\ell_i^<, i'\right]}\right) \mid \Phi_{<3t}\right) \,||\, \mathcal{U}\right)}, \\
&\qquad (\text{Claim 5.2 implies } \Phi_{<3t} \text{ fixes } \mathsf{PC}\left(\mathsf{f}_{\leq \ell_i^<}\right))
\end{aligned}$$

where $\mathsf{PC}\left(f_{\leq \ell_i^<}\right) \notin S$ is the value determined by our choice of $\Phi_{(3u,3t)}$. Now, recall that $u = t - 1$. Thus, $\Phi_{(3u,3t)} = (\Phi_{3t-2}, \Phi_{3t-1})$. We now argue that we can remove $(\Phi_{3t-2}, \Phi_{3t-1})$

from the conditioning. For this, recall from Lemma 5.6 that the event $\Phi_{\le 3u}$ and the event $\big(\forall C \in \{A,B\} : \mathsf{f}^C \in \mathsf{Rec}^C(\Phi_{\le 3u})\big)$ are equivalent. As the latter event is a combinatorial rectangle (it is of the form $\big(\mathsf{f}^A \in \mathcal{A}\big) \wedge \big(\mathsf{f}^B \in \mathcal{B}\big)$ for some sets $\mathcal{A}, \mathcal{B}$) and the random variables f^A and f^B are independent, we get that the random variables f^A and f^B are also independent conditioned on $\Phi_{\le 3u}$.

Next, recall that $\mathsf{f}_{(\ell_i^<,i']}$ is determined by f^A, and conclude that the random variables $\mathsf{f}_{\left(\ell_i^<,i'\right]}$ and f^B are also independent conditioned on $\Phi_{\le 3u}$. Finally, as $\sigma'_{3t-2} = \sigma'_{3t-1} = \sigma_t = \tau_{\ell_i^<} = B$ using $t = \mathsf{Emb}(\ell_i^<)$ and Lemma 5.3, conclude from Lemma 5.3 that $\Phi_{<3t}$ is determined by f^B conditioned on $\Phi_{\le 3u}$. Thus, we get that the random variable $\mathsf{f}_{(\ell_i^<,i']}$, and therefore also the random variable $\mathtt{PC}_{>\ell_i^<}(\mathtt{PC}(f_{\le \ell_i^<}), \mathsf{f}_{(\ell_i^<,i']})$, is independent of the random variable $(\Phi_{3t-2}, \Phi_{3t-1})$ conditioned on $\Phi_{\le 3u}$. Plugging in, we get:

$$\Pr(\Phi_{3t} = 1 \mid \Phi_{<3t}) \le \sum_{i' \in \left(\ell_i^<, \ell_i^\ge\right]} \sqrt{\mathbb{D}\Big(\mathsf{dist}\Big(\mathtt{PC}_{>\ell_i^<}\Big(\mathtt{PC}\Big(f_{\le \ell_i^<}\Big), \mathsf{f}_{\left(\ell_i^<, i'\right]}\Big) \mid \Phi_{\le 3u}\Big) \;\|\; \mathcal{U}\Big)}.$$

The proof is complete by the fact that $\mathtt{PC}(f_{\le \ell_i^<}) \notin S$ and Eq. (17).

□

Proof of Claim 5.11. From Definition A.5, the claim is equivalent to showing that for all $z \in [m]^{i-\ell_i^<}$, we have:

$$\Pr\Big(\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) = z \mid \Phi_{<3t}, \Phi_{3t} = 0\Big) \le 2^{-\mathsf{Rem}^i(t)}.$$

Owing to Claim 5.10, it suffices to show that for all $z \in [m]^{i-\ell_i^<}$, we have:

$$\Pr\Big(\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) = z, \Phi_{3t} = 0 \mid \Phi_{<3t}\Big) \le 2^{-1-\mathsf{Rem}^i(t)}.$$

The foregoing equation is trivial if $\Pr\Big(\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) = z \mid \Phi_{<3t}\Big) \le 2^{-1-\mathsf{Rem}^i(t)}$, so it suffices to consider $z \in [m]^{i-\ell_i^<}$ such that $\Pr\Big(\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) = z \mid \Phi_{<3t}\Big) > 2^{-1-\mathsf{Rem}^i(t)}$. Fix such a z. By Eq. (10) and our choice of z, we get that the event $\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) = z$ implies the event $\Phi_{3t} = 1$. This means that:

$$\Pr\Big(\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\le i}) = z, \Phi_{3t} = 0 \mid \Phi_{<3t}\Big) = 0,$$

and we are done. □

5.5 Finishing the Proof

We are now ready to prove Lemma 5.1

Proof of Lemma 5.1. Recall that we are showing the lemma in the contrapositive, assuming that τ is a not subsequence of σ. This means that $\mathsf{Emb}(n) > T$ implying by Eq. (6) that there exists $i \in [n]$ such that $i \in \mathsf{Good}(T)$. Fix such an i and apply Lemma 5.9 to conclude that $\Pr(\mathcal{E}) < \frac{1}{n^2}$, where $\mathcal{E}$ is the event $\Phi_{\leq 3T} \in \mathsf{Guess}^i(T)$. We now derive Eq. (4) as follows:

$$\begin{aligned} \Pr(\mathsf{res}_\Pi(\mathsf{F}) = \mathtt{PC}(\mathsf{F})) &\leq \Pr(\mathcal{E}) + \Pr\big(\mathsf{res}_\Pi(\mathsf{F}) = \mathtt{PC}(\mathsf{F}) \mid \overline{\mathcal{E}}\big) && \text{(Union bound)} \\ &< \frac{1}{n^2} + \Pr\big(\mathsf{res}_\Pi(\mathsf{F}) = \mathtt{PC}(\mathsf{F}) \mid \overline{\mathcal{E}}\big). \end{aligned}$$

Thus, it suffices to show that $\Pr\big(\mathsf{res}_\Pi(\mathsf{F}) = \mathtt{PC}(\mathsf{F}) \mid \overline{\mathcal{E}}\big) \leq \frac{1}{n^2}$. We show this holds under a stronger conditioning by conditioning on an arbitrary $\Phi_{\leq 3T}$ such that $\Phi_{\leq 3T} \notin \mathsf{Guess}^i(T)$. Fixing such a $\Phi_{\leq 3T}$ and noting that fixing $\Phi_{\leq 3T}$ also fixes $\mathsf{res}_\Pi(\mathsf{F})$ to some value res we get:

$$\begin{aligned} \Pr(\mathsf{res}_\Pi(\mathsf{F}) = \mathtt{PC}(\mathsf{F}) \mid \Phi_{\leq 3T}) &\leq \Pr(\mathtt{PC}(\mathsf{F}) = \mathsf{res} \mid \Phi_{\leq 3T}) \\ &\leq \Pr\Big(\mathtt{PC}_{>\ell_i^<}(\mathsf{f}_{\leq i}) = \mathsf{res}_{>\ell_i^<} \mid \Phi_{\leq 3T}\Big) \\ &\leq 2^{-\mathsf{Rem}^i(T)} && \text{(Eq. (11) as } \Phi_{\leq 3T} \notin \mathsf{Guess}^i(T)) \\ &\leq m^{-0.5} && \text{(Eq. (7))} \\ &\leq \frac{1}{n^2}. \end{aligned}$$

□

6 Proof of Theorem 4.2

In this section, we prove Theorem 4.2. For notational convenience, we define the constant $\eta = 10^{-5}$.

6.1 A Customized Concentration Inequality

Fact 6.1. *For all integers $1 \leq k \leq n$, we have:*

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{3n}{k}\right)^k.$$

Lemma 6.2. *Let $Z \subseteq \mathbb{N}$ and $n > 0$ be an integer. Also, let $p > 0$ and $\mathsf{X}_1, \mathsf{X}_2, \cdots, \mathsf{X}_n$ be random variables taking values in $\mathbb{N}$. Then, if $\delta > 0$ is such that for all $i \in [n]$ and all $x_1, x_2, \cdots, x_{i-1} \in \mathbb{N}$, we have:*

$$\Pr(\mathsf{X}_i = 1 \mid \forall i' \in [i-1] : \mathsf{X}_{i'} = x_{i'}) \leq 1 - p \cdot \mathbb{1}\left(\sum_{i'=1}^{i-1} x_{i'} \notin Z\right) \cdot \mathbb{1}\left(\sum_{i'=1}^{i-1} x_{i'} \leq (1+\delta) \cdot n - 1\right).$$

Then, it holds that:

$$\Pr\left(\sum_{i=1}^{n} \mathsf{X}_i \leq (1+\delta)\cdot n\right) \leq 2^{-pn+\delta n\cdot\log\frac{12}{\delta}+|Z|}.$$

