Interactive Channel Capacity

Gillat Kol* Ran Raz†

Abstract

We study the interactive channel capacity of an ϵ-noisy channel. The interactive channel capacity $C(\epsilon)$ is defined as the minimal ratio between the communication complexity of a problem (over a non-noisy channel), and the communication complexity of the same problem over the binary symmetric channel with noise rate ϵ, where the communication complexity tends to infinity.

Our main result is the upper bound $C(\epsilon) \leq 1 - \Omega\left(\sqrt{\mathbf{H}(\epsilon)}\right)$. This compares with Shannon's non-interactive channel capacity of $1 - \mathbf{H}(\epsilon)$. In particular, for a small enough ϵ, our result gives the first separation between interactive and non-interactive channel capacity, answering an open problem by Schulman [6].

We complement this result by the lower bound $C(\epsilon) \geq 1 - O\left(\sqrt{\mathbf{H}(\epsilon)}\right)$, proved for the case where the players take alternating turns.

1 Introduction

Few papers in the history of science have affected the way that people think in so many branches of science, as profoundly as Shannon's 1948 paper "A Mathematical Theory of Communication" [9]. One of the gems in that paper is an exact formula for the channel capacity of any communication channel. For example, for the binary symmetric channel with noise rate ϵ, the channel capacity is $1 - \mathbf{H}(\epsilon)$, where $\mathbf{H}$ denotes the binary entropy function. This means that one can reliably communicate n bits, with a negligible probability of error, by sending $\left(\frac{1}{1-\mathbf{H}(\epsilon)}\right) \cdot n + o(n)$ bits over the channel.

*Weizmann Institute of Science. Part of this work was done when the author was a visiting student at Princeton University and the Institute for Advanced Study. Research supported by an Israel Science Foundation grant, and by NSF grant number CCF-0832797.

†Weizmann Institute of Science. Part of this work was done when the author was a member at the Institute for Advanced Study, Princeton. Research supported by an Israel Science Foundation grant, by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation, and by NSF grants number CCF-0832797, DMS-0835373.

In this paper we study the *interactive channel capacity* of the binary symmetric channel. For a communication complexity problem f with communication complexity $\mathsf{CC}(f) = n$, we denote by $\mathsf{CC}_\epsilon(f)$ the expected number of communication bits needed to solve f probabilistically, with negligible error, where the communication is over the binary symmetric channel with noise rate ϵ. We define the channel capacity $C(\epsilon)$ by[1]

$$\lim_{n\to\infty} \min_{\{f:\mathsf{CC}(f)=n\}} \left\{ \frac{n}{\mathsf{CC}_\epsilon(f)} \right\}.$$

In 1992, Schulman showed how to translate any interactive communication protocol to an equivalent noise-resilient protocol that runs over the binary symmetric channel, with only a constant overhead in the communication complexity [6]. This shows that for any $\epsilon < 0.5$, the channel capacity $C(\epsilon) > 0$. Schulman's work was followed by a flow of other works [7, 8, 4, 5, 3, 2], culminating in the recent elegant result by Brakerski and Kalai [1] that shows how to efficiently simulate any interactive protocol in the presence of constant-rate adversarial noise, with only a constant overhead in the communication complexity.

However, none of these works gives any bound on the value of $C(\epsilon)$. Potentially, each of the above mentioned protocols gives a lower bound for $C(\epsilon)$, but since the protocols were typically proved with large constants and without paying attention to constants, these bounds are not explicit, and seem to be relatively weak. As for upper bounds, the only previously known upper bound on the value of $C(\epsilon)$ is the non-interactive capacity, $1-\mathbf{H}(\epsilon)$, proved by Shannon.

As in many previous works, for simplicity, we limit the discussion to protocols with "predetermined order of communication". That is, when considering deterministic protocols (or probabilistic protocols with fixed random strings), we will assume that it is known in advance which player sends a bit in each round. This is usually justified as follows. Since the noisy channel can change any transcript to any other transcript, if the order of communication is not predetermined, one can prove that with some non-zero probability both players will try to transmit a bit at the same time, contradicting the synchronous channel requirement that at each time-step exactly one player sends a bit.

[1] We note that $\mathsf{CC}(f)$ here stands for the *deterministic* communication complexity of f. It is as reasonable to make the same definition when considering the *probabilistic* communication complexity of f, that is, the number of communication bits needed to solve f probabilistically, with negligible error. All our results hold for both cases: Our upper bound considers a function with *deterministic* communication complexity n, and hence holds also for the probabilistic case. Our lower bound is proved by simulating *probabilistic* communication complexity protocols, and hence holds also for the deterministic case.

1.1 Our Results

Our main result is the first upper bound on the value of $C(\epsilon)$:

$$C(\epsilon) \leq 1 - \Omega\left(\sqrt{\mathbf{H}(\epsilon)}\right).$$

In particular, for small enough ϵ, this gives the first separation between interactive and non-interactive channel capacity, answering an open problem by Schulman [6].

We complement this result by a tight lower bound of

$$C(\epsilon) \geq 1 - O\left(\sqrt{\mathbf{H}(\epsilon)}\right),$$

for the alternating-turns case, where the first player sends bits in odd time-steps and the second player sends bits in even time-steps. More generally, the lower bound is obtained for the case where the pattern of turns taken by the two players is periodic with a small period. Moreover, the lower bound is valid for any pattern of turns in the asynchronous case, where one doesn't require that at each time-step exactly one player sends a bit, but rather the less restricting requirement that if both players try to send bits at the same time these bits are lost. As for the upper bound, while the main ideas of the proof may be valid for the asynchronous case as well, since the details of the proof are complicated in any case, we focus in this work on the synchronous case. We note however that our proof doesn't rely on any artifact of the synchronous channel and the ideas seem to be relevant for other cases as well.

To summarize, while there are still gaps between the validity models for the upper bound and the lower bound, they give a strong evidence that (in some models) the asymptotical behavior of $C(\epsilon)$, when ϵ tends to 0, is $1-\Theta\left(\sqrt{\mathbf{H}(\epsilon)}\right)$. This compares to the non-interactive capacity of $1 - \mathbf{H}(\epsilon)$.

Remark 1.1. *The exact type of channel considered may be important. Besides the case of predetermined order of communication, one can consider several other cases: The* alternating *case, where the players send bits alternately, or more generally, the* periodic *case, where the pattern of turns taken by the players is any periodic pattern with a small period; the* asynchronous *case, where if both players send bits at the same time these bits are lost; and the* two-channel *case, where each player can send bits over a separate channel whenever she wants, and we only count the number of bits that were actually sent.*

An interesting case, that captures a large class of protocols, is the case where we allow any protocol where the pattern of turns taken by the two players is periodic with a small period. We note that for this class of protocols (specifically, where we allow any pattern with period

of length up to $O\left(\sqrt{\epsilon^{-1}\log(\epsilon^{-1})}\right)$*, the upper bound and the lower bound are both valid (and match).*

1.2 Techniques

1.2.1 Upper Bound

Our upper bound is proved by considering the pointer jumping game on 2^k-ary trees of depth d, with edges directed from the root to the leaves. The pointer jumping game is a communication game with two parties. We think of the vertices in even layers of the tree as "owned" by the first player, and the vertices in odd layers of the tree as "owned" by the second player. Each player gets as an input exactly one edge going out of every vertex that she owns. We denote by X the input for the first player, and by Y the input for the second player. For a pair of inputs, X, Y, there exists a unique leaf of the tree, reachable from the root using the edges of both X, Y. The players' mutual goal is to find that leaf.

We fix the noise rate of the channel to be $\epsilon = \Theta\left(\frac{\log(k)}{k^2}\right)$. We think of ϵ as a fixed small constant, and the depth of the tree d tends to infinity. We consider probabilistic communication complexity protocols that run over the binary symmetric channel with noise rate ϵ and solve the pointer jumping game with probability of error δ. We prove a lower bound on the expected number of bits communicated by the players during the execution of any such protocol. Our lower bound shows that the expected number of bits communicated is at least

$$d \cdot (k + \Omega(\log(k))) \cdot (1 - 2\delta) - O(k).$$

Fixing $\delta = o(1)$, and since the communication complexity of the problem on a non-noisy channel is $\approx kd$, the obtained upper bound on the channel capacity is

$$1 - \Omega\left(\frac{\log(k)}{k}\right) = 1 - \Omega\left(\sqrt{\epsilon\log(\epsilon^{-1})}\right) = 1 - \Omega\left(\sqrt{\mathbf{H}(\epsilon)}\right).$$

The very high level idea that we use for the proof of the lower bound on the communication complexity of the pointer jumping game is very simple. Suppose that the two players want to solve the pointer jumping game, over the noisy channel, with as low communication as possible. The most reasonable thing to do is to start by having the first player sending the first edge in the path to the correct leaf, that is, the edge leaving the root. Note that this edge is part of the input X. Sending this edge requires a communication of k bits, and note that the probability that one of these bits is received incorrectly is $\Theta\left(\frac{\log(k)}{k}\right)$. At this point the players are faced with a dilemma. If they proceed by having the second player sending the k bits of the second edge, then with probability of $\Theta\left(\frac{\log(k)}{k}\right)$ these bits are lost

because the second player didn't have the correct value of the first edge. Thus, in expectation $\Theta(\log(k))$ bits are lost. On the other hand, if they proceed by having the first player sending additional bits in order to correct a possible error in the first k bits, then with probability close to 1 these additional bits are wasted, because the second player already had the correct value of the first edge.

While this seems like a simple and natural approach, attempts to formulate it must deal with several difficulties. First note that in order to obtain a meaningful lower bound, we need to show that the protocol "wastes" $\Omega(\log(k))$ bits not only once, but rather $\approx d$ times. The first time is easy because we can assume that the inputs X, Y are uniformly distributed on the set of all possible inputs. However, after conditioning the input distributions on a partial transcript of the protocol, the conditional distributions may be arbitrary. This means that some information is known about both inputs and in particular some information is known about the next edge that we consider. An important question is how to measure the amount of "information" known about the next edge, and more generally how to measure the amount of "progress" made by the two players towards the solution.

A first attempt may be to measure the information known on the next edge by Shannon entropy or by a variant such as relative entropy. The main problem with this approach is that even if the entropy of the edge is still large, say $k^{0.1}$ bits, it is still possible that a certain value is obtained with probability close to 1. Thus, the other player can guess that edge with high probability even though the entropy is still relatively large. A second attempt may be to measure the information known on the next edge by min-entropy or by the logarithm of the l_2 norm. The main problem with this approach is that these measures are not sub-additive. Therefore, the other player doesn't necessarily need to know the current edge in order to give information about the next edge, as she can give information about several branches of the tree simultaneously.

In addition to these difficulties, recall that we are proving a probabilistic lower bound so the probability of error must be taken into account, and that probability may be different on different branches of the communication protocol. Moreover, we are trying to prove a very tight lower bound, up to second order terms, and not up to a multiplicative constant as is usually done in probabilistic communication complexity.

To deal with all these issues we measure the progress made by the protocol, by several different measures. Given input distributions P_X, P_Y for the inputs of the two players, we denote by I_1 the relative entropy of P_X with respect to the uniform distribution and by I_2 the relative entropy of P_Y with respect to the uniform distribution. We denote $I = I_1 + I_2$. We denote by κ the min-entropy of the first edge. We say that a distribution P_X (or P_Y) is *flat* if it is roughly uniform on a subset of inputs. More precisely, a distribution is flat if

it gives any two elements the same probability, up to a multiplicative factor of $2^{0.01k}$. We say that the game is *nice* if $I_1 \leq 10k$; $I_2 \leq 20k$; $\kappa \geq 0.5k$; and P_X, P_Y are both flat. We inductively bound the communication complexity of any nice game by

$$d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - (k - \kappa) - 100k,$$

and the communication complexity of any game (not necessarily nice) by

$$d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - 100I - 1000k.$$

These two lemmas are proven simultaneously, by a mutual recursion, where each lemma assumes that the other lemma is correct for depth $d' < d$. Hence, both lemmas are correct for every d.

The proofs of the two lemmas are quite involved. To prove the first lemma (which is the main lemma and the more challenging one), we use an adversarial argument, where at each step we consider a block of the next $t = \kappa + 0.25k$ bits transmitted by the protocol. We separate to cases according to the number of bits transmitted by each player. Denote by t_1 the number of bits in that block that were sent by the first player. If $t_1 < \kappa + 0.5\log(k)$, then since the channel is noisy, the second player can still not guess with high enough probability the exact value of the first edge and then the $t - t_1$ bits that she sent are wasted with non-negligible probability. On the other hand if $t_1 \geq \kappa + 0.5\log(k)$, then the first player wasted $0.5\log(k)$ bits, since in our measure for the progress made by the protocol we measure the amount of information that is known about the first edge by $k - \kappa$.

In order to make this argument work, we need to "voluntarily" reveal to the two players some information about their inputs, even though that information is not given by the transcript of the communication protocol. This is done in order to make sure that the remaining game, that is, the game after conditioning on the partial transcript of the protocol and the additional information that we reveal, is nice with high probability. If the game that we get is nice, we recursively apply the first lemma and if it is not nice we recursively apply the second lemma.

As explained above, a major difficulty is that no measure of information is completely satisfying for our purpose. Shannon entropy has the disadvantage that even if the entropy is large it is possible that the variable can be guessed with high probability. Min-entropy and other similar measures have the disadvantage that they are not sub-additive. An idea that we extensively use in the proof of both lemmas is to "flatten" a distribution, that is, to make it flat. This is done by revealing to the two players the "flattening" values of certain variables. The flattening value is just a rounded estimate of the probability for a random

variable to get the value that it actually gets. By revealing the flattening value, we partition the support of a distribution so that the distribution is presented as a convex combination of flat distributions. Working with flat distributions is significantly easier since the entropy and min-entropy of a flat distribution are almost equal, so one can use min-entropy and switch to entropy whenever subadditivity is needed.

1.2.2 Lower Bound

We show that for any communication protocol Π of length n (where the players send bits in an alternating order), there exists a simulating protocol A that simulates Π over the binary symmetric channel with noise rate ϵ. The simulating protocol communicates $n\left(1+O\left(\sqrt{\mathbf{H}(\epsilon)}\right)\right)$ bits, and allows the players to retrieve the transcript of Π, except with probability negligible in n.