Proof. Define the set:

$$S = \left\{(x_1, x_2, \cdots, x_n) \in \mathbb{N}^n \mid \sum_{i=1}^{n} x_i \leq (1+\delta)\cdot n\right\},$$

to be the set of all n-tuples of positive integers that sum to at most $(1+\delta)\cdot n$. From a standard argument[10], it follows that $|S| = \binom{n(1+\delta)}{\delta n}$. Using this bound, we have:

$$\begin{aligned}\Pr\left(\sum_{i=1}^{n} \mathsf{X}_i \leq (1+\delta)\cdot n\right) &= \sum_{(x_1,x_2,\cdots,x_n)\in S} \Pr(\forall i \in [n] : \mathsf{X}_i = x_i) \\ &\leq \binom{n(1+\delta)}{\delta n} \cdot \max_{(x_1,x_2,\cdots,x_n)\in S} \Pr(\forall i \in [n] : \mathsf{X}_i = x_i).\end{aligned} \tag{20}$$

We now consider an arbitrary n-tuple $(x_1, x_2, \cdots, x_n) \in S$ and upper bound the probability term corresponding to this n-tuple. We have:

$$\begin{aligned}\Pr(\forall i \in [n] : \mathsf{X}_i = x_i) &= \prod_{i\in[n]} \Pr(\mathsf{X}_i = x_i \mid \forall i' \in [i-1] : \mathsf{X}_{i'} = x_{i'}) \\ &\leq \prod_{i\in[n]:x_i=1} \Pr(\mathsf{X}_i = 1 \mid \forall i' \in [i-1] : \mathsf{X}_{i'} = x_{i'}) \\ &\leq \prod_{i\in[n]:x_i=1} \left(1 - p\cdot \mathbb{1}\left(\sum_{i'=1}^{i-1} x_{i'} \notin Z\right)\right) \\ &\qquad\qquad \text{(Assumption in the lemma and the definition of } S\text{)} \\ &\leq \prod_{i\in[n]:x_i=1} 2^{-p\cdot\mathbb{1}\left(\sum_{i'=1}^{i-1} x_{i'}\notin Z\right)} \qquad \text{(As } 1-x \leq 2^{-x} \text{ for all } x \geq 0\text{)} \\ &\leq 2^{-p\cdot\sum_{i\in[n]:x_i=1}\mathbb{1}\left(\sum_{i'=1}^{i-1} x_{i'}\notin Z\right)}.\end{aligned}$$

To continue, let K be the number of i such that $x_i = 1$ and note that the definition of S

[10] We provide the argument here for completeness. The claim is that, for integers $n, r > 0$, the number of non-negative integer solutions of $\sum_{i\in[n]} z_i \leq r$ is equal to $\binom{n+r}{r}$. Indeed, every bit string of length $n+r$ that has n zeros can be interpreted as a solution, where for all $i \in [n]$, the number of 1s between the $(i-1)^{\text{th}}$ and the i^{th} zero is the value of z_i (the 0^{th} zero is assumed to be at location 0), and any solution can be written as such a bit string with the 1s after the n^{th} corresponding to the "slack" in $\sum_{i\in[n]} z_i \leq r$. Thus, the number of solutions is equal to the number of strings, which is equal to $\binom{n+r}{r}$.

requires that $K \geq (1-\delta) \cdot n$. We have:

$$\begin{aligned}\Pr(\forall i \in [n] : \mathsf{X}_i = x_i) &\leq 2^{-pK + \sum_{i \in [n] : x_i = 1} \mathbb{1}\left(\sum_{i'=1}^{i-1} x_{i'} \in Z\right)} \\ &\leq 2^{-pn + \delta n + \sum_{i \in [n] : x_i = 1} \mathbb{1}\left(\sum_{i'=1}^{i-1} x_{i'} \in Z\right)}.\end{aligned}$$

Next, note that the values $\sum_{i'=1}^{i-1} x_{i'}$ are distinct for all $i \in [n]$ and non-negative. Thus, we get the bound:

$$\Pr(\forall i \in [n] : \mathsf{X}_i = x_i) \leq 2^{-pn + \delta n + |Z|}.$$

As this bound was shown for an arbitrary n-tuple $(x_1, x_2, \cdots, x_n) \in S$, we can plug it into Eq. (20) and get:

$$\begin{aligned}\Pr\left(\sum_{i=1}^{n} \mathsf{X}_i \leq (1+\delta) \cdot n\right) &\leq \binom{n(1+\delta)}{\delta n} \cdot 2^{-pn + \delta n + |Z|} \\ &\leq \left(\frac{3(1+\delta)}{\delta}\right)^{\delta n} \cdot 2^{-pn + \delta n + |Z|} && \text{(Fact 6.1)} \\ &\leq 2^{-pn + \delta n \cdot \log \frac{12}{\delta} + |Z|}. && \text{(As } \delta \in (0,1))\end{aligned}$$

□

6.2 Basic Definitions

Recall that $\eta = 10^{-5}$. Also recall from Section 2.2 that we consider segments of geometrically increasing lengths. These segments will be parameterized by an integer $\ell > 0$. We will use L_ℓ to denote the length of segment ℓ, D_ℓ to denote the "delay" or the "lag" before the segment starts, and C_ℓ to denote the non-bullet symbols in the pattern for this segment. We set these parameters as follows:

$$L_\ell = \eta^{-2\ell-2} \qquad C_\ell = \lfloor \eta^6 L_\ell \rfloor \qquad D_\ell = \eta^4 L_\ell. \tag{21}$$

We also define $L_{\leq \ell} = \sum_{\ell'=1}^{\ell} L_{\ell'}$ and $L_{<\ell} = \sum_{\ell'=1}^{\ell-1} L_{\ell'}$. We adopt the convention that $L_{\leq 0} = 0$ and observe that all these parameters integers. Next, we define the set $\{A, B\}_\bullet$ to denote the set $\{A, B\}_\bullet = \{A, B\} \cup \{\bullet\}$. For $\rho \in \{A, B\}_\bullet^*$, we use $\mathsf{bull}(\rho)$ to denote the number of coordinates in the string ρ that are equal to the "bullet" symbol $\bullet$. Formally, we have $\mathsf{bull}(\rho) = |\{i \in [|\rho|] \mid \rho_i = \bullet\}|$. The following simple lemma counts the number of strings ρ with a given value of $\mathsf{bull}(\rho)$.

Lemma 6.3. *For all $0 \leq T' \leq T$, we have:*

$$\left|\left\{\rho \in \{A, B\}_\bullet^T \mid \mathsf{bull}(\rho) = T'\right\}\right| = 2^{T-T'} \cdot \binom{T}{T-T'}.$$

Proof. There are exactly $\binom{T}{T'} = \binom{T}{T-T'}$ of choosing the T' "bullet" coordinates and for each such choice, there are $2^{T-T'}$ way of choosing the other coordinates. $\square$

For strings $\rho \in \{A, B\}_\bullet^*$ and $\sigma \in \{A, B\}^{\mathsf{bull}(\rho)}$, we can insert the coordinates of σ into the bullet coordinates of ρ to get a string $\mathsf{ins}(\sigma, \rho) \in \{A, B\}^{|\rho|}$, whose i^{th} coordinate, for $i \in [|\rho|]$, is denoted by $\mathsf{ins}_i(\sigma, \rho)$ and defined as:

$$\mathsf{ins}_i(\sigma, \rho) = \begin{cases} \rho_i, & \text{if } \rho_i \neq \bullet \\ \sigma_{\mathsf{bull}(\rho_{\leq i})}, & \text{if } \rho_i = \bullet \end{cases}. \tag{22}$$

The function $\mathsf{ins}(\cdot)$ satisfies the following:

Lemma 6.4. *Let $\sigma, \sigma' \in \{A, B\}^*$ and define $T = |\sigma|$ and $T' = |\sigma'|$. For all (possibly empty[11]) sets $S \subseteq [T]$ such that $\mathsf{Emb}(\sigma, \sigma', \max(S)) \leq T'$, there exists a string $\rho \in \{A, B\}_\bullet^{T'}$ such that $\mathsf{bull}(\rho) = |S|$ and $\mathsf{ins}(\sigma_S, \rho) = \sigma'$.*

Proof. We start by defining the string ρ. For $i' \in [T']$, define:

$$\rho_{i'} = \begin{cases} \bullet, & \text{if } \exists i \in S : \mathsf{Emb}(\sigma, \sigma', i) = i' \\ \sigma'_{i'}, & \text{otherwise} \end{cases} \tag{23}$$

As Observation 3.1 implies that the values of $\mathsf{Emb}(\sigma, \sigma', i)$ are distinct for all $i \in S$, we have $\mathsf{bull}(\rho) = |S|$ finishing the first part of the proof. For the second part, let $s_1 < \cdots < s_{|S|}$ be the elements of S, and note that Observation 3.1 also implies that for all $i' \in [T']$ such that $\rho_{i'} = \bullet$, we have $i' = \mathsf{Emb}\Big(\sigma, \sigma', s_{\mathsf{bull}(\rho_{\leq i'})}\Big)$. We get that, for all $i' \in [T']$,

$$\begin{aligned} \mathsf{ins}_{i'}(\sigma_S, \rho) &= \begin{cases} \rho_{i'}, & \text{if } \rho_{i'} \neq \bullet \\ \sigma_{s_{\mathsf{bull}(\rho_{\leq i'})}}, & \text{if } \rho_{i'} = \bullet \end{cases} \\ &= \begin{cases} \rho_{i'}, & \text{if } \rho_{i'} \neq \bullet \\ \sigma'_{i'}, & \text{if } \rho_{i'} = \bullet \end{cases} && \Big(\text{As } i' = \mathsf{Emb}\Big(\sigma, \sigma', s_{\mathsf{bull}(\rho_{\leq i'})}\Big) \leq T'\Big) \\ &= \sigma'_{i'}. && (\text{Eq. (23)}) \end{aligned}$$

$\square$

Lemma 6.5. *Let $\sigma, \sigma' \in \{A, B\}^*$ and define $T = |\sigma|$ and $T' = |\sigma'|$. Let $i \in [T]$, $i' \in \{0\} \cup [T']$ be such that $\mathsf{Emb}(\sigma, \sigma', i-1) \leq i' < \mathsf{Emb}(\sigma, \sigma', i)$. For all $i'' \geq i' \in [T']$ and all $0 \leq b \leq |\mathsf{E}(\sigma, \sigma') \cap (i', i'']|$, there is a string $\rho \in \{A, B\}_\bullet^{i''-i'}$ such that $\mathsf{bull}(\rho) = b$ and:*

$$\mathsf{ins}\big(\sigma_{[i, i+b)}, \rho\big) = \sigma'_{(i', i'']}.$$

[11] We adopt the convention that $\max(\emptyset) = 0$.

Proof. To start, use Lemmas 3.2 and 3.5 to get:

$$\mathsf{Emb}\Big(\sigma_{\geq i}, \sigma'_{(i',i'']}, b\Big) + i' = \max(i', \mathsf{Emb}(\sigma, \sigma', b+i-1)) \leq i''.$$

It follows that $\mathsf{Emb}\Big(\sigma_{\geq i}, \sigma'_{(i',i'']}, b\Big) \leq i'' - i'$. This, together with the fact that $b \leq T - i + 1$ (which follows from $\mathsf{Emb}(\sigma, \sigma', i-1) \leq i'$) implies due to Lemma 6.4 that there exists a string $\rho \in \{A,B\}_{\bullet}^{i''-i'}$ such that $\mathsf{bull}(\rho) = b$ and $\mathsf{ins}\big(\sigma_{[i,i+b)}, \rho\big) = \sigma'_{(i',i'']}$, as desired. □

6.3 Predictable Indices

We are now ready to define the notion of predictable indices.

Definition 6.6 (Predictable indices)**.** *Let $\ell > 0$ and $\sigma \in \{A,B\}^*$ be given. Let i be an integer satisfying $D_\ell \leq i \leq |\sigma| - 2L_\ell$. For all integers $0 \leq j \leq L_\ell$ and $\rho \in \{A,B\}_{\bullet}^{L_\ell}$ satisfying $\mathsf{bull}(\rho) = L_\ell - C_\ell$, define the function*[12]*:*

$$\mathsf{Delay}(\ell, \sigma, i, j, \rho) = \mathsf{Emb}\Big(\sigma_{>i-D_\ell}, \mathsf{ins}\big(\sigma_{(i+j,i+j+L_\ell-C_\ell]}, \rho\big), (1-\eta)\cdot L_\ell\Big).$$

We say that i is ℓ-predictable in σ if there exist ρ, j as above for which $\mathsf{Delay}(\ell, \sigma, i, j, \rho) \leq L_\ell$ and use $\mathsf{Pred}_\ell(\sigma)$ to denote the set of all indices $D_\ell \leq i \leq T - 2L_\ell$ that are ℓ-predictable in σ.

Lemma 6.7. *Let integers $T, \ell > 0$ and $D_\ell \leq i \leq T - 2L_\ell$ be given. We have:*

$$\Pr_{\boldsymbol{\sigma}\sim\{A,B\}^T}(i \in \mathsf{Pred}_\ell(\boldsymbol{\sigma})) \leq 2^{-\frac{L_\ell}{4}}.$$

Proof. For brevity of notation, we omit writing $\boldsymbol{\sigma} \sim \{A,B\}^T$. Using Definition 6.6 and a union bound, we have:

$$\begin{aligned}
&\Pr(i \in \mathsf{Pred}_\ell(\boldsymbol{\sigma})) \\
&\quad\leq \Pr\Big(\exists 0 \leq j \leq L_\ell, \rho \in \{A,B\}_{\bullet}^{L_\ell} : \mathsf{bull}(\rho) = L_\ell - C_\ell \wedge \mathsf{Delay}(\ell, \boldsymbol{\sigma}, i, j, \rho) \leq L_\ell\Big) \\
&\quad\leq \sum_{j=0}^{L_\ell} \sum_{\substack{\rho\in\{A,B\}_{\bullet}^{L_\ell} \\ \mathsf{bull}(\rho)=L_\ell-C_\ell}} \Pr(\mathsf{Delay}(\ell, \boldsymbol{\sigma}, i, j, \rho) \leq L_\ell) \\
&\quad\leq 2L_\ell \cdot \binom{L_\ell}{C_\ell} \cdot 2^{C_\ell} \cdot \max_{0\leq j\leq L_\ell} \max_{\substack{\rho\in\{A,B\}_{\bullet}^{L_\ell} \\ \mathsf{bull}(\rho)=L_\ell-C_\ell}} \Pr(\mathsf{Delay}(\ell, \boldsymbol{\sigma}, i, j, \rho) \leq L_\ell). \qquad \text{(Lemma 6.3)}
\end{aligned}$$

□

To continue, note that if $\ell = 1 \implies C_\ell = 0$, we have $\binom{L_\ell}{C_\ell} \cdot 2^{C_\ell} \leq 1$. Otherwise, we get

[12] Note that Eq. (21) implies that $(1-\eta)\cdot L_\ell$ is an integer.

from Fact 6.1 and Eq. (21) that

$$\binom{L_\ell}{C_\ell} \cdot 2^{C_\ell} \leq 2^{72C_\ell}.$$

Thus, in either case, we have:

$$\Pr(i \in \mathsf{Pred}_\ell(\boldsymbol{\sigma})) \leq 2L_\ell \cdot 2^{72C_\ell} \cdot \max_{0 \leq j \leq L_\ell} \max_{\substack{\rho \in \{A,B\}_\bullet^{L_\ell} \\ \mathsf{bull}(\rho) = L_\ell - C_\ell}} \Pr(\mathsf{Delay}(\ell, \boldsymbol{\sigma}, i, j, \rho) \leq L_\ell). \tag{24}$$

We now fix an arbitrary $0 \leq j \leq L_\ell$ and $\rho \in \{A,B\}_\bullet^{L_\ell}$ such that $\mathsf{bull}(\rho) = L_\ell - C_\ell$ and upper bound the probability term on the right. To this end, we define the set $Z = \{0 \leq i < L_\ell \mid \rho_{i+1} \neq \bullet\}$ and observe that $|Z| = C_\ell$. We additionally, define, for $t \in [(1-\eta) \cdot L_\ell]$, the random variable:

$$\mathsf{X}_t = \mathsf{Emb}\big(\boldsymbol{\sigma}_{>i-D_\ell}, \mathsf{ins}\big(\boldsymbol{\sigma}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\big), t\big) - \mathsf{Emb}\big(\boldsymbol{\sigma}_{>i-D_\ell}, \mathsf{ins}\big(\boldsymbol{\sigma}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\big), t-1\big). \tag{25}$$

Observe from Observation 3.1 that for all $t \in [(1-\eta) \cdot L_\ell]$, the random variable X_t only takes values in $\mathbb{N}$. We claim that:

Claim 6.8. *For all $t \in [(1-\eta) \cdot L_\ell]$ and all $x_1, x_2, \cdots, x_{t-1} \in \mathbb{N}$, we have:*

$$\Pr(\mathsf{X}_t = 1 \mid \forall t' \in [t-1] : \mathsf{X}_{t'} = x_{t'}) \leq 1 - \frac{1}{2} \cdot \mathbb{1}\left(\sum_{t'=1}^{t-1} x_{t'} \notin Z\right) \cdot \mathbb{1}\left(\sum_{t'=1}^{t-1} x_{t'} \leq L_\ell - 1\right).$$

We prove Claim 6.8 later in Section 6.3.1 but assuming it for now, we can continue Eq. (24) as:

$$\begin{aligned}
\Pr(i \in \mathsf{Pred}_\ell(\boldsymbol{\sigma})) &\leq 2L_\ell \cdot 2^{72C_\ell} \cdot \max_{0 \leq j \leq L_\ell} \max_{\substack{\rho \in \{A,B\}_\bullet^{L_\ell} \\ \mathsf{bull}(\rho) = L_\ell - C_\ell}} \Pr\left(\sum_{t=1}^{(1-\eta)\cdot L_\ell} \mathsf{X}_t \leq L_\ell\right) \\
&\leq 2L_\ell \cdot 2^{72C_\ell} \cdot 2^{-\frac{2L_\ell}{5} + \eta \cdot L_\ell \cdot \log \frac{12}{\eta} + |Z|} && \left(\text{Lemma 6.2 with } \delta = \tfrac{\eta}{1-\eta}\right) \\
&\leq 2L_\ell \cdot 2^{72C_\ell} \cdot 2^{-\frac{L_\ell}{3} + C_\ell} && (\text{As } \eta = 10^{-5} \text{ and } |Z| = C_\ell) \\
&\leq 2^{-\frac{L_\ell}{4}}. && (\text{Eq. (21)})
\end{aligned}$$

6.3.1 Proof of Claim 6.8

We now prove Claim 6.8 that was used in the proof of Lemma 6.7.