By fixing the random string for Π we can assume without loss of generality that Π is deterministic. Denote by T the tree underlying the protocol Π (that is, the binary tree with vertices that correspond to the transcripts of Π). We will consider *partial simulating protocols* for Π, as follows. Before running a partial simulating protocol A, each of the players is assumed to have a vertex of the tree T. We call these vertices the start vertices. When A ends, each of the players holds another vertex of T. We call these vertices the end vertices. The end vertex of each of the players will be a descendant of her start vertex. In addition, if the players have the same start vertex, they reach the same end vertex with high probability. Moreover, if the two players have the same start vertex and the same end vertex then every edge on the path connecting the start vertex and the end vertex agrees with the execution of the protocol Π on the inputs X, Y of the two players. We denote the start vertices of the players by V_1 and V_2, and the end vertices by V_1'' and V_2''.

Fix $\epsilon = \frac{\log(k)}{k^2}$, for a sufficiently large constant k. We recursively construct a sequence of partial simulating protocols.

The protocol A_1 is defined as follows: In the first phase, the players run the protocol Π for k rounds, where each player runs Π starting from her start vertex. Denote by V_1' and V_2' the vertices in T reached by the players after the first phase.

The second phase is an error-detecting phase, where the players check if an error has occurred (that is, if at least one of the sent bits was received incorrectly). To do so, the players decide on a set $\mathcal{F}$ of $r = C\log(k)$ random hash functions $f : \{0,1\}^k \to \{0,1\}$, using the shared random string, where C is a large odd constant.

The players exchange the evaluation of the hash functions in $\mathcal{F}$ on the transcript of Π that they observed in the first phase, where each bit is sent C times. For every $f \in \mathcal{F}$, the first player computes the majority of the C (possibly noisy) copies of the bit that she

got from the second player, and compares the majority bit against her own evaluation. If the first player finds that all the r majority bits match her bits, she sets her end vertex to $V_1'' = V_1'$. Otherwise, she rolls-back and sets her end vertex to $V_1'' = V_1$. The second player operates the same.

The protocol A_{i+1} is defined as follows: In the first phase, the players run the protocol A_i, k consecutive times. The second phase is again an error-detecting phase, and it is similar to the second phase of A_1, except that the size of the set $\mathcal{F}$ of random hash functions is now $C^{i+1}\log(k)$ (instead of $C\log(k)$), and that each of the bits is sent C^{i+1} times (instead of just C times).

We show that for large enough s, the protocol A_s has a very small probability of error while the number of bits that it transmits is close to the expected number of bits of the protocol Π that are retrieved.

1.3 Organization of the Paper

The paper is organized as follows. In Section 2, we prove our lower bound on the interactive channel capacity of the binary symmetric channel with noise rate ϵ. As discussed above, this is done by presenting, for any communication protocol Π (where the players send bits in an alternating order), a simulating protocol A that simulates Π over the binary symmetric channel with noise rate ϵ. The result is stated in Theorem 1.

The rest of the paper is devoted for the upper bound on the interactive channel capacity of the binary symmetric channel with noise rate ϵ. As discussed above, this is done by proving a lower bound for the communication complexity of probabilistic protocols that solve the pointer jumping game over the binary symmetric channel with noise rate ϵ.

In Section 3, we present the pointer jumping game and the models of communication complexity that we consider. We state our main result (a lower bound for the probabilistic communication complexity of the pointer jumping game over the binary symmetric channel) in Theorem 2.

In Section 4, we give notation and preliminaries.

In Section 5, we give many lemmas that may be of independent interest and are used in our main proof. We present the notion of *flat distribution*, as well as the *"flattening value"* that is used in order to present a distribution as a convex combination of flat distributions. We prove several lemmas about the flattening value. We also prove several lemmas that bound the entropy loss that occurs when one presents a distribution as a convex combination of other distributions.

In Section 6, we present several operations on pointer jumping games that are used in

our main proof. This includes *conditioning a game on a feasible transcript*, and *reducing a game*.

In Section 7, we give the proof of Theorem 2, given two central lemmas, Lemma 7.2 and Lemma 7.3. As discussed above, these lemmas prove lower bounds on the communication complexity of *nice* pointer jumping games, and *general* pointer jumping games, respectively. Lemma 7.2 is proved in Section 8, and Lemma 7.3 is proved in Section 9. We note that the only place where we use the fact that the channel is noisy is in the proof of Claim 8.21. All other claims are true even if the channel is not noisy.

2 The Simulating Protocol (Lower Bound on the Channel Capacity)

In this section we are given a probabilistic communication protocol Π between two players, where the communication is over a non-noisy binary channel. Our goal is to construct a simulating protocol A that simulates Π over the binary symmetric channel with *noise rate* ϵ. We assume that $\epsilon = \frac{\log(k)}{k^2}$, for a sufficiently large constant k. The protocol A simulates Π in the sense that it allows the players to retrieve the transcript of Π.

We assume that the players in a communication protocol share a random string. We assume without loss of generality that the given protocol Π is deterministic, as the players of the simulating protocol can fix the random string used by Π to their own shared random string. For simplicity, we assume that Π stops after the same number of rounds in every execution. Denote this number by n. The simulating protocol A will also stop after the same number of rounds in every execution.

In this section, we restrict the discussion to the case where the two players in Π send bits in an alternating order. We will construct a simulating protocol A that also has the alternating order property. We note that the result could be extended to the case where the pattern of turns taken by the two players is periodic with a small period. Moreover, the result could be extended to include any pattern of turns, in the asynchronous channel case, where one doesn't require that at each time-step exactly one player sends a bit, but rather the less restricting requirement that if both players try to send bits at the same time then these bits are lost.

Theorem 1. *For any communication protocol* Π *of length* n*, as above, there exists a simulating protocol* A *that simulates* Π *over the binary symmetric channel with noise rate* ϵ*. The simulating protocol* A *communicates* $n \cdot \left(1 + O\left(\sqrt{\mathbf{H}(\epsilon)}\right)\right)$ *bits, and allows the players to retrieve the transcript of* Π*, except with probability negligible in* n*.*

Proof. Fix $\epsilon = \frac{\log(k)}{k^2}$, where k is a sufficiently large constant. Since the protocol Π is assumed without loss of generality to be deterministic, we can think of it as a binary tree T of depth n, with edges labeled by either 0 or 1 (where the two edges going out of the same inner vertex are labeled differently). We think of the vertices in even layers of the tree as "owned" by the first player, and the vertices in odd layers of the tree as "owned" by the second player. Each player gets as an input exactly one edge going out of every vertex that she owns. In each round of Π, the player that owns the current vertex sends to the other player the bit label of the unique edge in his input going out of the current vertex.

We start by describing *partial simulating protocols* for Π. Before running a partial simulating protocol A, each of the players is assumed to have a vertex of T. We call these vertices the start vertices. When A ends, each of the players holds another (possibly different) vertex of T. We call these vertices the end vertices. We require that the end vertex of each of the players is a descendant of her start vertex. In addition, if the players have the same start vertex, they reach the same end vertex with high probability. Moreover, if the two players have the same start vertex and the same end vertex then every edge on the path connecting the start vertex and the end vertex is contained in the input of one of the players, that is, the path agrees with the protocol Π.

We denote the start vertices of the players by V_1 and V_2, and the end vertices by V_1'' and V_2''. For a vertex V of T, we denote by $d(V)$ the depth of V in T, where the root has depth 0.

We measure the partial simulating protocol A using several parameters:

1. m is the number of bits communicated by the protocol in every execution.

2. α is the maximal probability that the players disagree on the end vertex, assuming that they agreed on the start vertex. Formally,

$$\alpha = \max_{\Pi,v} \Pr\left[V_1'' \neq V_2'' \mid V_1 = V_2 = v\right].$$

3. t is the minimal expected depth gain, assuming that the players agreed on the start vertex. Formally,

$$t = \min_{\Pi,v} \mathbf{E}\left[d(V_1'') - d(V_1) \mid V_1 = V_2 = v\right],$$

   (where the minimum is taken over protocols Π of infinite length, so that a leaf is never reached).

For example, we can consider the protocol A that runs Π for a single round. This protocol has parameters $m = 1$, $\alpha = \epsilon$, and $t = 1$.

We next recursively construct a sequence of partial simulating protocols $A_1, \ldots, A_s$ for Π, where $s = \lceil \log\log(n) \rceil$. The parameters of the protocol A_i are denoted m_i, α_i, t_i. We will

show that the parameters of the protocols in the sequence keep improving. Specifically, as i gets larger, m_i and t_i increase, while α_i decreases. We then construct the simulating protocol A using the protocol A_s.

Assume for simplicity and without loss of generality that k is even. We will assume that $d(V_1), d(V_2)$ are both odd or both even.

The protocol A_1. The protocol A_1 is defined as follows. In the first phase, the players run the protocol Π for k rounds, where each player runs Π starting from her start vertex. That is, the first player starts from the vertex V_1, and the second player starts from V_2. Denote by V_1' and V_2' the vertices in T reached by the players after the first phase.

The second phase is an error-detecting phase, where the players check if an error has occurred (that is, if at least one of the sent bits was received incorrectly). To do so, the players decide on a set $\mathcal{F}$ of $r = 101\log(k)$ random hash functions $f : \{0,1\}^k \to \{0,1\}$, using the shared random string. For concreteness, assume that each of the functions $f \in \mathcal{F}$ is obtained by randomly selecting $a \in \{0,1\}^k$, and setting $f(x) = \bigoplus_{i\in[k]} a_i \cdot x_i$.

The players exchange the evaluation of the hash functions in $\mathcal{F}$ on the transcript of Π that they observed in the first phase. Formally, for a pair of vertices V, V' of T, such that V' is a descendant of V, we denote by $P(V, V')$ the labels of the edges on the path connecting V and V'. For every $f \in \mathcal{F}$, the first player sends the bit $b_{f,1} = f(P(V_1, V_1'))$ to the second player 101 times. For every $f \in \mathcal{F}$, the second player sends the bit $b_{f,2} = f(P(V_2, V_2'))$ to the first player 101 times. (If $P(V_1, V_1')$ or $P(V_2, V_2')$ are shorter than k bits, the players pad).

For every $f \in \mathcal{F}$, the first player computes the majority of the 101 (possibly noisy) copies of the bit $b_{f,2}$ that she got from the second player, and compares the majority bit against her own $b_{f,1}$ bit. If the first player finds that all the r majority bits match her r bits, she sets her end vertex to $V_1'' = V_1'$. Otherwise, she rolls-back and sets her end vertex to $V_1'' = V_1$. The second player operates the same.

We calculate the parameters of the protocol A_1:

1. $m_1 = k + 2 \cdot 101^2 \log(k)$: The protocol Π is run for k rounds, and the error-detecting phase adds $2 \cdot 101 \cdot r = 2 \cdot 101^2 \cdot \log(k)$ rounds.

2. $\alpha_1 \leq k^{-20}$: Assume that the players agreed on the start vertex. They may disagree on the end vertex in one of two cases:

   The first case is when $b_{f,1} = b_{f,2}$ for every $f \in \mathcal{F}$, although an error has occurred in the k bits of the protocol Π that were sent in the first phase. This happens with probability at most $2^{-r} = k^{-101}$.

The second case is when one of the majorities got flipped. That is, there exists $f \in \mathcal{F}$, such that out of the 101 received copies of $b_{f,1}$ or of $b_{f,2}$, at least 51 were noisy. This happens with probability at most $2 \cdot r \cdot 2^{101} \cdot \epsilon^{51} \leq \epsilon^{20}$.

3. $t_1 \geq k\left(1 - \frac{2\log(k)}{k}\right)$: Assuming the players had the same start vertex, a roll-back can only occur if one of the m_1 bits exchanged by A_1 is noisy. This happens with probability of at most $m_1 \cdot \epsilon \leq \frac{2\log(k)}{k}$. Therefore, with probability $1 - \frac{2\log(k)}{k}$, the depth gain is k.

The protocol A_{i+1}. The protocol A_{i+1} is defined as follows. In the first phase, the players run the protocol A_i, k consecutive times. The start vertices for the first execution of A_i are V_1 and V_2. The start vertices for the $(j+1)^{th}$ execution of A_i are the end vertices of the j^{th} execution. Denote by V_1' and V_2' the vertices in T reached by the players after the first phase.

The second phase is again an error-detecting phase, and it is similar to the second phase of A_1, except that the size of the set $\mathcal{F}$ of random hash functions is now $r_{i+1} = 101^{i+1}\log(k)$ (instead of $r = 101\log(k)$), and that each of the bits $b_{f,1}$ and $b_{f,2}$ is sent 101^{i+1} times (instead of just 101 times).

We calculate the parameters of the protocol A_{i+1}:

1. $m_{i+1} = k \cdot m_i + 2 \cdot 101^{2i+2}\log(k)$: The protocol A_i is run k times, and the error-detecting phase adds $2 \cdot 101^{i+1} \cdot r_{i+1} = 2 \cdot 101^{2i+2} \cdot \log(k)$ rounds.

2. $\alpha_{i+1} \leq k^{-20^{i+1}}$: Assume that the players agreed on the start vertex. They may disagree on the end vertex in one of two cases:

   The first case is when $b_{f,1} = b_{f,2}$ for every $f \in \mathcal{F}$, although the strings $P(V_1, V_1')$ and $P(V_2, V_2')$ do not match. This happens with probability at most $2^{-r_{i+1}} = k^{-101^{i+1}}$.

   The second case is when one of the majorities got flipped. That is, there exists $f \in \mathcal{F}$, such that out of the 101^{i+1} received copies of $b_{f,1}$ or of $b_{f,2}$, more than half were noisy (and in particular, more than $0.5 \cdot 101^{i+1}$). This happens with probability at most $2 \cdot r_{i+1} \cdot 2^{101^{i+1}} \cdot \epsilon^{0.5 \cdot 101^{i+1}} \leq \epsilon^{20^{i+1}}$.

3. $t_{i+1} \geq k \cdot t_i\left(1 - k^{-10^i}\right)$: Assume that the players agreed on the start vertex. A roll-back can only occur in one of two cases:

   The first case is when one of the majorities got flipped. As computed above, this happens with probability at most $k^{-20^{i+1}}$.

   The second case is when in one of the k executions of A_i, the players agree on the start vertex, but disagree on the end vertex. This happens with probability at most $k \cdot \alpha_i$.

   Thus, a roll-back occurs with probability at most k^{-15^i}.

Note also that if the players agree on the start vertex of the j^{th} execution of A_i (an event that occurs when the second case doesn't occur, and in particular, it occurs with probability of at least $1-k^{-15^i}$), the expected depth gain from the j^{th} execution of A_i is at least t_i.

Therefore, the total gain from the k executions of A_i is at least $(1-k^{-15^i})\cdot k\cdot t_i$. A roll back occurs with probability of at most k^{-15^i}, and costs us at most $m_{i+1} < (2k)^{i+1}$ (and note that $k^{-15^i}\cdot(2k)^{i+1} \leq k^{-11^i}$). Thus, the total gain is at least $(1-k^{-15^i})\cdot k\cdot t_i - k^{-11^i} \geq (1-k^{-10^i})\cdot k\cdot t_i$.