Proof of Claim 6.8. For convenience, we define $x_{<t} = \sum_{t'=1}^{t-1} x_{t'}$. As the claim is trivial otherwise, we can assume that $x_{<t} \notin Z$ and $x_{<t} \leq L_\ell - 1$. By the definition of Z, this means that $\rho_{x_{<t}+1} = \bullet$, which implies that $\mathsf{bull}(\rho_{\leq x_{<t}+1}) = \mathsf{bull}(\rho_{\leq x_{<t}}) + 1$.

Define the value $z = i + j + \mathsf{bull}(\rho_{\le x_{<t}})$. Observe that:

Claim 6.9. *It holds that $i + t - D_\ell \le z$.*

Proof. We have:

$$\begin{aligned}
i + t - D_\ell &\le i + j + t - D_\ell && \text{(As } j \ge 0\text{)}\\
&\le i + j + t - C_\ell - 1 && \text{(Eq. (21))}\\
&\le i + j - C_\ell + x_{<t} && \text{(As } x_{t'} \in \mathbb{N}\text{)}\\
&\le i + j + \mathsf{bull}(\rho) - L_\ell + x_{<t} && \text{(As } \mathsf{bull}(\rho) = L_\ell - C_\ell\text{)}\\
&\le i + j + \mathsf{bull}(\rho) - |\rho_{>x_{<t}}| && \text{(As } x_{<t} \le L_\ell - 1\text{)}\\
&\le i + j + \mathsf{bull}(\rho) - \mathsf{bull}(\rho_{>x_{<t}}) &&\\
&\le i + j + \mathsf{bull}(\rho_{\le x_{<t}}) &&\\
&\le z. &&
\end{aligned}$$

□

Next, note that whether or not the event $\forall t' \in [t-1] : \mathsf{X}_{t'} = x_{t'}$ is determined by the value of $\boldsymbol{\sigma}_{\le z}$. Indeed, for any two strings $\sigma^{(1)}$ and $\sigma^{(2)}$ that agree on the first z coordinates, say $\sigma^{(1)}_{\le z} = \sigma^{(2)}_{\le z} = \tau$, we can apply Lemma 3.4 and Claim 6.9 with $\sigma = \sigma^{(1)}_{>i-D_\ell}$, $\tau = \sigma^{(2)}_{>i-D_\ell}$, $\sigma' = \tau' = \mathsf{ins}(\tau_{>i+j}, \rho_{\le x_{<t}})$ to get that for all $t' \in \{0\} \cup [t-1]$, we have:

$$\mathsf{Emb}\Big(\sigma^{(1)}_{>i-D_\ell}, \mathsf{ins}(\tau_{>i+j}, \rho_{\le x_{<t}}), t'\Big) \le x_{<t} \iff \mathsf{Emb}\Big(\sigma^{(2)}_{>i-D_\ell}, \mathsf{ins}(\tau_{>i+j}, \rho_{\le x_{<t}}), t'\Big) \le x_{<t}.$$

By the definition of ins, this gives:

$$\begin{aligned}
&\mathsf{Emb}\Big(\sigma^{(1)}_{>i-D_\ell}, \mathsf{ins}_{\le x_{<t}}\Big(\sigma^{(1)}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\Big), t'\Big) \le x_{<t}\\
&\qquad\qquad \iff \mathsf{Emb}\Big(\sigma^{(2)}_{>i-D_\ell}, \mathsf{ins}_{\le x_{<t}}\Big(\sigma^{(2)}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\Big), t'\Big) \le x_{<t}.
\end{aligned}$$

Again applying Lemma 3.4, we get for all $t' \in \{0\} \cup [t-1]$:

$$\begin{aligned}
&\mathsf{Emb}\Big(\sigma^{(1)}_{>i-D_\ell}, \mathsf{ins}\Big(\sigma^{(1)}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\Big), t'\Big) \le x_{<t}\\
&\qquad\qquad \iff \mathsf{Emb}\Big(\sigma^{(2)}_{>i-D_\ell}, \mathsf{ins}\Big(\sigma^{(2)}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\Big), t'\Big) \le x_{<t}.
\end{aligned}$$

By Eq. (25), this means that:

$$\forall t' \in [t-1] : \mathsf{X}_{t'} = x_{t'} \mid \boldsymbol{\sigma} = \sigma^{(1)} \iff \forall t' \in [t-1] : \mathsf{X}_{t'} = x_{t'} \mid \boldsymbol{\sigma} = \sigma^{(2)},$$

as claimed. As we showed that whether or not the event $\forall t' \in [t-1] : \mathsf{X}_{t'} = x_{t'}$ is determined by the value of $\boldsymbol{\sigma}_{\le z}$, the claim follows if we show that for all $\tau \in \{A, B\}^z$ that $\forall t' \in [t-1] :$

$\mathsf{X}_{t'} = x_{t'}$ happens when $\boldsymbol{\sigma}_{\le z} = \tau$, we have:

$$\Pr(\mathsf{X}_t = 1 \mid \boldsymbol{\sigma}_{\le z} = \tau) \le \frac{1}{2}.$$

However, this is because

$$\begin{aligned}
\Pr(\mathsf{X}_t = 1 \mid \boldsymbol{\sigma}_{\le z} = \tau) &\le \Pr\big(\mathsf{Emb}\big(\boldsymbol{\sigma}_{>i-D_\ell}, \mathsf{ins}\big(\boldsymbol{\sigma}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\big), t\big) = x_{<t} + 1 \mid \boldsymbol{\sigma}_{\le z} = \tau\big) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Eq. (25) and choice of } \tau\text{)} \\
&\le \Pr\big(\mathsf{ins}_{x_{<t}+1}\big(\boldsymbol{\sigma}_{(i+j,i+j+L_\ell-C_\ell]}, \rho\big) = \boldsymbol{\sigma}_{i+t-D_\ell} \mid \boldsymbol{\sigma}_{\le z} = \tau\big) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(As } x_{<t} \le L_\ell - 1\text{)} \\
&\le \Pr\Big(\boldsymbol{\sigma}_{i+j+\mathsf{bull}\left(\rho_{\le x_{<t}+1}\right)} = \boldsymbol{\sigma}_{i+t-D_\ell} \mid \boldsymbol{\sigma}_{\le z} = \tau\Big) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Eq. (22) and } \rho_{x_{<t}+1} = \bullet\text{)} \\
&\le \Pr(\boldsymbol{\sigma}_{z+1} = \boldsymbol{\sigma}_{i+t-D_\ell} \mid \boldsymbol{\sigma}_{\le z} = \tau) \\
&\qquad \text{(As } \mathsf{bull}(\rho_{\le x_{<t}+1}) = \mathsf{bull}(\rho_{\le x_{<t}}) + 1 \text{ and } z = i + j + \mathsf{bull}(\rho_{\le x_{<t}})\text{)} \\
&\le \frac{1}{2}. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Claim 6.9)}
\end{aligned}$$

□

6.4 Strings With Small $\mathsf{Pred}_\ell(\cdot)$ Exist

Lemma 6.10. *For all integers $T > 0$, there exists $\sigma \in \{A, B\}^T$ such that for all $\ell > 0$, we have:*

$$|\mathsf{Pred}_\ell(\sigma)| \le 2^{-\frac{L_\ell}{8}} \cdot T.$$

Proof. It suffices to show that:

$$\Pr_{\boldsymbol{\sigma} \sim \{A,B\}^T}\Big(\exists \ell > 0 : |\mathsf{Pred}_\ell(\boldsymbol{\sigma})| > 2^{-\frac{L_\ell}{8}} \cdot T\Big) < 1.$$

This is because:

$$\begin{aligned}
\Pr_{\boldsymbol{\sigma} \sim \{A,B\}^T}\Big(\exists \ell > 0 : |\mathsf{Pred}_\ell(\boldsymbol{\sigma})| > 2^{-\frac{L_\ell}{8}} \cdot T\Big) &\le \sum_{\ell > 0} \Pr_{\boldsymbol{\sigma} \sim \{A,B\}^T}\Big(|\mathsf{Pred}_\ell(\boldsymbol{\sigma})| > 2^{-\frac{L_\ell}{8}} \cdot T\Big) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(Union bound)} \\
&\le \sum_{\ell > 0} \Pr_{\boldsymbol{\sigma} \sim \{A,B\}^T}\left(\sum_{i=1}^{T} \mathbb{1}(i \in \mathsf{Pred}_\ell(\boldsymbol{\sigma})) > 2^{-\frac{L_\ell}{8}} \cdot T\right) \\
&\le \sum_{\ell > 0} \sum_{i=1}^{T} \frac{\Pr_{\boldsymbol{\sigma} \sim \{A,B\}^T}(i \in \mathsf{Pred}_\ell(\boldsymbol{\sigma}))}{2^{-\frac{L_\ell}{8}} \cdot T} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(Markov inequality)}
\end{aligned}$$

$$\begin{aligned} &\leq \sum_{\ell > 0} 2^{-\frac{L_\ell}{8}} && \text{(Lemma 6.7)} \\ &\leq \frac{1}{2}. \end{aligned}$$

□

6.5 Structure of Long Subsequences

For the remainder of this section, readers may like to recall the definition of the set $\mathsf{E}_{\sigma,\sigma'}$ in Eq. (2). We borrow the following lemma from [Sch93].

Lemma 6.11 ([Sch93], Lemma 6)**.** *Let $T' > 0$ be an integer. Also, let $\mathcal{I}$ be an indexing set and a collection of pairs $\{t_i', t_i\}_{i \in \mathcal{I}}$ be given. Assume that $0 \leq t_i' < t_i \leq T'$ for all $i \in \mathcal{I}$. There exists a set $\mathcal{I}' \subseteq \mathcal{I}$ such that the intervals $\{(t_i', t_i]\}_{i \in \mathcal{I}'}$ are mutually disjoint and satisfy:*

$$\left| \bigcup_{i \in \mathcal{I}} (t_i', t_i] \right| \leq 2 \cdot \left| \bigcup_{i \in \mathcal{I}'} (t_i', t_i] \right|.$$

Lemma 6.12. *Let $\sigma, \sigma' \in \{A, B\}^*$ be such that σ is a subsequence of σ'. If $|\sigma'| \leq (1 + \eta^{20}) \cdot |\sigma|$, then:*

$$\left| \left\{ i' \in [|\sigma'|] \mid \exists 0 < k \leq |\sigma'| - i' : |\mathsf{E}_{\sigma,\sigma'} \cap (i', i' + k]| \leq \left(1 - \eta^8\right) \cdot k \right\} \right| \leq \eta^8 \cdot |\sigma'|.$$

Proof. For convenience, let S be the set in the lemma statement and define $T = |\sigma'|$, $T' = |\sigma'|$. Also, define the set of pairs:

$$\mathcal{I} = \left\{ (l, r) \mid 0 \leq l < r \leq T' \wedge |\mathsf{E}_{\sigma,\sigma'} \cap (l, r]| \leq \left(1 - \eta^8\right) \cdot |(l, r]| \right\}.$$

Observe that, if $i' \in S$, then $i' + 1 \in \bigcup_{(l,r) \in \mathcal{I}} (l, r]$. This implies that $|S| \leq \left| \bigcup_{(l,r) \in \mathcal{I}} (l, r] \right|$. Now, use Lemma 6.11 to get a $\mathcal{I}' \subseteq \mathcal{I}$ such that the intervals $(l, r]$, for $(l, r) \in \mathcal{I}'$ are pairwise disjoint and satisfy $|S| \leq 2 \cdot \left| \bigcup_{(l,r) \in \mathcal{I}'} (l, r] \right|$. We get:

$$\begin{aligned} \eta^{19} \cdot T' &\geq T' - T && \text{(As } T' \leq (1 + \eta^{20}) \cdot T\text{)} \\ &\geq |[T'] \setminus \mathsf{E}_{\sigma,\sigma'}| && \text{(As } \sigma \text{ is a subsequence of } \sigma'\text{)} \\ &\geq \sum_{(l,r) \in \mathcal{I}'} |(l, r] \setminus \mathsf{E}_{\sigma,\sigma'}| && \text{(As the intervals } (l, r]\text{, for } (l, r) \in \mathcal{I}' \text{ are pairwise disjoint)} \\ &\geq \sum_{(l,r) \in \mathcal{I}'} \eta^8 \cdot |(l, r]| && \text{(As } \mathcal{I}' \subseteq \mathcal{I}\text{)} \\ &\geq \eta^8 \cdot \left| \bigcup_{(l,r) \in \mathcal{I}'} (l, r] \right| && \text{(As the intervals } (l, r]\text{, for } (l, r) \in \mathcal{I}' \text{ are pairwise disjoint)} \end{aligned}$$

$$\geq \eta^9 \cdot |S|.$$

□

6.6 Proof of Theorem 4.2

Proof of Theorem 4.2. We define σ to be the string promised by Lemma 6.10. Thus, for all $\ell > 0$, we have:

$$|\mathsf{Pred}_\ell(\sigma)| \leq 2^{-\frac{L_\ell}{8}} \cdot T. \tag{26}$$

Fix an arbitrary $\sigma' \in \{A, B\}^*$ such that σ is a strong subsequence of σ' and let $T' = |\sigma'|$. Assume for the sake of contradiction that $T' < (1 + \eta^{20}) \cdot T$. As σ is a strong subsequence of σ', we must have $T' \geq T + 1$ and $|\mathsf{E}_{\sigma,\sigma'}| = T$. From these, we conclude that $2T \geq T' \geq T \geq \eta^{-20}$ and $|\mathsf{E}_{\sigma,\sigma'}| \geq (1 - \eta^{10}) \cdot T'$.

Next, we use Definition 3.7 to get a set $I \subseteq [T']$ such that $|I| \geq \frac{T'}{10}$ and for all $i' \in I$ we have that σ is a subsequence of $\sigma'_{-i'}$. Define the following sets:

$$\begin{aligned} I_1 &= [T'] \setminus [\lfloor 0.999T' \rfloor] \\ I_2 &= [T'] \setminus \mathsf{E}_{\sigma,\sigma'} \\ I_3 &= \left\{ i' \in [T'] \mid 0 < k \leq T' - i' : |\mathsf{E}_{\sigma,\sigma'} \cap (i', i' + k]| \leq \left(1 - \eta^8\right) \cdot k \right\} \end{aligned} \tag{27}$$

Also, define, for $\ell > 0$, the set:

$$I_{3+\ell} = \{ i' \in [T'] \mid \mathsf{bound}_\ell(i') \in \mathsf{Pred}_\ell(\sigma) \}, \tag{28}$$

where, for $i' \in [T']$, we define

$$\mathsf{bound}_\ell(i') = \max\{ i \in \{0\} \cup [T] \mid \mathsf{Emb}(\sigma, \sigma', i) \leq i' \} + (1 - \eta) \cdot L_{<\ell} + D_\ell - \ell. \tag{29}$$

We claim that:

Claim 6.13. *For all $\ell > 0$, it holds that:*

$$|I_{3+\ell} \setminus I_2| \leq 2^{-\frac{L_\ell}{8}} \cdot T.$$

Proof. Due to Eq. (26), it is sufficient to show that $\mathsf{bound}_\ell(\cdot)$ is a one to one function from the set $I_{3+\ell} \setminus I_2$ to the set $\mathsf{Pred}_\ell(\sigma)$. By Eq. (28), the function $\mathsf{bound}_\ell(\cdot)$ indeed maps the set $I_{3+\ell} \setminus I_2$ to the set $\mathsf{Pred}_\ell(\sigma)$. This function is also one-to-one, as if there exists $i'_1 < i'_2$ such that $\mathsf{bound}_\ell(i'_1) = \mathsf{bound}_\ell(i'_2)$, then, we have:

$$\max\{ i \in \{0\} \cup [T] \mid \mathsf{Emb}(\sigma, \sigma', i) \leq i'_1 \} = \max\{ i \in \{0\} \cup [T] \mid \mathsf{Emb}(\sigma, \sigma', i) \leq i'_2 \},$$

then in particular, there is no $i \in \{0\} \cup [T]$ such that $\mathsf{Emb}(\sigma, \sigma', i) = i'_2$, a contradiction to the fact that $i_2' \notin I_2$. □

Claim 6.14. *We have:*

$$\left|\bigcup_{j>0} I_j\right| \leq \frac{T'}{100}.$$

Proof. We have:

$$\begin{aligned}
\left|\bigcup_{j>0} I_j\right| &\leq |I_1| + |I_2| + |I_3| + \sum_{\ell>0} |I_{3+\ell} \setminus I_2| \\
&\leq \frac{T'}{100} + |I_2| + |I_3| + \sum_{\ell>0} |I_{3+\ell} \setminus I_2| && (\text{Eq. (27) and } T \geq \eta^{-20}) \\
&\leq \frac{2T'}{500} + |I_3| + \sum_{\ell>0} |I_{3+\ell} \setminus I_2| && (\text{Eq. (27) and } |\mathsf{E}_{\sigma,\sigma'}| \geq (1-\eta^{10}) \cdot T') \\
&\leq \frac{3T'}{500} + \sum_{\ell>0} |I_{3+\ell} \setminus I_2| && (\text{Eq. (27) and Lemma 6.12}) \\
&\leq \frac{3T'}{500} + \sum_{\ell>0} 2^{-\frac{L_\ell}{8}} \cdot T && (\text{Claim 6.13}) \\
&\leq \frac{4T'}{500}. && (\text{Eq. (21) and } 2T \geq T')
\end{aligned}$$

□

Conclude from Claim 6.14 and the fact that $|I| \geq \frac{T'}{10}$ that there exists an index $z' \in I \setminus \bigcup_{j>0} I_j$. We show that this leads to a contradiction. As $z' \in I$, we have that σ is a subsequence of $\sigma'_{-z'}$. Recall that this implies that $\left|\mathsf{E}\left(\sigma, \sigma'_{-z'}\right)\right| = T$ or, equivalently, $\mathsf{Emb}\left(\sigma, \sigma'_{-z'}, T\right) \leq T' - 1 < T'$. Henceforth, for notational convenience, we abbreviate $\mathsf{Emb}\left(\sigma, \sigma'_{-z'}, \cdot\right)$ to $\mathsf{Emb}^*(\cdot)$ and $\mathsf{E}\left(\sigma, \sigma'_{-z'}\right)$ to E^*. We also abbreviate $\mathsf{Emb}(\sigma, \sigma', \cdot)$ to $\mathsf{Emb}(\cdot)$ and $\mathsf{E}(\sigma, \sigma')$ to E.