We explicitly bound the parameters of A_s. There exists a constant $c\in\mathbb{R}^+$ such that $m_{i+1} \leq k\cdot m_i + c^{i+1}\log(k)$. By induction on i it holds that

$$m_i \leq k^i + k^{i-1}c^1\log(k) + k^{i-2}c^2\log(k) + \ldots + k^0c^i\log(k) \leq k^i + 2c\cdot k^{i-1}\log(k).$$

Thus,

$$m_s \leq k^s\left(1+O\left(\tfrac{\log(k)}{k}\right)\right).$$

In addition,

$$\begin{aligned} t_s &\geq k^s\left(1-\tfrac{2\log(k)}{k}\right)\prod_{i\in\{1,\ldots,s\}}\left(1-k^{-10^i}\right) \\ &\geq k^s\left(1-\tfrac{2\log(k)}{k}-\sum_{i\in\{1,\ldots,s\}}k^{-10^i}\right) \geq k^s\left(1-O\left(\tfrac{\log(k)}{k}\right)\right). \end{aligned}$$

Moreover,

$$\alpha_s \leq k^{-20^{\log\log(n)}} \leq 2^{-\log^4(n)},$$

which is negligible in n.

The protocol A. The simulating protocol A for Π runs the protocol A_s sequentially $a = \frac{n}{t_s}\cdot\left(1+\frac{\log(k)}{k}\right)$ times.

We calculate the parameters of the protocol A:

1. $m = a\cdot m_s = n\cdot\frac{m_s}{t_s}\cdot\left(1+\frac{\log(k)}{k}\right) = n\cdot\left(1+O\left(\frac{\log(k)}{k}\right)\right)$.
2. $\alpha \leq n\cdot\alpha_s \leq 2^{-\log^3(n)}$.
3. $t \geq a\cdot t_s\cdot(1-n\cdot\alpha_s) > n\cdot\left(1+\frac{\log(k)}{2k}\right)$.

Since the bound that we have on t_s applies to every protocol Π and every start vertex v, we get by Azuma's inequality, that the depth of the end vertex reached by A is with high probability close to its expectation, and in particular is at least n. That is, except with negligible probability in n, the protocol A retrieves the transcript of Π completely. Note that

$$\frac{m}{n} = 1 + O\left(\frac{\log(k)}{k}\right) = 1 + O\left(\sqrt{\epsilon \log(\epsilon^{-1})}\right) = 1 + O\left(\sqrt{\mathbf{H}(\epsilon)}\right).$$

□

Remark 2.1. *By setting s to a higher value, we can further decrease the error probability of A.*

Remark 2.2. *If the noise rate of the channel is large, one can first reduce it by repetition and then, when it is small enough, apply our protocol.*

3 Pointer Jumping Games

3.1 Games

Let $k, d \in \mathbb{N}$, and let T be the 2^k-ary tree of depth d, with edges directed from the root to the leaves. Denote the vertex set of T by V, and the edge set of T by E. Denote by $\text{Even}(T) \subseteq V$ the set of non-leaf vertices at an even depth of T, and by $\text{Odd}(T) \subseteq V$ the set of non-leaf vertices at an odd depth of T (where the depth of the root is 0).

The pointer jumping game is a communication game with two parties, called the *first player* and the *second player*. We think of the vertices in $\text{Even}(T)$ as "owned" by the first player, and the vertices in $\text{Odd}(T)$ as "owned" by the second player. Each player gets as an input exactly one edge going out of every node that she owns. We denote by x the input for the first player, and by y the input for the second player. That is, the input x is a set of edges that contains exactly one edge leaving each vertex in an even layer, and the input y is a set of edges that contains exactly one edge leaving each vertex in an odd layer. We denote by $\mathcal{X}(T)$ the set of all possible inputs x for the first player, and by $\mathcal{Y}(T)$ the set of all possible inputs y for the second player.

For a pair of inputs $x \in \mathcal{X}(T)$ and $y \in \mathcal{Y}(T)$, we denote by $L(x, y)$ the unique leaf of T reachable from the root using the edges of $x \cup y$. The players' mutual goal is to find the leaf $L(x, y)$.

For a random variable Z, we denote by P_Z the distribution of Z.

Definition 3.1 (**Pointer Jumping Game**). *Let $k, d \in \mathbb{N}$, and let T be the 2^k-ary tree of depth d. Let $P_X : \mathcal{X}(T) \to [0, 1]$ and $P_Y : \mathcal{Y}(T) \to [0, 1]$ be a pair of distributions.*

The pointer jumping game *G with parameters (k, d, P_X, P_Y) is the following two players communication game: A set $X \in \mathcal{X}(T)$ is drawn according to P_X, and is given as input to the first player. A set $Y \in \mathcal{Y}(T)$ is (independently) drawn according to P_Y, and is given as input to the second player. It is assumed that both players know k, d, P_X, P_Y. The players' mutual goal is to both output the leaf $L(X, Y)$.*

We will sometimes write the parameters of the game G as (k, d, X, Y) instead of (k, d, P_X, P_Y).

3.2 Protocols

We will consider the communication complexity of pointer jumping games (or simply "games"), in the case where the players communicate through an *ϵ-noisy channel*, and where they are allowed to err with probability δ, for some $\epsilon, \delta \in [0, 1]$. An ϵ-noisy channel is a channel that flips each communicated bit (independently) with probability ϵ.

Definition 3.2 (**Protocol**)**.** *Let G be a game with parameters (k, d, P_X, P_Y), and let $\epsilon, \delta \in [0, 1]$. A* protocol Π *for G with* noise rate ϵ *and* error δ *is a pair of probabilistic strategies, one for each player (if the strategies are deterministic, we will say that the protocol is deterministic).*

The protocol proceeds in rounds. In each round (exactly) one of the players sends a bit to the other player through an ϵ-noisy channel. At the end of the protocol both players output the correct vertex $L(X, Y)$ with probability at least $1 - \delta$. The probability here is taken over the selection of inputs, the randomness of the players' strategies, and the channel's noise.

Predetermined Turns. When considering deterministic protocols (or probabilistic protocols with fixed random strings), we will assume that it is known in advance which player sends a bit in each round. That is, for a given protocol, the order of communication is predetermined, and does not depend on the inputs and on the transcript of the communication (that is, the bits sent so far).

This is justified as follows: Note that in each round both players must know who speaks next, because we require that in each round (with probability 1) *exactly* one of the players sends a bit. Moreover, since the channel is noisy, every transcript can be changed by the channel to any other transcript. Therefore, for fixed inputs x and y, the identity of the player who speaks next cannot depend on the transcript. Since we consider a product distribution over the inputs, the order of communication must be the same for every pair x, y.

Balanced Protocols. In our main lower bound proof, it will be convenient to assume that every deterministic protocol satisfies the following property: At every stage of the protocol,

if the protocol ends within the next $2k$ rounds with probability greater than 0, then it ends within the next $2k$ rounds with probability 1, where the probability is over the selection of inputs and the channel's noise. Protocols that satisfy the above property are called *balanced*.

We remark that every protocol can be converted into a balanced protocol by adding $2k$ dummy rounds at the end of the protocol. Therefore, any lower bound proven for balanced protocols holds for general protocols, up to an additive $2k$ term. Hence, it suffices to only consider balanced protocols when proving our lower bound.

In all that comes below, when considering deterministic protocols, we will refer to a protocol that is not necessarily balanced as a *general protocol*, and refer to a balanced protocol simply as a *protocol*.

Bounded Number of Rounds. For simplicity, we will only consider protocols with some finite bound on the number of rounds. The bound can be arbitrarily large (say, double exponential in kd) so its affect on the probability of error is negligible. The reason that this simplifies the presentation is that this way the number of deterministic protocols is finite, so the deterministic communication complexity of a game can be defined as the minimum over these protocols, rather than the infimum.

3.3 Communication Complexity

Definition 3.3 (**Communication Complexity**). *Let G be a game with parameters (k, d, P_X, P_Y). Let $\epsilon, \delta \in [0, 1]$. Denote by $\mathcal{P}^*_{\epsilon,\delta}$ the set of all probabilistic protocols for G with noise rate ϵ and error δ, and by $\mathcal{P}_{\epsilon,\delta}$ the set of all (balanced) deterministic protocols for G with noise rate ϵ and error δ.*

Let $\Pi \in \mathcal{P}^*_{\epsilon,\delta}$ be a protocol. The (expected) communication complexity of the protocol Π*, denoted $\mathsf{CC}(\Pi)$, is the expected number of bits communicated by the players during the execution of the protocol. The expectation here is taken over the selection of inputs, the randomness of the players' strategies, and the channel's noise.*

The (expected) probabilistic communication complexity of the game G*, denoted $\mathsf{CC}^*_{\epsilon,\delta}(G)$, is given by*

$$\mathsf{CC}^*_{\epsilon,\delta}(G) = \inf_{\Pi \in \mathcal{P}^*_{\epsilon,\delta}} \{\mathsf{CC}(\Pi)\}.$$

The (expected) deterministic communication complexity of the game G*, denoted $\mathsf{CC}_{\epsilon,\delta}(G)$, is given by*

$$\mathsf{CC}_{\epsilon,\delta}(G) = \min_{\Pi \in \mathcal{P}_{\epsilon,\delta}} \{\mathsf{CC}(\Pi)\}.$$

3.4 Our Lower Bound Result

Theorem 2 (**Main, Lower Bound**). *Let G be a pointer jumping game with parameters $(k, d, U_{\mathcal{X}}, U_{\mathcal{Y}})$, where $U_{\mathcal{X}}$ and $U_{\mathcal{Y}}$ are the uniform distributions over the sets of possible inputs for the first and second players (respectively). Let $\epsilon = \frac{2000\log(k)}{k^2}$ and $\delta \in [0,1]$. Then,*

$$\mathsf{CC}^*_{\epsilon,\delta}(G) \geq d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - 102k.$$

4 Definitions and Preliminaries for the Communication Complexity Lower Bound

4.1 General Notation

Throughout the paper, unless stated otherwise, sets denoted by Ω will always be finite. All logarithms are taken with base 2, and we define $0\log(0) = 0$. We use the fact that the function $-x\log(x)$ is monotone increasing for $0 \leq x \leq \frac{1}{e}$.

4.2 Random Variables and their Distributions

We will use capital letters to denote random variables, and we will use lower case letters to denote values. For example, X, Y will denote random variables, and x, y will denote values that these random variables can take.

For a random variable X, we denote by P_X the distribution of X. For an event U we use the notation $P_{X|U}$ to denote the distribution of $X|U$, that is, the distribution of X conditioned on the event U. If Z is an additional random variable that is fixed (e.g., inside an expression where an expectation over Z is taken), we denote by $P_{X|Z}$ the distribution of X conditioned on Z. In the same way, for two (or more) random variables X, Y, we denote their joint distribution by P_{XY}, and we use the same notation as above to denote conditional distributions. For example, for an event U, we write $P_{XY|U}$ to denote the distribution of X, Y conditioned on the event U, i.e., $P_{XY|U}(x, y) = \Pr(X = x, Y = y|U)$.

If A and B are events, and B occurs with probability 0, we set $\Pr[A|B] = 0$. In general, we will many times condition on an event U that may occur with probability 0. This may cause conditional probabilities and distributions, such as $P_{X|U}$, to be undefined. Nevertheless, the undefined values will usually be multiplied by 0 to give 0. A statement that uses distributions or values that may be undefined should be interpreted as correct in the case that all the involved values and distributions are well defined (or are multiplied by 0). For example, we

may argue about a distribution $P_{X|Z=z}$, without necessarily mentioning that we assume that $z \in \text{supp}(Z)$.

4.3 Information Theory

4.3.1 Information

Definition 4.1 (**Information**). *Let $\mu : \Omega \to [0,1]$ be a distribution. The* information *of μ, denoted $\mathbf{I}(\mu)$, is defined by*

$$\mathbf{I}(\mu) = \sum_{x \in \text{supp}(\mu)} \mu(x) \log\left(\frac{\mu(x)}{\frac{1}{|\Omega|}}\right) = \sum_{x \in \text{supp}(\mu)} \mu(x) \log\left(|\Omega|\mu(x)\right).$$

Equivalently,

$$\mathbf{I}(\mu) = \log(|\Omega|) - \mathbf{H}(\mu),$$

where $\mathbf{H}(\mu)$ denotes the Shannon entropy of μ.

For a random variable X taking values in Ω, with distribution $P_X : \Omega \to [0,1]$, we define $\mathbf{I}(X) = \mathbf{I}(P_X)$.

Proposition 4.2 (**Supper-Additivity of Information**). *Let $X_1, \ldots, X_m$ be m random variables, taking values in $\Omega_1, \ldots, \Omega_m$, respectively. Consider the random variable $(X_1, \ldots, X_m)$, taking values in $\Omega_1 \times \ldots \times \Omega_m$. Then,*

$$\mathbf{I}\left((X_1, \ldots, X_m)\right) \geq \sum_{i \in [m]} \mathbf{I}(X_i).$$

Proof. Using the sub-additivity of the Shannon entropy function, we have

$$\begin{aligned} \mathbf{I}\left((X_1, \ldots, X_m)\right) &= \log(|\Omega_1 \times \ldots \times \Omega_m|) - \mathbf{H}(X_1, \ldots, X_m) \geq \sum_{i \in [m]} \log(|\Omega_i|) - \sum_{i \in [m]} \mathbf{H}(X_i) \\ &= \sum_{i \in [m]} \left(\log(|\Omega_i|) - \mathbf{H}(X_i)\right) = \sum_{i \in [m]} \mathbf{I}(X_i). \end{aligned}$$

□

4.3.2 Min-Entropy

Definition 4.3 (**Min-Entropy**). *Let $\mu : \Omega \to [0,1]$ be a distribution. The* min-entropy *of μ, denoted $\mathbf{H}_\infty(\mu)$, is defined by*

$$\mathbf{H}_\infty(\mu) = \min_{x \in \mathrm{supp}(\mu)} \{-\log(\mu(x))\}.$$

For a random variable X with distribution P_X, we define $\mathbf{H}_\infty(X) = \mathbf{H}_\infty(P_X)$.

4.3.3 Relative Entropy

Definition 4.4 (**Relative Entropy**). *Let $\mu_1, \mu_2 : \Omega \to [0,1]$ be two distributions, where Ω is discrete (but not necessarily finite). The* relative entropy *between μ_1 and μ_2, denoted $\mathbf{D}(\mu_1 \| \mu_2)$, is defined as*

$$\mathbf{D}(\mu_1 \| \mu_2) = \sum_{x \in \Omega} \mu_1(x) \log\left(\tfrac{\mu_1(x)}{\mu_2(x)}\right).$$

Proposition 4.5. *Let $\mu_1, \mu_2 : \Omega \to [0,1]$ be two distributions. Then,*

$$\mathbf{D}(\mu_1 \| \mu_2) \geq 0.$$

The following relation is called *Pinsker's Inequality*, and it relates the relative entropy to the L^1 distance.

Proposition 4.6 (**Pinsker's Inequality**). *Let $\mu_1, \mu_2 : \Omega \to [0,1]$ be two distributions. Then,*

$$2\ln(2) \cdot \mathbf{D}(\mu_1 \| \mu_2) \geq \|\mu_1 - \mu_2\|^2,$$

where

$$\|\mu_1 - \mu_2\| = \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)| = 2 \max_{E \subseteq \Omega} \{\mu_1(E) - \mu_2(E)\}.$$

5 Operations on Distributions

5.1 Flattening a Distribution

Definition 5.1 (**Flat Distribution**). *Let $\mu : \Omega \to [0,1]$ be a distribution. Let $a \geq 0$. We say that μ is a-*flat *if*

$$\forall x_1, x_2 \in \mathrm{supp}(\mu) : \tfrac{\mu(x_1)}{\mu(x_2)} \leq 2^a.$$

Proposition 5.2. *Let $\mu : \Omega \to [0,1]$ be an a-flat distribution. Then,*

$$\mathbf{H}(\mu) - a \leq \mathbf{H}_\infty(\mu) \leq \mathbf{H}(\mu).$$

Given a (possibly not flat) distribution $\mu : \Omega \to [0,1]$, we would like to turn it into a convex combination of flat distributions, plus a residue distribution (which, in our applications, will have a small coefficient in the convex combination). We will do that by conditioning the distribution on the following *"flattening value"* $f_{\mu,a,r}(x)$, where $x \in \text{supp}(\mu)$, and $a > 0$, $r \in \mathbb{N}$. First consider the function $f'_{\mu,a} : \text{supp}(\mu) \to \mathbb{Z}$ given by

$$f'_{\mu,a}(x) = \left\lceil \frac{\log(|\Omega| \cdot \mu(x))}{a} \right\rceil.$$

Note that $f'_{\mu,a}(x)$ gives the rounded value of the logarithm of the ratio between $\mu(x)$ and the uniform distribution over Ω. The function $f_{\mu,a,r} : \text{supp}(\mu) \to \{-r, \ldots, r+1\}$ is given by

$$f_{\mu,a,r}(x) = \begin{cases} -r & \text{if } f'_{\mu,a}(x) \leq -r \\ f'_{\mu,a}(x) & \text{if } -r < f'_{\mu,a}(x) \leq r \\ r+1 & \text{if } f'_{\mu,a}(x) > r \end{cases}$$

For every $i \in \{-r, \ldots, r+1\}$, we define the set $S_i \subseteq \Omega$ to be the set of all elements $x \in \Omega$ such that $f_{\mu,a,r}(x) = i$. That is, for $i = -r$, the set S_i is the set of elements x such that

$$\mu(x) \leq 2^{-ar} \cdot \tfrac{1}{|\Omega|}.$$

For $i \in \{-r+1, \ldots, r\}$, the set S_i is the set of elements x such that

$$2^{a(i-1)} \cdot \tfrac{1}{|\Omega|} < \mu(x) \leq 2^{ai} \cdot \tfrac{1}{|\Omega|}.$$

For $i = r+1$, the set S_i is the set of elements x such that

$$2^{ar} \cdot \tfrac{1}{|\Omega|} < \mu(x).$$

For every $i \in \{-r, \ldots, r+1\}$, define $\mu_i = \mu|_{S_i} : \Omega \to [0,1]$ to be the conditional distribution μ conditioned on S_i, and define $\alpha_i = \mu(S_i)$. (If $\alpha_i = 0$ we define μ_i to be the uniform distribution over Ω). Then we have

$$\mu = \sum_{i \in \{-r, \ldots, r+1\}} \alpha_i \mu_i.$$

Lemma 5.3. *For every $i \in \{-r+1, \ldots, r\}$, the distribution μ_i defined above is a-flat.*

Proof. Assume $\alpha_i > 0$. Let $x_1, x_2 \in \text{supp}(\mu_i) = S_i$. It holds that

$$\frac{\mu_i(x_1)}{\mu_i(x_2)} = \frac{\left(\frac{\mu(x_1)}{\alpha_i}\right)}{\left(\frac{\mu(x_2)}{\alpha_i}\right)} = \frac{\mu(x_1)}{\mu(x_2)} \leq \frac{2^{ai} \cdot \frac{1}{|\Omega|}}{2^{a(i-1)} \cdot \frac{1}{|\Omega|}} = 2^a.$$

□

Lemma 5.4. *For $i = r+1$, if $\alpha_{r+1} > 0$, the distribution μ_{r+1} defined above satisfies $\mathbf{I}(\mu_{r+1}) > ar$.*

Proof.

$$\mathbf{I}(\mu_{r+1}) = \sum_{x \in \text{supp}(\mu_{r+1})} \mu_{r+1}(x) \log\left(|\Omega| \mu_{r+1}(x)\right) > \sum_{x \in \text{supp}(\mu_{r+1})} \left(\mu_{r+1}(x) \cdot ar\right) = ar.$$

□

Lemma 5.5. *For $i = -r$, the coefficient α_{-r} defined above satisfies $\alpha_{-r} \leq 2^{-ar}$.*

Proof. Assume $\alpha_{-r} > 0$.

$$\alpha_{-r} = \sum_{\{x \in \Omega \,:\, \mu(x) \leq 2^{-ar}/|\Omega|\}} \mu(x) \leq 2^{-ar}.$$

□

Lemma 5.6. *Let $\mu : \Omega \to [0,1]$ be a distribution. Consider the random variable $f_{\mu,a,r}(X)$, where X is randomly selected according to μ, and $a, r \geq 1$. Let $I = \max\{\mathbf{I}(\mu), 2\}$. Then,*

1. $\mathbf{H}(f_{\mu,a,r}(X)) \leq \log(I) + 11$.
2. $\forall m \geq 1 : \Pr_X\left[f_{\mu,a,r}(X) > mI\right] < \frac{10}{a \cdot m}$.

Proof. We first show Part (2). Observe that if $mI \geq r+1$ then Part (2) is trivial, as it is always the case that $f_{\mu,a,r}(X) \leq r+1$. Thus, we assume that $mI < r+1$. Assume for a contradiction that

$$\delta := \Pr_X\left[f_{\mu,a,r}(X) > mI\right] \geq \frac{10}{a \cdot m}.$$

Let $p = 2^{a(mI-1)}$. Let $S \subseteq \Omega$ be the set of elements x with $\mu(x) > \frac{p}{|\Omega|}$. Observe that

$$\delta = \Pr_X\left[f_{\mu,a,r}(X) > mI\right] \leq \Pr_X\left[\mu(X) > \tfrac{p}{|\Omega|}\right] = \mu(S).$$

Let $\delta' := \mu(S)$. Consider the distribution μ' that assigns each of the elements in S a probability of $\frac{\delta'}{|S|} \geq \frac{p}{|\Omega|}$, and each of the other elements a probability of $\frac{1-\delta'}{|\Omega|-|S|} \geq \frac{1-\delta'}{|\Omega|}$. That is, μ' is obtained by re-distributing the weight of μ, so it will be uniform on S and on $\Omega \setminus S$. Observe that $\mathbf{H}(\mu') \geq \mathbf{H}(\mu)$, and hence $\mathbf{I}(\mu') \leq \mathbf{I}(\mu)$. In addition, $\forall x \in (0,1] : x \log(x) > -1$. This leads to a contradiction as follows

$$\begin{aligned} I \geq \mathbf{I}(\mu) \geq \mathbf{I}(\mu') &\geq \delta' \log(p) + (1-\delta') \log(1-\delta') \geq \tfrac{10}{a \cdot m} \cdot a(mI-1) - 1 \\ &= 10I - \tfrac{10}{m} - 1 \geq 10I - 11 \geq 2I. \end{aligned}$$

We turn to prove Part (1). Consider the random variable $Z = Z(X)$ that gets the value $m \in \mathbb{Z}$ if

$$100mI < f_{\mu,a,r}(X) \leq 100(m+1)I.$$

Observe that

$$\mathbf{H}(f_{\mu,a,r}(X)) = \mathbf{H}(Z) + \mathbf{H}(f_{\mu,a,r}(X) \mid Z) \leq \mathbf{H}(Z) + \log(100I),$$

where the last inequality holds since for a fixed $Z = m$, the random variable $f_{\mu,a,r}(X)$ takes at most $100I$ possible values. The rest of the proof is devoted to bounding $\mathbf{H}(Z) \leq 4$.

Let $P : \mathbb{Z} \to [0,1]$ be the distribution of the random variable Z. By Part (2) we know that

$$\forall m \geq 1 : \Pr_X [f_{\mu,a,r}(X) > 100mI] < \tfrac{10}{a \cdot 100m} = \tfrac{1}{10a \cdot m}.$$

Thus,

$$\forall m \geq 1 : \sum_{i \geq m} P(i) \leq \tfrac{1}{10a \cdot m}.$$

Let $F : \mathbb{N} \to [0,1]$ be a function (not necessarily a distribution). Define the entropy of the function F by

$$\mathbf{H}^*(F) = -\sum_{i \in \mathbb{N}} F(i) \log(F(i)).$$

Let $\mathcal{F}$ be the set of functions $F : \mathbb{N} \to [0,1]$ satisfying the condition

$$\forall m \in \mathbb{N} : \sum_{i \geq m} F(i) \leq \tfrac{1}{10a \cdot m}. \tag{1}$$

Consider the function $Q : \mathbb{N} \to [0,1]$ defined by

$$Q(i) = \tfrac{1}{10ai} - \tfrac{1}{10a(i+1)}.$$

Observe that $Q \in \mathcal{F}$. The following claim shows that among the functions in $\mathcal{F}$, the function Q maximizes the entropy $\mathbf{H}^*$.

Claim 5.7. $\mathbf{H}^*(Q) = \max_{F \in \mathcal{F}}\{\mathbf{H}^*(F)\}$.

Proof. Let $F \in \mathcal{F}$ be a function. We will show that $\mathbf{H}^*(F) \leq \mathbf{H}^*(Q)$. Observe that $S := \sum_{i \in \mathbb{N}} Q(i) = \frac{1}{10a} \leq \frac{1}{10}$. We assume without loss of generality that $\sum_{i \in \mathbb{N}} F(i) = S$, as otherwise we can enlarge $F(1)$ and get a function that still satisfies the condition in Equation 1, and has a larger entropy. Next, we observe that $\frac{1}{S}F$ and $\frac{1}{S}Q$ are distributions, and hence using Proposition 4.5,

$$0 \leq \mathbf{D}\left(\tfrac{1}{S}F \| \tfrac{1}{S}Q\right) = \sum_{i \in \mathbb{N}} \tfrac{1}{S}F(i) \log\left(\frac{\frac{1}{S}F(i)}{\frac{1}{S}Q(i)}\right) = \tfrac{1}{S}\sum_{i \in \mathbb{N}} F(i) \log\left(\tfrac{F(i)}{Q(i)}\right).$$

This implies

$$\mathbf{H}^*(F) = -\sum_{i \in \mathbb{N}} F(i)\log(F(i)) \leq -\sum_{i \in \mathbb{N}} F(i)\log(Q(i)).$$

Observe that in order to get the desired result $\mathbf{H}^*(F) \leq \mathbf{H}^*(Q)$, it suffices to prove the following inequity:

$$-\sum_{i \in \mathbb{N}} F(i)\log(Q(i)) \leq -\sum_{i \in \mathbb{N}} Q(i)\log(Q(i)) = \mathbf{H}^*(Q).$$

The rest of the proof is devoted to showing the last inequality.

We first note that the set $\mathcal{F}$, equipped with the L_1 metric, is a compact metric space. We assume without loss of generality that the function F maximizes the expression $-\sum_{i \in \mathbb{N}} F(i)\log(Q(i))$ among the functions in $\mathcal{F}$. We will show that F satisfies all the constraints of Equation 1 with equality, that is, $\forall m \in \mathbb{N} : \sum_{i \geq m} F(i) = \frac{1}{10a \cdot m}$. Thus, $F = Q$.