We now use the fact that $z' \notin \bigcup_{j>0} I_j$ to get more information about z'. From Eq. (27), we get that $z' \leq 0.999T'$ and $z' \in \mathsf{E}$. Due to Eq. (2), this implies that there exists $z \in [T]$ such that $\mathsf{Emb}(z) = z'$. We claim that:

$$z \leq (1-\eta) \cdot T. \tag{30}$$

Indeed, if not, we have from Observation 3.1 that $0.999T' \geq z' \geq z \geq (1-\eta) \cdot T$, a contradiction to $T' < (1+\eta^{20}) \cdot T$. Next, Eq. (27) also says that for all $0 < k \leq T' - z'$, we have (as the left hand side is an integer):

$$|\mathsf{E} \cap (z', z'+k]| \geq \left\lceil \left(1-\eta^8\right) \cdot k \right\rceil. \tag{31}$$

Finally, use Eq. (28) and Observation 3.1 and $\mathsf{Emb}(z) = z'$ to get that, for all $\ell > 0$:

$$\mathsf{bound}_\ell(z') = z + (1-\eta) \cdot L_{<\ell} + D_\ell - \ell \notin \mathsf{Pred}_\ell(\sigma). \tag{32}$$

To derive a contradiction, we shall show that:

Lemma 6.15. *For all $\ell \geq 0$ such that $z \leq T - 3L_{\leq \ell}$, we have:*

$$\mathsf{Emb}^*(z + (1-\eta) \cdot L_{\leq \ell} - \ell) \geq z' + L_{\leq \ell}.$$

Before showing Lemma 6.15, we finish the proof of Theorem 4.2 by showing that it implies a contradiction. For this, define $\ell^* = \lfloor \log_{10^{10}}(\eta^4 T') \rfloor$ and note that $T' \geq \eta^{-20}$ implies that $\ell^* \geq 5$. We get from Eq. (21) that

$$L_{\leq \ell^*} \leq 2L_{\ell^*} \leq 2\eta^2 T' \qquad\qquad L_{\leq \ell^*} \geq L_{\ell^*} \geq \eta^4 T'. \tag{33}$$

Due to Eqs. (30) and (33), we can use Lemma 6.15 with ℓ^* to get:

$$\begin{aligned}
\mathsf{Emb}^*(T) &\geq \mathsf{Emb}^*(z + (1-\eta) \cdot L_{\leq \ell^*} - \ell^*) + T - z - (1-\eta) \cdot L_{\leq \ell^*} + \ell^* && \text{(Observation 3.1)} \\
&\geq z' + \eta \cdot L_{\leq \ell^*} + T - z \\
&\geq \eta \cdot L_{\leq \ell^*} + T && \text{(As Observation 3.1 implies } z' \geq z\text{)} \\
&\geq \eta^5 T' + T. && \text{(As } L_{\leq \ell^*} \geq \eta^4 T'\text{)}
\end{aligned}$$

As we know that $\mathsf{Emb}^*(T) < T'$, this contradicts $T' < (1 + \eta^{20}) \cdot T$. □

It remains to show Lemma 6.15.

Proof of Lemma 6.15. We prove the lemma by induction on ℓ. For the base case $\ell = 0$, we have $\mathsf{Emb}^*(z) > z' - 1$ by Lemma 3.4, and the result follows. For the inductive step, we show the result for $\ell > 0$ by assuming it holds for $\ell - 1$. By our assumption, we have:

$$\mathsf{Emb}^*(z + (1-\eta) \cdot L_{<\ell} - \ell + 1) \geq z' + L_{<\ell}. \tag{34}$$

We now claim that:

Claim 6.16. *We have $z' + 2L_{\leq \ell} \leq T'$.*

Proof. If not, we have:

$$\begin{aligned}
\mathsf{Emb}^*(T) &\geq \mathsf{Emb}^*(z + (1-\eta) \cdot L_{<\ell} - \ell + 1) + T - z - (1-\eta) \cdot L_{<\ell} + \ell - 1 \\
& && \text{(Observation 3.1)} \\
&\geq z' + \eta \cdot L_{<\ell} + T - z && \text{(Eq. (34) and } \ell > 0\text{)} \\
&\geq z' + \eta \cdot L_{<\ell} + 2L_{\leq \ell} && \text{(Assumption in the lemma)} \\
&\geq T' + \eta \cdot L_{<\ell}.
\end{aligned}$$

As we know that $\mathsf{Emb}^*(T) < T'$, this is a contradiction. □

Claim 6.17. *It holds that* $D_\ell \leq \mathsf{bound}_\ell(z') \leq T - 2L_{\leq \ell}$.

Proof. For the first inequality, note that $\mathsf{bound}_\ell(z') \geq 1 + \frac{1}{2} \cdot L_{<\ell} - \ell \geq 0$. For the second, note that $\mathsf{bound}_\ell(z') \leq z + L_{<\ell} + D_\ell \leq z + L_{\leq \ell} \leq T - 2L_{\leq \ell}$. □

Claim 6.18. *There exists* $0 \leq j \leq L_\ell$ *and* $\rho \in \{A, B\}_{\bullet}^{L_\ell}$ *satisfying* $\mathsf{bull}(\rho) = L_\ell - C_\ell$ *and:*

$$\mathsf{ins}\big(\sigma_{(\mathsf{bound}_\ell(z')+j, \mathsf{bound}_\ell(z')+j+L_\ell-C_\ell]}, \rho\big) = \sigma'_{(z'+L_{<\ell}, z'+L_{\leq \ell}]}.$$

Proof. We start by using Eq. (31) to derive the following inequalities (the condition in Eq. (31) is satisfied due to Claim 6.16):

$$\begin{aligned}\big\lceil (1-\eta^8) \cdot L_{\leq \ell} \big\rceil &\leq |\mathsf{E} \cap (z', z' + L_{\leq \ell}]|, \\ \big\lceil (1-\eta^8) \cdot L_{<\ell} \big\rceil &\leq |\mathsf{E} \cap (z', z' + L_{<\ell}]| \leq L_{<\ell},\end{aligned}$$

This implies that:

$$\begin{aligned}|\mathsf{E} \cap (z' + L_{<\ell}, z' + L_{\leq \ell}]| &\geq \big\lceil (1-\eta^8) \cdot L_{\leq \ell} - L_{<\ell} \big\rceil \\ &\geq \big\lceil (1-\eta^8) \cdot L_\ell - \eta^8 \cdot L_{<\ell} \big\rceil \\ &\geq \big\lceil (1-2\eta^8) \cdot L_\ell \big\rceil \\ &\geq L_\ell - \big\lfloor 2\eta^8 \cdot L_\ell \big\rfloor \qquad \text{(As for all } x\text{, we have } \lfloor -x \rfloor = -\lceil x \rceil) \\ &\geq L_\ell - C_\ell. \qquad \text{(Eq. (21))}\end{aligned}$$

We define j to be:

$$j = |\mathsf{E} \cap (z', z' + L_{<\ell}]| - (1-\eta) \cdot L_{<\ell} - D_\ell + \ell. \tag{35}$$

We now show that $0 \leq j \leq L_\ell$. For the first inequality, note using Eq. (21) that

$$j \geq \big\lceil (1-\eta^8) \cdot L_{<\ell} \big\rceil - (1-\eta) \cdot L_{<\ell} - D_\ell + 1 \geq \left\lceil \frac{\eta}{2} \cdot L_{<\ell} \right\rceil - D_\ell + 1 \geq 0.$$

For the second inequality, note that $j \leq L_{<\ell} + \ell \leq L_\ell$. Next, observe that due to Lemmas 3.2 and 3.3, we have that:

$$\mathsf{Emb}(z + |\mathsf{E} \cap (z', z' + L_{<\ell}]|) \leq z' + L_{<\ell} < \mathsf{Emb}(z + 1 + |\mathsf{E} \cap (z', z' + L_{<\ell}]|).$$

Equivalently, by Eqs. (32) and (35), we have:

$$\mathsf{Emb}(\mathsf{bound}_\ell(z') + j) \leq z' + L_{<\ell} < \mathsf{Emb}(\mathsf{bound}_\ell(z') + j + 1).$$