The reason that F satisfies all the constraints of Equation 1 with equality is the following. First note that $\sum_{i \geq 1} F(i) = \frac{1}{10a}$, as otherwise we can enlarge $F(1)$ and get a function that still satisfies the condition in Equation 1, while increasing the expression $-\sum_{i \in \mathbb{N}} F(i)\log(Q(i))$, in contradiction to the maximality of F. Assume for a contradiction that there exists $m \in \mathbb{N}$ such that $\sum_{i \geq m+1} F(i) < \frac{1}{10a \cdot (m+1)}$. Then, since Q is a monotone decreasing function, we are better off "moving" some of the weight of F from m to $m+1$. Formally, consider the function $F' : \mathbb{N} \to [0,1]$, defined by $\forall i \in \mathbb{N} \setminus \{m, m+1\} : F'(i) = F(i)$, and $F'(m) = F(m) - \epsilon$, and $F'(m+1) = F(m+1) + \epsilon$, where $\epsilon > 0$ is sufficiently small so that $\sum_{i \geq m+1} F'(i) \leq \frac{1}{10a \cdot (m+1)}$. Note that F' still satisfies the condition in Equation 1. We then derive a contradiction to

the maximality of F as

$$-\sum_{i\in\mathbb{N}} F(i)\log(Q(i)) < -\sum_{i\in\mathbb{N}} F'(i)\log(Q(i)).$$

□

We can now bound $\mathbf{H}(P)$. First observe that $Q(i) = \frac{1}{10ai} - \frac{1}{10a(i+1)} = \frac{1}{10a\cdot i(i+1)} < \frac{1}{10i^2}$. Therefore,

$$\mathbf{H}^*(Q) < \sum_{i\in\mathbb{N}} \tfrac{1}{10i^2}\cdot\log\left(10i^2\right) \leq 1.$$

In addition, for $i \in \{2,3,\ldots\}$ it holds that

$$P(-i) \leq \Pr_X\left[f_{\mu,a,r}(X) \leq -100(i-1)I\right] \leq \Pr_X\left[\mu(X) \leq 2^{-100(i-1)aI}\cdot\tfrac{1}{|\Omega|}\right] \leq 2^{-100(i-1)I}.$$

Thus,

$$\begin{aligned}\mathbf{H}(P) &\leq \mathbf{H}^*(Q) - P(0)\log(P(0)) - P(-1)\log(P(-1)) - \sum_{i\in\{2,3,\ldots\}} P(-i)\log(P(-i)) \\ &\leq 1 + 2 + \sum_{i\in\{2,3,\ldots\}} 2^{-100(i-1)I}(100(i-1)I) \leq 4.\end{aligned}$$

□

5.2 Partitioning a Distribution

Lemma 5.8. *Let $\mu : \Omega \to [0,1]$ be a distribution satisfying $\mathbf{H}_\infty(\mu) \geq \log(|\Omega|) - a$, where $a \in \mathbb{R}^+$. Let $c \in \mathbb{N}$. Let $\mu_1,\ldots,\mu_{2^c} : \Omega \to [0,1]$ be 2^c distributions. Let $\alpha_1,\ldots,\alpha_{2^c} \in [0,1]$. Assume that*

$$\mu = \sum_{i\in[2^c]} \alpha_i\mu_i.$$

For $i \in [2^c]$, let $I_i = \mathbf{I}(\mu_i)$. Let $h \in \mathbb{R}^+$. Let $\mathcal{B} \subseteq [2^c]$ be the set of indices i such that $I_i > a + c + h$. Then,

$$\sum_{i\in\mathcal{B}} \alpha_i < 2^{-h}.$$

Proof. Recall that we assume

$$\mathbf{H}_\infty(\mu) = \min_{x\in\mathrm{supp}(\mu)}\{-\log(\mu(x))\} \geq \log(|\Omega|) - a.$$

Therefore, for every $x \in \mathrm{supp}(\mu)$, it holds that

$$a \geq \log(|\Omega|) + \log(\mu(x)) = \log(|\Omega|\mu(x)). \tag{2}$$

Let $i \in [2^c]$ be such that $\alpha_i \geq 2^{-(h+c)}$. Let $x \in \Omega$. It holds that

$$\mu(x) \geq \alpha_i \mu_i(x) \geq 2^{-(h+c)} \mu_i(x). \tag{3}$$

Using Equations 3 and 2, we have

$$\begin{aligned} I_i &= \sum_{x \in \mathrm{supp}(\mu_i)} \mu_i(x) \log\left(|\Omega|\mu_i(x)\right) \\ &\leq \sum_{x \in \mathrm{supp}(\mu_i)} \mu_i(x) \log\left(|\Omega| \cdot 2^{h+c}\mu(x)\right) \\ &\leq \left(\sum_{x \in \mathrm{supp}(\mu_i)} \mu_i(x) \log(|\Omega|\mu(x)) \right) + h + c \\ &\leq a + h + c. \end{aligned}$$

Therefore, $i \in \mathcal{B}$ implies $\alpha_i < 2^{-(h+c)}$. As there are 2^c possible indices i, it holds that $|\mathcal{B}| \leq 2^c$. Thus,

$$\sum_{i \in \mathcal{B}} \alpha_i < |\mathcal{B}| \cdot 2^{-(h+c)} \leq 2^{-h}.$$

□

Lemma 5.9. *Let $\mu : \Omega \to [0,1]$ be a distribution. Let $c \in \mathbb{N}$. Let $\mu_1, \ldots, \mu_{2^c} : \Omega \to [0,1]$ be 2^c distributions. Let $\alpha_1, \ldots, \alpha_{2^c} \in [0,1]$. Assume that*

$$\mu = \sum_{i \in [2^c]} \alpha_i \mu_i.$$

Let $h \in \mathbb{R}^+$. Let $\mathcal{B} \subseteq [2^c]$ be the set of indices i such that $\mathbf{H}_\infty(\mu_i) < \mathbf{H}_\infty(\mu) - h - c$. Then,

$$\sum_{i \in \mathcal{B}} \alpha_i < 2^{-h}.$$

Therefore, also

$$\sum_{i \in [2^c]} \alpha_i \mathbf{H}_\infty(\mu_i) \geq \mathbf{H}_\infty(\mu) - c - 4.$$

Proof. Let $i \in [2^c]$ be such that $\alpha_i \geq 2^{-(h+c)}$. Let $x \in \Omega$. It holds that

$$\mu(x) \geq \alpha_i \mu_i(x) \geq 2^{-(h+c)} \mu_i(x).$$

Hence,

$$\mathbf{H}_\infty(\mu_i) = \min_{x \in \mathrm{supp}(\mu_i)} \{-\log(\mu_i(x))\} \geq \min_{x \in \mathrm{supp}(\mu)} \left\{-\log\left(2^{h+c}\mu(x)\right)\right\} = \mathbf{H}_\infty(\mu) - h - c.$$

Therefore, $i \in \mathcal{B}$ implies $\alpha_i < 2^{-(h+c)}$. As there are 2^c possible indices i, it holds that $|\mathcal{B}| \leq 2^c$. Thus,

$$\sum_{i \in \mathcal{B}} \alpha_i < |\mathcal{B}| \cdot 2^{-(h+c)} \leq 2^{-h}.$$

The last bound also implies that

$$\left(\sum_{i \in [2^c]} \alpha_i \mathbf{H}_\infty(\mu_i)\right) - \mathbf{H}_\infty(\mu) + c = \sum_{i \in [2^c]} \alpha_i \left(\mathbf{H}_\infty(\mu_i) - \mathbf{H}_\infty(\mu) + c\right) \geq \sum_{h \in \mathbb{N}} 2^{-(h-1)}(-h) = -4$$

(where the last inequality follows by partitioning the sum according to $-h = \lfloor \mathbf{H}_\infty(\mu_i) - \mathbf{H}_\infty(\mu) + c \rfloor$ and using the previous bound). □

Lemma 5.10. *Let $\mu : \Omega \to [0,1]$ be an a-*flat *distribution, where $a \geq 1$. Let $c \in \mathbb{N}$. Let $\mu_1, \ldots, \mu_{2^c} : \Omega \to [0,1]$ be 2^c distributions. Let $\alpha_1, \ldots, \alpha_{2^c} \in [0,1]$. Assume that*

$$\mu = \sum_{i \in [2^c]} \alpha_i \mu_i.$$

For $i \in [2^c]$, let $I_i = \mathbf{I}(\mu_i)$, and let $I = \mathbf{I}(\mu)$. Let $h \in \mathbb{R}^+$. Let $\mathcal{B} \subseteq [2^c]$ be the set of indices i such that $I_i > I + c + h + a$. Then,

$$\sum_{i \in \mathcal{B}} \alpha_i < 2^{-h}.$$

Proof. Let $S = \mathrm{supp}(\mu) \subseteq \Omega$. Consider the distribution $\mu_S : S \to [0,1]$ obtained by restricting the domain of μ to S. That is, μ_S is given by $\forall x \in S : \mu_S(x) = \mu(x)$.

We first claim that μ_S satisfies $\mathbf{H}_\infty(\mu_S) \geq \log(|S|) - a$. Let $p_{max} = \max_{x \in S}\{\mu_S(x)\}$ be the maximal probability of an element according to μ_S. Let $x \in S$. Using the flatness of μ,

$$\mu_S(x) = \mu(x) \geq 2^{-a} p_{max}.$$

Therefore $\mu_S(S) = 1 \geq 2^{-a}|S|p_{max}$, which means $p_{max} \leq \frac{2^a}{|S|}$. We can now bound the

min-entropy of μ_S:

$$\mathbf{H}_\infty(\mu_S) = -\log(p_{max}) \geq -\log\left(\tfrac{2^a}{|S|}\right) = \log(|S|) - a. \tag{4}$$

For every $i \in [2^c]$, let $\mu_{S,i} : S \to [0,1]$ be the distribution obtained by restricting the domain of μ_i to S. That is, $\mu_{S,i}$ is given by $\forall x \in S : \mu_{S,i}(x) = \mu_i(x)$. Observe that $\mathrm{supp}(\mu_i) \subseteq S$, and hence $\mu_{S,i}$ is in fact a distribution (unless $\alpha_i = 0$). The distribution μ_S can be decomposed as follows:

$$\mu_S = \sum_{i \in [2^c]} \alpha_i \mu_{S,i}.$$

We bound I_i using $\mathbf{I}(\mu_{S,i})$ as follows (for i such that $\alpha_i > 0$)

$$\begin{aligned} I_i &= \log(|\Omega|) - \mathbf{H}(\mu_i) = (\log(|\Omega|) - \log(|S|)) + (\log(|S|) - \mathbf{H}(\mu_{S,i})) \\ &\leq (\log(|\Omega|) - \mathbf{H}(\mu)) + \mathbf{I}(\mu_{S,i}) = I + \mathbf{I}(\mu_{S,i}). \end{aligned}$$

Recall that $\mathcal{B}$ is the set of indices i such that $I_i > I+c+h+a$, which implies $\mathbf{I}(\mu_{S,i}) > c+h+a$. By applying Lemma 5.8 to μ_S we get the desired

$$\sum_{i \in \mathcal{B}} \alpha_i < 2^{-h}.$$

□

Lemma 5.11. *Let $\mu : \Omega \to [0,1]$ be a distribution satisfying $I = \mathbf{I}(\mu) \leq 0.01$. Let $\mathcal{A} \subseteq \Omega$ be the set of elements with $\mu(x) < \frac{1}{|\Omega|}$. Denote*

$$I^{neg}(\mu) = -\sum_{x \in \mathcal{A}} \mu(x) \log(|\Omega|\mu(x)).$$

Then, $I^{neg}(\mu) < 4I^{0.25} \log\left(\frac{1}{I^{0.25}}\right) < 4I^{0.1}$.

Proof. Let $\beta \in [0,1]$ be such that $|\mathcal{A}| = \beta|\Omega|$, that is, β is the density of $\mathcal{A}$. Let $\alpha = \mu(\mathcal{A})$, that is, α is the weight of $\mathcal{A}$. Consider the distribution $\mu' : \Omega \to [0,1]$ that gives each element in $\mathcal{A}$ the probability $\frac{\alpha}{|\Omega|\beta}$, and each of the other elements, the probability $\frac{1-\alpha}{|\Omega|(1-\beta)}$. In particular, μ' is uniform over $\mathcal{A}$, and is uniform over $\Omega \setminus \mathcal{A}$. Observe that the set of elements with $\mu'(x) < \frac{1}{|\Omega|}$ is exactly $\mathcal{A}$. By the concavity of the logarithm function, it holds that $I^{neg}(\mu') \geq I^{neg}(\mu)$, and note also that $\mathbf{I}(\mu') \leq \mathbf{I}(\mu) = I$. Therefore, it suffices to show $I^{neg}(\mu') < 4I^{0.25} \log\left(\frac{1}{I^{0.25}}\right)$.

Observe that

$$\mathbf{H}(\mu') = -\alpha \log\left(\tfrac{\alpha}{|\Omega|\beta}\right) - (1-\alpha)\log\left(\tfrac{(1-\alpha)}{|\Omega|(1-\beta)}\right) = -\mathbf{D}\left((\alpha, 1-\alpha)\|(\beta, 1-\beta)\right) + \log(|\Omega|).$$

Therefore,

$$\mathbf{I}(\mu') = \log(|\Omega|) - \mathbf{H}(\mu') = \mathbf{D}((\alpha, 1-\alpha)\|(\beta, 1-\beta)).$$

Using Pinsker's Inequality (Proposition 4.6) it holds that

$$|\alpha - \beta| < \sqrt{\mathbf{I}(\mu')} \leq \sqrt{I}. \tag{5}$$

We turn to bound $I^{neg}(\mu')$. Observe that

$$I^{neg}(\mu') = -\alpha \log\left(\tfrac{\alpha}{\beta}\right).$$

Recall that for $x \in \mathcal{A}$ it holds that $\mu(x) < \frac{1}{|\Omega|}$. Therefore $\alpha = \mu(\mathcal{A}) < |\mathcal{A}| \cdot \frac{1}{|\Omega|} = \beta$. We consider the following two cases:

Case 1: $\beta > I^{0.25}$**.** By Equation 5 it holds that $\alpha > \beta - \sqrt{I}$. Therefore, $\frac{\alpha}{\beta} > \frac{\beta - \sqrt{I}}{\beta} = 1 - \frac{\sqrt{I}}{\beta} > 1 - I^{0.25}$. Since $\alpha \leq 1$, it holds that

$$I^{neg}(\mu') \leq -\log\left(1 - I^{0.25}\right) \leq 2I^{0.25}.$$

Case 2: $\beta \leq I^{0.25}$**.** Since $\alpha < \beta$, it is also the case that $\alpha < I^{0.25}$. Since $\beta \leq 1$, it holds that

$$I^{neg}(\mu') \leq -\alpha \log(\alpha) < -I^{0.25} \log\left(I^{0.25}\right).$$

□

Lemma 5.12. *Let $\mu : \Omega \to [0,1]$ be a distribution satisfying $I = \mathbf{I}(\mu) \leq 0.01$. Let $\mathcal{A} \subseteq \Omega$ be the set of elements with $\mu(x) \geq \frac{2}{|\Omega|}$. Then, $\mu(\mathcal{A}) < 4I^{0.25}\log\left(\frac{1}{I^{0.25}}\right) + I < 5I^{0.1}$.*

Proof. Denote

$$A = \sum_{x \in \mathcal{A}} \mu(x) \log(|\Omega|\mu(x)).$$

Observe that for every $x \in \mathcal{A}$ it holds that $\log(|\Omega|\mu(x)) \geq \log(2) = 1$, thus, $A \geq \mu(\mathcal{A})$. Let $\mathcal{B} \subseteq \Omega$ be the set of elements with $\frac{1}{|\Omega|} \leq \mu(x) < \frac{2}{|\Omega|}$. Denote

$$B = \sum_{x \in \mathcal{B}} \mu(x) \log(|\Omega|\mu(x)).$$

Observe that for every $x \in \mathcal{B}$ it holds that $\log(|\Omega|\mu(x)) \geq \log(1) = 0$, thus, $B \geq 0$. The claim now follows using Lemma 5.11, as

$$I = A + B - I^{neg}(\mu) > \mu(\mathcal{A}) - 4I^{0.25} \log\left(\tfrac{1}{I^{0.25}}\right),$$

where $I^{neg}(\mu)$ is defined as in Lemma 5.11. □

Lemma 5.13. *Let $\mu : \Omega \to [0,1]$ be a distribution satisfying $I = \mathbf{I}(\mu) \leq 0.01$. Let $c \in \mathbb{N}$. Let $\mu_1, \ldots, \mu_{2^c} : \Omega \to [0,1]$ be 2^c distributions. Let $\alpha_1, \ldots, \alpha_{2^c} \in [0,1]$. Assume that*

$$\mu = \sum_{i \in [2^c]} \alpha_i \mu_i.$$

For $i \in [2^c]$, let $I_i = \mathbf{I}(\mu_i)$. Let $h \in \mathbb{R}^+$. Let $\mathcal{B} \subseteq [2^c]$ be the set of indices i such that $I_i > c + h + 1$. Then,

$$\sum_{i \in \mathcal{B}} \alpha_i < 2^{-h} + 5I^{0.1}.$$

Proof. For $i \in [2^c]$ define

$$M_i = \sum_{x \in \mathrm{supp}(\mu)} \mu_i(x) \log(|\Omega|\mu(x)).$$

Let $i \in [2^c]$ be such that $\alpha_i \geq 2^{-(h+c)}$. Let $x \in \Omega$. It holds that

$$\mu(x) \geq \alpha_i \mu_i(x) \geq 2^{-(h+c)} \mu_i(x). \tag{6}$$

Using Equation 6, we have

$$\begin{aligned} I_i &= \sum_{x \in \mathrm{supp}(\mu_i)} \mu_i(x) \log\left(|\Omega|\mu_i(x)\right) \\ &\leq \sum_{x \in \mathrm{supp}(\mu)} \mu_i(x) \log\left(|\Omega| \cdot 2^{h+c} \mu(x)\right) \\ &\leq M_i + h + c. \end{aligned}$$

Therefore, $i \in \mathcal{B}$ implies $\alpha_i < 2^{-(h+c)}$ or $M_i > 1$.

Let $\mathcal{M} \subseteq [2^c]$ be the set of indices i such that $M_i > 1$. As was done in Lemma 5.11, define $\mathcal{A} \subseteq \Omega$ to be the set of elements $x \in \Omega$ with $\mu(x) < \frac{1}{|\Omega|}$. Note that for such elements it holds that $\log(|\Omega|\mu(x)) < 0$. Also, recall the definition of I^{neg} formulated in Lemma 5.11:

$$I^{neg}(\mu) = -\sum_{x \in \mathcal{A}} \mu(x) \log(|\Omega|\mu(x)).$$

Using the bound on $I^{neg}(\mu)$ offered by Lemma 5.11, we get

$$\begin{aligned}\sum_{i\in\mathcal{M}}\alpha_i &\leq \sum_{i\in\mathcal{M}}\alpha_i M_i = \sum_{i\in\mathcal{M}}\alpha_i\left(\sum_{x\in\text{supp}(\mu)}\mu_i(x)\log(|\Omega|\mu(x))\right)\\ &\leq \sum_{x\in(\text{supp}(\mu)\setminus\mathcal{A})}\left(\sum_{i\in\mathcal{M}}\alpha_i\mu_i(x)\right)\log(|\Omega|\mu(x))\\ &\leq \sum_{x\in(\text{supp}(\mu)\setminus\mathcal{A})}\mu(x)\log(|\Omega|\mu(x))\\ &\leq I + I^{neg}(\mu) \leq I + 4I^{0.1} \leq 5I^{0.1}.\end{aligned}$$

As mentioned above, $i\in\mathcal{B}$ implies $\alpha_i < 2^{-(h+c)}$ or $M_i > 1$. Therefore,

$$\sum_{i\in\mathcal{B}}\alpha_i < 2^{-(h+c)}\cdot 2^c + \sum_{i\in\mathcal{M}}\alpha_i < 2^{-h} + 5I^{0.1}.$$

□

6 Operations Over Games

6.1 General Notation

Decomposing the inputs. Let G be a game with parameters (k, d, P_X, P_Y), with underlying tree T. Let v be a vertex of T. We label each of the 2^k edges leaving v by a unique label from the set $[2^k]$. The input X can be written as $X = \{X_v\}_{v\in\text{Even}(T)}$, where $X_v \in [2^k]$ is the label of the unique edge leaving v that is contained in X. Similarly, we can write Y as $Y = \{Y_v\}_{v\in\text{Odd}(T)}$, where $Y_v \in [2^k]$ is the label of the unique edge leaving v that is contained in Y.

Let v be a vertex of T. We denote by T_v the subtree of T rooted at v. Slightly abusing notation, we denote the root of the tree by the number 0, and the 2^k children of the root by the numbers $1,\ldots,2^k$ (such that the name of a vertex is consistent with the label of the edge that reaches that vertex). In particular, for $j\in[2^k]$, we denote by T_j the subtree of T rooted at the j^{th} child of the root. We will often write X as

$$X = \left(X_0, X_{T_1}, \ldots, X_{T_{2^k}}\right),$$

where $X_{T_v} = \{X_{v'}\}_{v'\in T_v\cap\text{Even}(T)}$ is the restriction of the information in X to vertices of T_v,

and X_0 is $X_{root(T)}$. Similarly, we will often write Y as

$$Y = \left(Y_{T_1}, \ldots, Y_{T_{2^k}}\right),$$

where $Y_{T_v} = \{Y_{v'}\}_{v' \in T_v \cap \mathrm{Odd}(T)}$ is the restriction of the information in Y to vertices of T_v. For a vertex v of T, we denote by $(X,Y)_{T_v}$ the pair (X_{T_v}, Y_{T_v}).

The "correct" path. Given a game G and inputs X and Y, let $V_0, \ldots, V_d$ be the d vertices on the path from the root to the leaf of the underlying tree, defined by the inputs X and Y, where V_0 is the root and $V_d = L(X,Y)$ is the leaf. Note that $V_0, \ldots, V_d$ are random variables that depend on the inputs X and Y, and note that $V_1 = X_0$. Let $E_1, \ldots, E_d$ be the edges $E_1 = (V_0, V_1), E_2 = (V_1, V_2), \ldots, E_d = (V_{d-1}, V_d)$.

Constructing a game given inputs. Given a game G with parameters (k, d, P_X, P_Y), we sometimes want to consider variants of G played with different input distributions (for example, when the distributions are conditioned on an event). For that reason we introduce the following notation. Let $k, d' \in \mathbb{N}$, and let T be the 2^k-ary tree of depth d'. Recall that we denote by $\mathcal{X}(T)$ the set of possible inputs for the first player, and by $\mathcal{Y}(T)$ the set of possible inputs for the second player. For a pair of independent random variables X', Y', over the sets $\mathcal{X}(T)$ and $\mathcal{Y}(T)$ respectively, we denote by $G_{X'Y'}$ the game with parameters $(k, d', P_{X'}, P_{Y'})$.

Distribution over games. In the proof we often apply an operation to a given game G, and obtain a "distribution $\mathcal{G}$ over games". By that we mean that we reach a distribution $\mathcal{G}$ whose domain is a set of new games $G_1, \ldots, G_m$ (not necessarily different), where $m \in \mathbb{N}$. For every $i \in [m]$, the game G_i is obtained with probability $\alpha_i \in [0,1]$, where $\sum_{i \in [m]} \alpha_i = 1$.

6.2 Conditioning a Game

Conditioning a game on a random variable. Let G be a game with parameters (k, d, P_X, P_Y). Let W be a random variable that is conditionally independent of the input X given the input Y. We think of W as a probabilistic function of Y (independent of X), that is, without loss of generality we think of W as determined by Y and an independent random string R (that is independent of X, Y). The operation of conditioning G on W results in a distribution $\mathcal{G}$ over games, obtained as follows: For every $w \in \mathrm{supp}(W)$ we will have a game $G_{XY|W=w}$ with parameters $(k, d, P_X, P_{Y|W=w})$. The distribution $\mathcal{G}$ will have the

domain $\left\{G_{XY|W=w}\right\}_{w\in\text{supp}(W)}$. For every $w \in \text{supp}(W)$, the game $G_{XY|W=w}$ is obtained with probability $\Pr[W = w]$.

In the same way, if W be a random variable that is conditionally independent of the input Y given the input X, the operation of conditioning G on W results in a distribution $\mathcal{G}$ over games, obtained as follows: For every $w \in \text{supp}(W)$ we will have a game $G_{XY|W=w}$ with parameters $(k, d, P_{X|W=w}, P_Y)$. The distribution $\mathcal{G}$ will have the domain $\left\{G_{XY|W=w}\right\}_{w\in\text{supp}(W)}$. For every $w \in \text{supp}(W)$, the game $G_{XY|W=w}$ is obtained with probability $\Pr[W = w]$.

In both cases, we denote by $G_{XY|W}$ the random game chosen according to the distribution $\mathcal{G}$, such that the game $G_{XY|W=w}$ is chosen when $W = w$. In particular, for every $w \in \text{supp}(W)$, the game $G_{XY|W}$ is $G_{XY|W=w}$ with probability $\Pr[W = w]$. The notation $G_{XY|W=w}$ will sometimes be abbreviated as $G_{XY|w}$.

Conditioning a game on a sequence of random variables. Let G be a game with parameters (k, d, P_X, P_Y). Let $W = W_1, \ldots, W_m$ be a sequence of random variables, where $m \in \mathbb{N}$. We say that W is a *feasible transcript* if for every $i \in [m]$, either the variables W_i and X are conditionally independent given $W_1, \ldots, W_{i-1}, Y$, or the variables W_i and Y are conditionally independent given $W_1, \ldots, W_{i-1}, X$. The operation of conditioning G on W results in a distribution $\mathcal{G}$ over games, obtained by conditioning on $W_1, \ldots, W_m$ one by one.

Intuitively, a feasible transcript is a possible transcript of a communication protocol between two players that hold X, Y respectively.

Lemma 6.1. *Let G be a game with inputs X and Y. Let W be a feasible transcript. Then,*

$$\begin{aligned} P_X &= \sum_{w\in\text{supp}(W)} \Pr_W[W = w] \cdot P_{X|W=w}, \\ P_Y &= \sum_{w\in\text{supp}(W)} \Pr_W[W = w] \cdot P_{Y|W=w}, \\ P_{X,Y} &= \sum_{w\in\text{supp}(W)} \Pr_W[W = w] \cdot P_{XY|W=w}. \end{aligned}$$

Proof. Follows immediately from the complete probability formula. □

Lemma 6.2. *Let G be a game with inputs X and Y. Let $W = W_1, \ldots, W_m$ be a feasible transcript, where $m \in \mathbb{N}$. Assume without loss of generality that $W_1, \ldots, W_m$ are bits. Let $m_1 \in \mathbb{N}$ be the number of indices $i \in [m]$, such that the variables W_i and X are* not *conditionally independent given $W_1, \ldots, W_{i-1}, Y$.*

Let $h \in \mathbb{R}^+$. Let $\mathcal{B} \subseteq \text{supp}(W)$ be the set of strings w such that $\mathbf{H}_\infty(P_{X_0|W=w}) <$

$\mathbf{H}_\infty(P_{X_0}) - m_1 - h$. *Then,*

$$\sum_{w \in \mathcal{B}} \Pr_W[W = w] < 2^{-h}.$$

Moreover,

$$\mathop{\mathbf{E}}_W \left[\mathbf{H}_\infty(P_{X_0|W})\right] \geq \mathbf{H}_\infty(P_{X_0}) - m_1 - 4.$$

Proof. Denote by $\mathcal{A}$ the set of indices $i \in [m]$, such that the variables W_i and X are *not* conditionally independent given $W_1, \ldots, W_{i-1}, Y$. Note that $|\mathcal{A}| = m_1$. Denote $W_\mathcal{A} = \{W_i\}_{i \in \mathcal{A}}$ and $W_{\bar{\mathcal{A}}} = \{W_i\}_{i \in \bar{\mathcal{A}}}$. Thus, each bit of $W_{\bar{\mathcal{A}}}$ is conditionally independent of X, given all previous bits of W and Y.

Without loss of generality we can assume that each bit $i \in W_{\bar{\mathcal{A}}}$ is determined by $W_1, \ldots, W_{i-1}, Y$ and an independent random string R_i (that is independent of X, Y and all other random strings), and in the same way, each bit $i \in W_\mathcal{A}$ is determined by $W_1, \ldots, W_{i-1}, X$ and an independent random string R_i (that is independent of X, Y and all other random strings). Denote $R_\mathcal{A} = \{R_i\}_{i \in \mathcal{A}}$ and $R_{\bar{\mathcal{A}}} = \{R_i\}_{i \in \bar{\mathcal{A}}}$. Denote $Y' = (Y, R_{\bar{\mathcal{A}}})$.

For every $y' \in \mathrm{supp}(Y')$, the distribution $P_{X_0|Y'=y'}$ is just the distribution P_{X_0}, since Y' is independent of X. Moreover, since W is a feasible transcript, the random variables $(X, R_\mathcal{A})$ and $(Y, R_{\bar{\mathcal{A}}})$ are conditionally independent given W (this is a standard fact in communication complexity and is easily proved by induction on m). Hence, for every $(w, y') \in \mathrm{supp}((W, Y'))$ we have $P_{X_0|W=w,Y'=y'} = P_{X_0|W=w}$.

Fix $y' \in \mathrm{supp}(Y')$. The distribution P_{X_0} can be written as a sum

$$P_{X_0} = P_{X_0|Y'=y'} = \sum_{w \in \mathrm{supp}(W|Y'=y')} \Pr[W = w \mid Y' = y'] \cdot P_{X_0|W=w,Y'=y'}.$$

$$= \sum_{w \in \mathrm{supp}(W|Y'=y')} \Pr[W = w \mid Y' = y'] \cdot P_{X_0|W=w}.$$

Note that in the last sum at most 2^{m_1} summands are non-zero, because every bit of $W_{\bar{\mathcal{A}}}$ is determined by Y' and all previous bits of W (so by induction on m we have that $\Pr[W = w \mid Y' = y']$ is non-zero for at most 2^{m_1} possibilities of w).