Thus, we can apply Lemma 6.5 to get a string $\rho \in \{A,B\}_{\bullet}^{L_\ell}$ such that $\mathsf{bull}(\rho) = L_\ell - C_\ell$ and:

$$\mathsf{ins}\Big(\sigma_{(\mathsf{bound}_\ell(z')+j,\mathsf{bound}_\ell(z')+j+L_\ell-C_\ell]}, \rho\Big) = \sigma'_{(z'+L_{<\ell}, z'+L_{\leq\ell}]},$$

as claimed. □

We continue our proof of Lemma 6.15. Due to Eq. (32), we have $\mathsf{bound}_\ell(z') \notin \mathsf{Pred}_\ell(\sigma)$. Due to Definition 6.6 and Claim 6.18 (recall that $D_\ell \leq \mathsf{bound}_\ell(z') \leq T - 2L_{\leq\ell}$ due to Claim 6.17), this means that:

$$\mathsf{Emb}\Big(\sigma_{>\mathsf{bound}_\ell(z')-D_\ell}, \sigma'_{(z'+L_{<\ell},z'+L_{\leq\ell}]}, (1-\eta)\cdot L_\ell\Big) > L_\ell.$$

By definition of $\sigma'_{-z'}$ and Eq. (32), we then get:

$$\mathsf{Emb}\Big(\sigma_{\geq z+(1-\eta)\cdot L_{<\ell}-\ell+1}, \big(\sigma'_{-z'}\big)_{[z'+L_{<\ell},z'+L_{\leq\ell})}, (1-\eta)\cdot L_\ell\Big) > L_\ell. \tag{36}$$

To continue, we recall Eq. (34) which says that $\mathsf{Emb}^*(z+(1-\eta)\cdot L_{<\ell}-\ell+1) \geq z' + L_{<\ell}$. As $\mathsf{Emb}^*(0) = 0$ and we have Observation 3.1, this means that there exists a unique $c \in [z+(1-\eta)\cdot L_{<\ell}-\ell+1]$ such that:

$$\mathsf{Emb}^*(c-1) \leq z' + L_{<\ell} - 1 < \mathsf{Emb}^*(c). \tag{37}$$

Now, assume for the sake of contradiction that $\mathsf{Emb}^*(z+(1-\eta)\cdot L_{\leq\ell}-\ell) < z' + L_{\leq\ell}$. Together with Eq. (37) and Observation 3.1, this means that

$$|\mathsf{E}^* \cap [z'+L_{<\ell}, z'+L_{\leq\ell})| \geq z+(1-\eta)\cdot L_{\leq\ell}-\ell+1-c.$$

Applying Lemma 3.6 with $a = z+(1-\eta)\cdot L_{<\ell}-\ell+1-c$ and $b = z+(1-\eta)\cdot L_{\leq\ell}-\ell+1-c$, we get:

$$\mathsf{Emb}\Big(\sigma_{\geq z+(1-\eta)\cdot L_{<\ell}-\ell+1}, \big(\sigma'_{-z'}\big)_{[z'+L_{<\ell},z'+L_{\leq\ell})}, (1-\eta)\cdot L_\ell\Big) \leq L_\ell,$$

a contradiction to Eq. (36).

□

References

[AGS16] Shweta Agrawal, Ran Gelles, and Amit Sahai. Adaptive protocols for interactive communication. In *International Symposium on Information Theory (ISIT)*, pages 595–599, 2016. 5

[BKOS21] Assaf Ben-Yishai, Young-Han Kim, Or Ordentlich, and Ofer Shayevitz. A lower bound on the essential interactive capacity of binary memoryless symmetric channels. *IEEE Transactions on Information Theory*, 67(12):7639–7658, 2021. 4

[BR11] Mark Braverman and Anup Rao. Towards coding for maximum errors in interactive communication. In *Symposium on Theory of computing (STOC)*, pages 159–166, 2011. 4

[CS19] Gil Cohen and Shahar Samocha. Capacity-approaching deterministic interactive coding schemes against adversarial errors. *Electronic Colloquium on Computational Complexity: ECCC*, page 147, 2019. 1, 2, 4

[EHK20] Klim Efremenko, Elad Haramaty, and Yael Tauman Kalai. Interactive coding with constant round and communication blowup. In Thomas Vidick, editor, *Innovations in Theoretical Computer Science Conference (ITCS)*, volume 151, pages 7:1–7:34, 2020. 5

[Gel17] Ran Gelles. Coding for interactive communication: A survey. *Foundations and Trends® in Theoretical Computer Science*, 13(1–2):1–157, 2017. 2

[GH14] Ran Gelles and Bernhard Haeupler. Capacity of interactive communication over erasure channels and channels with feedback. In *Symposium on Discrete Algorithms (SODA)*, pages 1296–1311, 2014. 1, 2, 4

[GHK$^+$16] Ran Gelles, Bernhard Haeupler, Gillat Kol, Noga Ron-Zewi, and Avi Wigderson. Towards optimal deterministic coding for interactive communication. In *Symposium on Discrete Algorithms (SODA)*, pages 1922–1936. Society for Industrial and Applied Mathematics, 2016. 4

[Hae14] Bernhard Haeupler. Interactive channel capacity revisited. In *Foundations of Computer Science (FOCS)*, pages 226–235, 2014. 1, 2, 4, 5

[HSV18] Bernhard Haeupler, Amirbehshad Shahrasbi, and Ellen Vitercik. Synchronization strings: Channel simulations and interactive coding for insertions and deletions. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 107, pages 75:1–75:14, 2018. 1, 2, 4

[HV17] Bernhard Haeupler and Ameya Velingker. Bridging the capacity gap between interactive and one-way communication. In *Symposium on Discrete Algorithms (SODA)*, pages 2123–2142, 2017. 4

[KR13] Gillat Kol and Ran Raz. Interactive channel capacity. In *Symposium on Theory of computing (STOC)*, pages 715–724, 2013. 1, 2, 4, 5

[NW91] Noam Nisan and Avi Widgerson. Rounds in communication complexity revisited. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 419–429, 1991. 3, 5, 6, 26

[Pan13] Denis Pankratov. On the power of feedback in interactive channels. *Manuscript*, 2013. 1, 2, 4

[Sch92] Leonard J Schulman. Communication on noisy channels: A coding theorem for computation. In *Foundations of Computer Science (FOCS)*, pages 724–733, 1992. 1, 3, 4

[Sch93] Leonard J Schulman. Deterministic coding for interactive communication. In *Symposium on Theory of computing (STOC)*, pages 747–756, 1993. 4, 38

[Sha48] Claude E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001. Originally appeared in *Bell System Tech. J.* 27:379–423, 623–656, 1948. 1

A Information Theory Preliminaries

Recall that we use sans-serif letters to denote random variables. We reserve E to denote an arbitrary event. All random variables will be assumed to be discrete and we shall adopt the convention $0 \log \frac{1}{0} = 0$. All logarithms are taken with base 2.

A.1 Entropy

Definition A.1 (Entropy)**.** *The (binary) entropy of* X *is defined as:*

$$\mathbb{H}(\mathsf{X}) = \sum_{x \in \mathsf{supp}(\mathsf{X})} \Pr(x) \cdot \log \frac{1}{\Pr(x)}.$$

The entropy of X *conditioned on* E *is defined as:*

$$\mathbb{H}(\mathsf{X} \mid E) = \sum_{x \in \mathsf{supp}(\mathsf{X})} \Pr(x \mid E) \cdot \log \frac{1}{\Pr(x \mid E)}.$$

Definition A.2 (Conditional Entropy)**.** *We define the conditional entropy of* X *given* Y *and* E *as:*

$$\mathbb{H}(\mathsf{X} \mid \mathsf{Y}, E) = \sum_{y \in \mathsf{supp}(\mathsf{Y})} \Pr(y \mid E) \cdot \mathbb{H}(\mathsf{X} \mid \mathsf{Y} = y, E).$$

Henceforth, we shall omit writing the $\mathsf{supp}(\cdot)$ when it is clear from context.