Recall that $\mathcal{B} \subseteq \mathrm{supp}(W)$ is the set of $w \in \mathrm{supp}(W)$ such that

$$\mathbf{H}_\infty(P_{X_0|W=w}) < \mathbf{H}_\infty(P_{X_0}) - m_1 - h.$$

Using Lemma 5.9, it holds that

$$\sum_{w \in \mathcal{B}} \Pr[W = w \mid Y' = y'] < 2^{-h}.$$

The first part of the lemma follows as

$$\begin{aligned}\sum_{w\in\mathcal{B}} \Pr_W[W=w] &= \sum_{y'\in\mathrm{supp}(Y')} \Pr[Y'=y'] \sum_{w\in\mathcal{B}} \Pr[W=w \mid Y'=y'] \\ &< \sum_{y'\in\mathrm{supp}(Y')} \Pr[Y'=y'] \cdot 2^{-h} = 2^{-h}.\end{aligned}$$

Lemma 5.9 also implies that

$$\sum_{w\in\mathrm{supp}(W|Y'=y')} \Pr[W=w \mid Y'=y'] \cdot \mathbf{H}_\infty(P_{X_0|W=w}) \geq \mathbf{H}_\infty(P_{X_0}) - m_1 - 4.$$

The second part of the lemma follows as

$$\begin{aligned}&\mathop{\mathbf{E}}_W\left[\mathbf{H}_\infty(P_{X_0|W})\right] = \\ &\sum_{y'\in\mathrm{supp}(Y')} \Pr[Y'=y'] \sum_{w\in\mathrm{supp}(W|Y'=y')} \Pr[W=w \mid Y'=y'] \cdot \mathbf{H}_\infty(P_{X_0|W=w}) \geq \\ &\sum_{y'\in\mathrm{supp}(Y')} \Pr[Y'=y'] \left(\mathbf{H}_\infty(P_{X_0}) - m_1 - 4\right) = \\ &\mathbf{H}_\infty(P_{X_0}) - m_1 - 4.\end{aligned}$$

□

Lemma 6.3. *Let G be a game with inputs X and Y. Let $\epsilon, \delta \in [0,1]$. Let W be a feasible transcript. Then, there exists $\{\delta_w\}_{w\in\mathrm{supp}(W)}$, $\delta_w \in [0,1]$, such that $\mathbf{E}_W[\delta_W] = \delta$, and*

$$\mathsf{CC}_{\epsilon,\delta}(G) \geq \mathop{\mathbf{E}}_W\left[\mathsf{CC}_{\epsilon,\delta_W}(G_{XY|W})\right].$$

Proof. Let $\Pi \in \mathcal{P}_{\epsilon,\delta}$ be a deterministic protocol such that $\mathsf{CC}(\Pi) = \mathsf{CC}_{\epsilon,\delta}(G)$. Let $w \in \mathrm{supp}(W)$. We can view Π as a protocol for the game $G_{XY|W=w}$. Denote this protocol by Π_w. Define δ_w to be the probability that Π_w errs on the game $G_{XY|W=w}$, where the probability is over the inputs and the channel's noise. Recall that

$$P_{X,Y} = \sum_{w\in\mathrm{supp}(W)} \Pr_W[W=w] \cdot P_{XY|W=w}.$$

Hence,

$$\mathsf{CC}_{\epsilon,\delta}(G) = \mathsf{CC}(\Pi) = \mathop{\mathbf{E}}_W\left[\mathsf{CC}(\Pi_W)\right] \geq \mathop{\mathbf{E}}_W\left[\mathsf{CC}_{\epsilon,\delta_W}(G_{XY|W})\right],$$

and $\mathbf{E}_W[\delta_W] = \delta$. □

Let G be a game with inputs X and Y, and let Π be a deterministic protocol for G. We will often consider the first t communication bits *received* by the players during an execution of the protocol Π on inputs X and Y. Denote these bits by $W = W_1, \ldots, W_t$. We refer to W as the *received transcript* of the first t rounds of Π. Note that the bits received by the players may differ from the bits sent by the players due to the channel's error. Observe that W is a feasible transcript, and that W is a random variable that depends on X and Y.

Lemma 6.4. *Let G be a game with inputs X and Y. Let $\epsilon, \delta \in [0,1]$. Let $\Pi \in \mathcal{P}_{\epsilon,\delta}$ be a deterministic protocol for G. Let $t \in \mathbb{N}$. Assume that, with probability 1, the protocol Π has at least t rounds, and let W be the received transcript of the first t rounds of Π. Then, there exists $\{\delta_w\}_{w \in \mathrm{supp}(W)}$, $\delta_w \in [0,1]$, such that $\mathbf{E}_W[\delta_W] = \delta$, and*

$$\mathsf{CC}(\Pi) \geq t + \mathop{\mathbf{E}}_{W}\left[\mathsf{CC}_{\epsilon,\delta_W}(G_{XY|W})\right].$$

Proof. Let $w \in \mathrm{supp}(W)$. We denote by Π_w the protocol consisting of rounds $t+1, t+2, \ldots$ of Π, when the received transcript of the first t rounds of Π is w. We view Π_w as a protocol for the game $G_{XY|W=w}$. Define δ_w to be the probability that Π_w errs on the game $G_{XY|W=w}$, where the probability is over the inputs and the channel's noise. Recall that

$$P_{X,Y} = \sum_{w \in \mathrm{supp}(W)} \Pr_W[W = w] \cdot P_{XY|W=w}.$$

Hence,

$$\mathsf{CC}(\Pi) = t + \mathop{\mathbf{E}}_{W}\left[\mathsf{CC}(\Pi_W)\right] \geq t + \mathop{\mathbf{E}}_{W}\left[\mathsf{CC}_{\epsilon,\delta_W}(G_{XY|W})\right],$$

and $\mathbf{E}_W[\delta_W] = \delta$. □

6.3 Reducing a Game

Let $k, d' \in \mathbb{N}$, and let T' be the 2^k-ary tree of depth d'. Recall that we denote by $\mathcal{X}(T')$ the set of possible inputs for the first player, and by $\mathcal{Y}(T')$ the set of possible inputs for the second player. Recall that, for a pair of independent random variables X', Y', over the sets $\mathcal{X}(T')$ and $\mathcal{Y}(T')$ respectively, we denote by $G_{X'Y'}$ the game with parameters $(k, d', P_{X'}, P_{Y'})$.

Let G be a game with parameters (k, d, P_X, P_Y), with underlying tree T. Let v be a vertex of T. We often consider the *reduced game*, obtained by restricting G to the subtree T_v, and played with inputs X_{T_v} and Y_{T_v}. Formally, if $v \in \mathrm{Even}(T)$, then the reduced game is $G_{(X,Y)_{T_v}}$. That is, the game with parameters $(k, d', P_{X_{T_v}}, P_{Y_{T_v}})$, where d' is the depth of T_v. If $v \in \mathrm{Odd}(T)$, then the reduced game is $G_{(Y,X)_{T_v}}$. That is, the game with parameters $(k, d', P_{Y_{T_v}}, P_{X_{T_v}})$, where d' is the depth of T_v. Note that in both cases the depth of the

game G is reduced by the depth of v, and that in the case where $v \in \mathrm{Odd}(T)$, the roles of the two inputs X and Y are switched, as odd layers of T are even layers of T_v, and vice versa.

The vertex v will sometimes be a random variable V, in which case, the reduced game $G_{(X,Y)_{T_V}}$ or $G_{(Y,X)_{T_V}}$ can be viewed as a random variable, or as a distribution over games.

Recall that we denote by V_1 the first non-root vertex on the correct path. We will often reduce the game G to the subtree T_{V_1} after conditioning G on the value of V_1, and obtain the game $G_{(Y,X|V_1)_{T_{V_1}}}$. Note that conditioning G on the value of V_1 can be viewed as "revealing" the value of V_1 to the second player (the first player already knows this value). Since the first edge is now known, the players can turn to play the game G reduced to T_{V_1}.

Formally, we can view $G_{(Y,X|V_1)_{T_{V_1}}}$ as a distribution $\mathcal{G}$ over games, obtained as follows: For every $v_1 \in \mathrm{supp}(V_1)$ we will have a game $G_{(Y,X|V_1=v_1)_{T_{v_1}}}$ with parameters $(k, d-1, P_{(Y|V_1=v_1)_{T_{v_1}}}, P_{(X|V_1=v_1)_{T_{v_1}}})$. The distribution $\mathcal{G}$ will have the domain $\left\{G_{(Y,X|V_1=v_1)_{T_{v_1}}}\right\}_{v_1 \in \mathrm{supp}(V_1)}$. For every $v_1 \in \mathrm{supp}(V_1)$, the game $G_{(Y,X|V_1=v_1)_{T_{v_1}}}$ is obtained with probability $\Pr[V_1 = v_1]$.

We also view $G_{(Y,X|V_1)_{T_{V_1}}}$ as the random game chosen according to the distribution $\mathcal{G}$, such that the game $G_{(Y,X|V_1=v_1)_{T_{v_1}}}$ is chosen when $V_1 = v_1$. In particular, for every $v_1 \in \mathrm{supp}(V_1)$, the game $G_{(Y,X|V_1)_{T_{V_1}}}$ is $G_{(Y,X|V_1=v_1)_{T_{v_1}}}$ with probability $\Pr[V_1 = v_1]$. The notation $G_{(Y,X|V_1=v_1)_{T_{v_1}}}$ will sometimes be abbreviated as $G_{(Y,X|v_1)_{T_{v_1}}}$.

Remark 6.5. *We remark that since the random variables Y and V_1 are independent, the distribution $P_{(Y|V_1=v_1)_{T_{v_1}}}$ is the same as $P_{Y_{T_{v_1}}}$. Nevertheless, we will usually prefer not to omit the conditioning on $V_1 = v_1$ as it may become significant when further conditioning on an additional random variable (specifically, if the additional variable is a function of both Y and V_1).*

Lemma 6.6. *Let $Z_1, \ldots, Z_m$ be m random variables, and let J be a random variable taking values in $[m]$. Assume that J is independent of $Z_1, \ldots, Z_m$. Then,*

$$\mathop{\mathbf{E}}_{J}\left[\mathbf{I}(Z_J)\right] \leq 2^{-\mathbf{H}_\infty(J)}\mathbf{I}(Z_1, \ldots, Z_m).$$

Proof. Using the supper-additivity of information (Proposition 4.2)

$$\mathop{\mathbf{E}}_{J}\left[\mathbf{I}(Z_J)\right] = \sum_{j \in [m]} \Pr[J = j] \cdot \mathbf{I}(Z_j) \leq \max_{j \in [m]}\{\Pr[J = j]\} \sum_{j \in [m]} \mathbf{I}(Z_j) \leq 2^{-\mathbf{H}_\infty(J)}\mathbf{I}(Z_1, \ldots, Z_m).$$

□

Lemma 6.7. *Let G be a game with parameters (k, d, P_X, P_Y), with underlying tree T. Consider the game $G_{(Y,X|V_1)_{T_{V_1}}}$, with parameters $(k, d-1, P_{(Y|V_1)_{T_{V_1}}}, P_{(X|V_1)_{T_{V_1}}})$. Recall that*

$V_1 = X_0$. *Then,*

1. $\mathbf{E}_{V_1}\left[\mathbf{I}((Y|V_1)_{T_{V_1}})\right] \leq 2^{-\mathbf{H}_\infty(X_0)}\mathbf{I}(Y)$.
2. $\mathbf{E}_{V_1}\left[\mathbf{I}((X|V_1)_{T_{V_1}})\right] \leq \mathbf{I}(X) - \mathbf{I}(X_0)$.

Proof. The first part of the lemma follows directly from Lemma 6.7, since $V_1 = X_0$ is independent of Y (see also Remark 6.5).

We turn to prove the second part:

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((X|V_1)_{T_{V_1}})\right] = \\
&\sum_{v_1\in[2^k]} \Pr[V_1 = v_1]\cdot \mathbf{I}\left((X|V_1 = v_1)_{T_{v_1}}\right) = \\
&\sum_{v_1\in[2^k]} \Pr[V_1 = v_1]\cdot \mathbf{I}\left(X_{T_{v_1}}|V_1 = v_1\right) \leq \\
&\sum_{v_1,j\in[2^k]} \Pr[V_1 = v_1]\cdot \mathbf{I}\left(X_{T_j}|V_1 = v_1\right) = \\
&\sum_{v_1,j\in[2^k]} \Pr[V_1 = v_1]\cdot \left(\log(|\mathcal{Y}(T_j)|) - \mathbf{H}\left(X_{T_j}|V_1 = v_1\right)\right) = \\
&2^k\log(|\mathcal{Y}(T_1)|) - \sum_{j\in[2^k]}\sum_{v_1\in[2^k]} \Pr[V_1 = v_1]\cdot \mathbf{H}\left(X_{T_j}|V_1 = v_1\right) = \\
&(\log(|\mathcal{X}(T)|) - k) - \sum_{j\in[2^k]} \mathbf{H}\left(X_{T_j}|V_1\right) = \\
&\log(|\mathcal{X}(T)|) - \mathbf{I}(X_0) - \mathbf{H}(X_0) - \sum_{j\in[2^k]} \mathbf{H}\left(X_{T_j}|X_0\right) \leq \\
&\log(|\mathcal{X}(T)|) - \mathbf{I}(X_0) - \mathbf{H}(X_0) - \mathbf{H}\left(X_{T_1},\ldots,X_{T_{2^k}}|X_0\right) = \\
&\log(|\mathcal{X}(T)|) - \mathbf{I}(X_0) - \mathbf{H}\left(X_0, X_{T_1},\ldots,X_{T_{2^k}}\right) = \\
&\mathbf{I}(X) - \mathbf{I}(X_0).
\end{aligned}
$$

□

7 Main Lemmas and Proof of Theorem 2

For a game G with inputs X and Y, we define: $I_1(G) = \mathbf{I}(X)$, $I_2(G) = \mathbf{I}(Y)$, $I(G) = I_1(G) + I_2(G)$, and $\kappa(G) = \mathbf{H}_\infty(X_0)$. In all that comes below, we assume that k is sufficiently large (say, $k > 10^{100}$).

Definition 7.1 (**Nice Game**). *Let G be a game with parameters (k, d, P_X, P_Y). The game G is called* nice *if it satisfies all the following conditions:*

1. $\kappa(G) \geq 0.5k$.

2. $I_1(G) \leq 10k$.

3. $I_2(G) \leq 20k$.

4. *The distributions* P_X, P_Y *are* $(0.01k)$*-flat.*

Lemma 7.2 (**Communication Lower-Bound for Nice Games**). *Let* G *be a nice game with parameters* (k, d, P_X, P_Y). *Let* $\epsilon = \frac{2000\log(k)}{k^2}$ *and* $\delta \in [0,1]$. *Then,*

$$\mathsf{CC}_{\epsilon,\delta}(G) \geq d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - (k - \kappa(G)) - 100k.$$

Lemma 7.3 (**Communication Lower-Bound for Non-Nice Games**). *Let* G *be a game with parameters* (k, d, P_X, P_Y) *(*G *may not be nice). Let* $\epsilon = \frac{2000\log(k)}{k^2}$ *and* $\delta \in [0,1]$. *Then,*

$$\mathsf{CC}_{\epsilon,\delta}(G) \geq d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - 100I(G) - 1000k.$$

Lemmas 7.2 and 7.3 are proven simultaneously, by a mutual recursion, where the two lemmas use each other recursively. The proof is by induction on d, the depth of the game G. To prove Lemma 7.2 for a given d, we assume the validity of Lemma 7.3 for $d' < d$. To prove Lemma 7.3 for a given d, we assume the validity of Lemma 7.2 for $d' < d$. Hence, both lemmas are correct for every d.