Lemma A.3 (Chain Rule for Entropy)**.** *It holds for all* X, Y, Z *and* E *that:*

$$\mathbb{H}(\mathsf{XY} \mid \mathsf{Z}, E) = \mathbb{H}(\mathsf{X} \mid \mathsf{Z}, E) + \mathbb{H}(\mathsf{Y} \mid \mathsf{X}, \mathsf{Z}, E).$$

Proof. We have:

$$\begin{aligned}
\mathbb{H}(\mathsf{XY} \mid \mathsf{Z}, E) &= \sum_z \Pr(z \mid E) \cdot \mathbb{H}(\mathsf{XY} \mid z, E) \\
&= \sum_z \Pr(z \mid E) \cdot \sum_{x,y} \Pr(x, y \mid z, E) \cdot \log \frac{1}{\Pr(x, y \mid z, E)} \\
&= \sum_z \Pr(z \mid E) \cdot \sum_{x,y} \Pr(x, y \mid z, E) \cdot \left(\log \frac{1}{\Pr(x \mid z, E)} + \log \frac{1}{\Pr(y \mid x, z, E)} \right) \\
&= \mathbb{H}(\mathsf{X} \mid \mathsf{Z}, E) + \sum_{x,z} \Pr(x, z \mid E) \cdot \sum_y \Pr(y \mid x, z, E) \cdot \log \frac{1}{\Pr(y \mid x, z, E)} \\
&= \mathbb{H}(\mathsf{X} \mid \mathsf{Z}, E) + \mathbb{H}(\mathsf{Y} \mid \mathsf{X}, \mathsf{Z}, E).
\end{aligned}$$

□

Lemma A.4 (Conditioning reduces Entropy)**.** *It holds for all* X*,* Y*,* Z *and* E *that:*

$$\mathbb{H}(\mathsf{X} \mid \mathsf{Y}, \mathsf{Z}, E) \leq \mathbb{H}(\mathsf{X} \mid \mathsf{Z}, E).$$

Equality holds if and only if X *and* Y *are independent conditioned on* Z*,* E*.*

Proof. We have:

$$\begin{aligned}
\mathbb{H}(\mathsf{X} \mid \mathsf{Y}, \mathsf{Z}, E) &= \sum_{y,z} \Pr(y, z \mid E) \cdot \mathbb{H}(\mathsf{X} \mid \mathsf{Y} = y, \mathsf{Z} = z, E) \\
&= \sum_{x,y,z} \Pr(y, z \mid E) \cdot \Pr(x \mid y, z, E) \cdot \log \frac{1}{\Pr(x \mid y, z, E)} \\
&= \sum_{x,y,z} \Pr(x, z \mid E) \cdot \Pr(y \mid x, z, E) \cdot \log \frac{\Pr(y, z \mid E)}{\Pr(x, z \mid E) \cdot \Pr(y \mid x, z, E)} \\
&\leq \sum_{x,z} \Pr(x, z \mid E) \cdot \log \frac{\Pr(z \mid E)}{\Pr(x, z \mid E)} && \text{(Concavity of } \log(\cdot)\text{)} \\
&= \sum_z \Pr(z \mid E) \cdot \sum_x \Pr(x \mid z, E) \cdot \log \frac{1}{\Pr(x \mid z, E)} \\
&= \sum_z \Pr(z \mid E) \cdot \mathbb{H}(\mathsf{X} \mid \mathsf{Z} = z, E) \\
&= \mathbb{H}(\mathsf{X} \mid \mathsf{Z}, E).
\end{aligned}$$

□

A.2 Min-Entropy

Definition A.5 (Min-Entropy)**.** *The min-entropy of a discrete random variable* X *is*

$$\mathbb{H}_\infty(\mathsf{X}) = \min_{x : \Pr(x) > 0} \log \frac{1}{\Pr(x)}.$$

Fact A.6. *If the random variable* X *takes values in the set* Ω*, it holds that*

$$0 \leq \mathbb{H}_\infty(\mathsf{X}) \leq \mathbb{H}(\mathsf{X}) \leq \log|\Omega|$$

Lemma A.7. *Let* Ω, A, B *be (finite) sets and* X *be a random variable that takes values in the set* Ω*. Let* $f : \Omega \to A$ *and* $g : \Omega \to B$ *be functions. For an event* E *and* $t > 0$*, define the set:*

$$B' = \{b \in B \mid \mathbb{H}_\infty(f(\mathsf{X}) \mid E, g(\mathsf{X}) = b) \leq \mathbb{H}_\infty(f(\mathsf{X}) \mid E) - t\}.$$

It holds that:

$$\Pr(g(\mathsf{X}) \in B' \mid E) \leq |B| \cdot 2^{-t}.$$

Proof. For all $a \in A$, $b \in B$, we have by the chain rule that:

$$\Pr(g(\mathsf{X}) = b \mid E) \cdot \Pr(f(\mathsf{X}) = a \mid E, g(\mathsf{X}) = b) \leq \Pr(f(\mathsf{X}) = a \mid E).$$

Maximizing over all a and using Definition A.5, we have:

$$\Pr(g(\mathsf{X}) = b \mid E) \cdot 2^{-\mathbb{H}_\infty(f(\mathsf{X}) \mid E, g(\mathsf{X}) = b)} \leq 2^{-\mathbb{H}_\infty(f(\mathsf{X}) \mid E)}.$$

This means that for all $b \in B'$, we have:

$$\Pr(g(\mathsf{X}) = b \mid E) \leq 2^{-t}.$$

We get:

$$\Pr(g(\mathsf{X}) \in B' \mid E) = \sum_{b \in B'} \Pr(g(\mathsf{X}) = b \mid E) \leq |B'| \cdot 2^{-t} \leq |B| \cdot 2^{-t}.$$

□

A.3 KL Divergence

Definition A.8 (KL Divergence)**.** *If* μ, ν *are two distributions over the same (finite) set* Ω*, the Kullback-Leibler (KL) Divergence between* μ *and* ν *is defined as:*

$$\mathbb{D}(\mu \mid\mid \nu) = \sum_{\omega \in \Omega} \mu(\omega) \cdot \log \frac{\mu(\omega)}{\nu(\omega)}.$$

For a finite non-empty set S, we shall use $\mathcal{U}(S)$ to denote the uniform distribution over S. We omit S from the notation when it is clear from the context. We use $\mathsf{dist}(\mathsf{X} \mid E)$ to denote the distribution of the random variable X conditioned on the event E.

Lemma A.9. *Let* X *be a random variable uniformly distributed over a set* Ω *and* $S \subseteq \Omega$ *be given:*

$$\mathbb{D}(\mathsf{dist}(\mathsf{X} \mid \mathsf{X} \in S) \mid\mid \mathcal{U}) = \log \frac{|\Omega|}{|S|}.$$

Proof. As X is distributed uniformly, we have:

$$\mathbb{D}(\mathsf{dist}(\mathsf{X} \mid \mathsf{X} \in S) \mid\mid \mathcal{U}) = \sum_{x \in S} \frac{1}{|S|} \cdot \log \frac{|\Omega|}{|S|} = \log \frac{|\Omega|}{|S|}. \qquad \square$$

Lemma A.10. *It holds for all* X *and* E *that:*

$$\mathbb{D}(\mathsf{dist}(\mathsf{X} \mid E) \mid\mid \mathcal{U}) = \log(|\mathsf{supp}(\mathsf{X})|) - \mathbb{H}(\mathsf{X} \mid E).$$

Proof. We have:

$$\begin{aligned}\mathbb{D}(\mathsf{dist}(\mathsf{X} \mid E) \mid\mid \mathcal{U}) &= \sum_{x \in \mathsf{supp}(\mathsf{X})} \Pr(x \mid E) \cdot \log(\Pr(x \mid E) \cdot |\mathsf{supp}(\mathsf{X})|) \\ &= \sum_{x \in \mathsf{supp}(\mathsf{X})} \Pr(x \mid E) \cdot \log \Pr(x \mid E) + \sum_{x \in \mathsf{supp}(\mathsf{X})} \Pr(x \mid E) \cdot \log(|\mathsf{supp}(\mathsf{X})|) \\ &= \log(|\mathsf{supp}(\mathsf{X})|) - \mathbb{H}(\mathsf{X} \mid E). \qquad \square\end{aligned}$$

Lemma A.11. *It holds for all* X, Y *and* E *that:*

$$\mathbb{D}(\mathsf{dist}(\mathsf{XY} \mid E) \mid\mid \mathcal{U}) \geq \mathbb{D}(\mathsf{dist}(\mathsf{X} \mid E) \mid\mid \mathcal{U}) + \mathbb{D}(\mathsf{dist}(\mathsf{Y} \mid E) \mid\mid \mathcal{U}).$$

Proof. We have:

$$\begin{aligned}\mathbb{D}(\mathsf{dist}(\mathsf{XY} \mid E) \mid\mid \mathcal{U}) &= \log(|\mathsf{supp}(\mathsf{X})|) + \log(|\mathsf{supp}(\mathsf{Y})|) - \mathbb{H}(\mathsf{XY} \mid E) && \text{(Lemma A.10)} \\ &\geq \log(|\mathsf{supp}(\mathsf{X})|) + \log(|\mathsf{supp}(\mathsf{Y})|) - \mathbb{H}(\mathsf{X} \mid E) - \mathbb{H}(\mathsf{Y} \mid E) \\ & && \text{(Lemma A.3 and Lemma A.4)} \\ &\geq \mathbb{D}(\mathsf{dist}(\mathsf{X} \mid E) \mid\mid \mathcal{U}) + \mathbb{D}(\mathsf{dist}(\mathsf{Y} \mid E) \mid\mid \mathcal{U}). && \text{(Lemma A.10)}\end{aligned}$$

□

A.4 Total Variation Distance

Definition A.12 (Total variation distance)**.** *Let* μ, ν *be two distributions over the same (finite) set* Ω*. The total variation distance between* μ *and* ν *is defined as:*

$$\|\mu - \nu\|_{\mathrm{TV}} = \max_{\Omega' \subseteq \Omega} \sum_{\omega \in \Omega'} \mu(\omega) - \nu(\omega).$$

Fact A.13 (Pinsker's inequality)**.** *Let* μ, ν *be two distributions over the same set* Ω*. It holds that:*

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} \cdot \mathbb{D}(\mu \mid\mid \nu)}.$$