Observe that the base case $d < 90$ is trivial for both Lemma 7.2 and Lemma 7.3, since the bounds obtained are negative. Moreover, in both lemmas we can assume that $\delta \leq 0.5$ since otherwise the bounds obtained are negative.

Equipped with Lemma 7.2 we can now prove Theorem 2.

Proof of Theorem 2. Recall that any lower bound proven for balanced deterministic protocols for G holds for general deterministic protocols, up to an additive $2k$ term. Therefore, by Lemma 7.2, since G is a nice game with $\kappa(G) = k$, for any general deterministic protocol for G with noise rate ϵ and error δ (not necessarily balanced), it holds that

$$\mathsf{CC}(\Pi) \geq d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - 102k.$$

Let $\Pi \in \mathcal{P}^*_{\epsilon,\delta}$ be a probabilistic protocol for G. Let R be the (finite) random string used by Π. For every possible value r of R, we denote by Π_r the deterministic protocol obtained from Π when given the random string $R = r$. Denote by δ_r the error of the deterministic

protocol Π_r. Observe that $\mathbf{E}_R[\delta_R] = \delta$. It holds that

$$\begin{aligned}\mathsf{CC}(\Pi) = \mathop{\mathbf{E}}_{R}[\mathsf{CC}(\Pi_R)] &\geq d \cdot (k + 0.1\log(k)) \cdot (1 - 2\mathop{\mathbf{E}}_{R}[\delta_R]) - 102k \\ &\geq d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - 102k,\end{aligned}$$

and the assertion follows. □

8 Communication Lower-Bound for Nice Games (Proof of Lemma 7.2)

8.1 Proof of Lemma 7.2

Let G be a nice game with parameters (k, d, P_X, P_Y), and underlying tree T. Let Π be a protocol for G with noise rate ϵ and error δ, where ϵ and δ are as specified by Lemma 7.2. We assume that $d \geq 90$ and $\delta \leq 0.5$, otherwise, as explained before, the claim of the lemma holds trivially. Our goal is to bound $\mathsf{CC}(\Pi)$ in the other cases.

8.1.1 Ensuring $2k$ Rounds

We first claim that Π *always* communicates at least $2k$ bits. Recall from Subsection 3.2, that we only deal with balanced protocols. Thus, if Π communicates at most $2k$ bits with positive probability, then it must communicate at most $2k$ bits with probability 1, and hence $\mathsf{CC}(\Pi) \leq 2k$. The following lemma shows that if $\mathsf{CC}(\Pi) \leq 2k$ then $\delta > 0.5$ or $d < 90$, which contradicts our assumption.

Lemma 8.1. *If $d \geq 90$ and $\delta \leq 0.5$ then $\mathsf{CC}(\Pi) > 2k$.*

Proof. Assume that $d \geq 90$ and $\mathsf{CC}(\Pi) \leq 2k$, we will prove that $\delta > 0.5$. Consider the input X of the first player. Since G is a nice game, it holds that $\mathbf{I}(X) \leq 10k$, and that P_X is $(0.01k)$-flat. Using Lemma 5.2,

$$\mathbf{H}_\infty(X) \geq \mathbf{H}(X) - 0.01k = \log(|\mathcal{X}(T)|) - \mathbf{I}(X) - 0.01k \geq \log(|\mathcal{X}(T)|) - 10.01k.$$

Recall from Subsection 3.2, that we only deal with balanced protocols. Since Π is a balanced protocol and $\mathsf{CC}(\Pi) \leq 2k$, it must be the case that Π *always* ends after at most $2k$ rounds. We assume without loss of generality that Π always ends after exactly $2k$ rounds. Denote by $W = W_1, \ldots, W_{2k}$ the received transcript of the $2k$ rounds of Π (the received transcript is defined in Subsection 6.2).

We assume for the rest of the proof of Lemma 8.1 that no noise occurred in the channel, and hence all the bits that were sent were received correctly. This event occurs with probability of at least $1 - 2\epsilon k > 0.99$. The argument below assumes that this event occurs.

Let $\mathcal{B}$ be the set of strings $w \in \{0,1\}^{2k}$, such that when the received transcript of the $2k$ rounds of Π is w, the protocol Π errs with probability at most 0.75. For every $w \in \mathcal{B}$ the protocol Π declares an answer (since no noise occurred in the channel, both players declare the same answer), and this answer must be correct with probability at least 0.25. The answer of the protocol is a leaf of the tree T.

Fix $w \in \mathcal{B}$ and let $e = \{e_1, e_2, e_3, \ldots\}$ be the edges on the path to the leaf declared by the protocol. Let $e' = \{e_1, e_3, e_5, \ldots\}$. Since the answer of the protocol is correct with probability of at least 0.25, we have that, conditioned on the event $W = w$, the edges in e are contained in $X \cup Y$ with probability of at least 0.25. That is, $\Pr[e \subset X \cup Y | W = w] \geq 0.25$. Hence, $\Pr[e' \subset X | W = w] \geq 0.25$.

Denote by X' a random variable over $\mathcal{X}(T)$ chosen according to the distribution $P_{X|W=w}$ and denote by U a random variable over $\mathcal{X}(T)$ chosen according to the uniform distribution. Thus, $\Pr[e' \subset X'] \geq 0.25$, while $\Pr[e' \subset U] = 2^{-\lceil 0.5d \rceil k} \leq 2^{-45k}$. The last two inequalities imply that for every $w \in \mathcal{B}$, we have $\mathbf{H}_\infty(X|W = w) \leq \log(|\mathcal{X}(T)|) - (45k - 2)$.

Recall that

$$P_X = \sum_{w \in \{0,1\}^{2k}} \Pr_W[W = w] \cdot P_{X|W=w}.$$

Using Lemma 5.9, with $c = 2k$ and $h = 30k$, it holds that

$$\sum_{w \in \mathcal{B}} \Pr_W[W = w] < 2^{-30k}.$$

Thus, the error of the protocol Π satisfies $\delta > 0.75 \cdot (1 - 2^{-30k}) \cdot 0.99 > 0.5$. □

8.1.2 Steps in the Analysis

Consider the first $t = \lfloor \kappa(G) + 0.25k \rfloor$ bits communicated by the protocol. Recall that we assume that it is known in advance which player sends a bit in each round (see Subsection 3.2). Let t_1 and t_2 be the number of bits sent by each player in this block of t bits, $t_1 + t_2 = t$. Denote by $W = W_1, \ldots, W_t$ the received transcript of the first t rounds of Π (the received transcript is defined in Subsection 6.2).

Revealing the first correct vertex V_1. Recall that we denote by V_1 the first non-root vertex on the correct path. We first "reveal" the value of V_1 to the second player (the first player already knows this value). That is, we condition the game G on the value of V_1. We

then reduce the game to the subtree T_{V_1}. That is, we consider the game $G_{(Y,X|V_1)_{T_{V_1}}}$, with parameters $(k, d-1, P_{(Y|V_1)_{T_{V_1}}}, P_{(X|V_1)_{T_{V_1}}})$ (see Subsection 6.3).

Denote by $\mathcal{B}_1 \subseteq [2^k]$ the set of all vertices v_1 such that $\mathbf{I}\left((Y|V_1 = v_1)_{T_{v_1}}\right) \geq 2^{-0.25k}$. Denote $\bar{\mathcal{B}_1} = [2^k] \setminus \mathcal{B}_1$.

Revealing B. Fix $v_1 \in \bar{\mathcal{B}_1}$. For every $v_2 \in \mathrm{supp}(P_{V_2|V_1=v_1})$, we define the value $B(v_2|v_1) \in \{0,1\}$ as follows: $B(v_2|v_1) = 1$ if and only if $\Pr_{V_2}[V_2 = v_2|V_1 = v_1] > 2 \cdot 2^{-k}$. (For $v_1 \in \mathcal{B}_1$, it will be convenient to define $B(v_2|v_1) = 0$ for every v_2, although this value will never be used). Define $B = B(V_2|V_1)$. Roughly speaking, the bit B indicates whether the probability of the correct second vertex is significantly larger than the the probability of a random child of the correct first vertex.

We "reveal" the value of B to the first player (the second player already knows this value). That is, we condition the game $G_{(Y,X|V_1)_{T_{V_1}}}$ on the value of B, and consider the game $G_{(Y,X|V_1)_{T_{V_1}}|B} = G_{(Y,X|V_1,B)_{T_{V_1}}}$, with parameters $(k, d-1, P_{(Y|V_1,B)_{T_{V_1}}}, P_{(X|V_1,B)_{T_{V_1}}})$ (see Subsection 6.2).

Denote $\mathcal{B}_2 = \{1\}$. Denote $\bar{\mathcal{B}_2} = \{0,1\} \setminus \mathcal{B}_2 = \{0\}$.

Revealing the received transcript W. Next, we "reveal" the value of W to both players. That is, we condition the game $G_{(Y,X|V_1,B)_{T_{V_1}}}$ on the value of W, and consider the game $G_{(Y,X|V_1,B)_{T_{V_1}}|W} = G_{(Y,X|V_1,B,W)_{T_{V_1}}}$, with parameters $(k, d-1, P_{(Y|V_1,B,W)_{T_{V_1}}}, P_{(X|V_1,B,W)_{T_{V_1}}})$ (see Subsection 6.2).

Revealing the noise indicator E. Recall that t_1 is the number of bits sent by the first player in the block of t bits that we consider. We define the random variable $E \in \{0,1\}$ as follows: $E = 1$ if and only if *exactly* one of the t_1 bits sent by the first player, was received incorrectly by the second player, due to the noise of the channel.

Since the bits sent by the first player are a deterministic function of the input X and the received transcript W, one can compare these bits to the received transcript W and compute E deterministically, given X and W. Therefore, $E = E(X,W)$. Thus, the first player already knows the value of E.

We "reveal" the value of E to the second player (the first player already knows this value). That is, we condition the game $G_{(Y,X|V_1,B,W)_{T_{V_1}}}$ on the value of E, and consider the game $G_{(Y,X|V_1,B,W)_{T_{V_1}}|E} = G_{(Y,X|V_1,B,W,E)_{T_{V_1}}}$, with parameters $(k, d-1, P_{(Y|V_1,B,W,E)_{T_{V_1}}}, P_{(X|V_1,B,W,E)_{T_{V_1}}})$.

Revealing B'. Fix $v_1 \in [2^k]$, and $w \in \mathrm{supp}(W)$. For every $v_2 \in \mathrm{supp}(P_{V_2|V_1=v_1})$, we define the value $B'(v_2|v_1,w) \in \{0,1\}$ as follows: $B'(v_2|v_1,w) = 1$ if and only if $\Pr_{V_2}[V_2 = v_2|V_1 =$

$v_1, W = w] > 2^{0.01k} \cdot 2^{-k}$. Define $B' = B'(V_2|V_1, W)$.

We "reveal" the value of B' to the first player (the second player already knows this value). That is, we condition the game $G_{(Y,X|V_1,B,W,E)_{T_{V_1}}}$ on the value of B', and consider the game $G_{(Y,X|V_1,B,W,E)_{T_{V_1}}|B'} = G_{(Y,X|V_1,B,W,E,B')_{T_{V_1}}}$, with parameters
$(k, d-1, P_{(Y|V_1,B,W,E,B')_{T_{V_1}}}, P_{(X|V_1,B,W,E,B')_{T_{V_1}}})$.

Revealing the flattening values F_1, F_2. Let $a = 0.01k$, and $r = \frac{20k}{a} = 2000$. Denote $\mu_1 = P_{(Y|V_1,B,W,E,B')_{T_{V_1}}}$, and $\mu_2 = P_{(X|V_1,B,W,E,B')_{T_{V_1}}}$. If $V_1 \in \bar{\mathcal{B}}_1$ and $B \in \bar{\mathcal{B}}_2$, we denote $F_1 = f_{\mu_1,a,r}(Y_{T_{V_1}})$ and $F_2 = f_{\mu_2,a,r}(X_{T_{V_1}})$ (see Subsection 5.1). (If $V_1 \in \mathcal{B}_1$ or $B \in \mathcal{B}_2$, it will be convenient to define $F_1 = F_2 = 0$).

We "reveal" the value of F_1 to the first player (the second player already knows this value), and the value of F_2 to the second player (the first player already knows this value). That is, we condition the game $G_{(Y,X|V_1,B,W,E,B')_{T_{V_1}}}$ on the values of F_1, F_2, and consider the game $G_{(Y,X|V_1,B,W,E,B')_{T_{V_1}}|F_1,F_2} = G_{(Y,X|V_1,B,W,E,B',F_1,F_2)_{T_{V_1}}}$, with parameters
$(k, d-1, P_{(Y|V_1,B,W,E,B',F_1,F_2)_{T_{V_1}}}, P_{(X|V_1,B,W,E,B',F_1,F_2)_{T_{V_1}}})$.

$V_1, B, W, E, B', F_1, F_2$ is a feasible transcript. Note that $V_1, B, W, E, B', F_1, F_2$ is a feasible transcript (see Subsection 6.2). Moreover, these variables form a feasible transcript even when they appear in several different orders. In general, the variables form a feasible transcript if all the following conditions are satisfied: F_1, F_2 appear at he end; B appears after V_1; and B' appears after both V_1, W.

8.1.3 Bounding $\mathsf{CC}(\Pi)$

We next bound $\mathsf{CC}(\Pi)$, and thus complete the proof of Lemma 7.2. To do so, we will use the following three main lemmas. For simplicity of the notation we denote by Z the tuple of random variables $(V_1, B, W, E, B', F_1, F_2)$, and we denote by z a tuple of values $(v_1, b, w, e, b', f_1, f_2) \in \mathrm{supp}(Z)$ that these random variables can take.

Lemma 8.2. *The game $G_{(Y,X|Z)_{T_{V_1}}}$ is nice with probability of at least $1 - 2^{-0.005k}$ (where the probability is over the selection of Z).*

Lemma 8.3.

$$\mathop{\mathbf{E}}_{Z}\left[\kappa\left(G_{(Y,X|Z)_{T_{V_1}}}\right)\right] > 0.75k + 0.25\log(k).$$

Lemma 8.4. *Let $h \in \mathbb{R}^+$. Then,*

$$\Pr_{Z}\left[I\left(G_{(Y,X|Z)_{T_{V_1}}}\right) > 40k + 2h\right] < 2 \cdot 2^{-h}.$$

Lemma 8.2 is proved in Subsection 8.2, Lemma 8.3 is proved in Subsection 8.3, and Lemma 8.4 is proved in Subsection 8.4.

Equipped with the above lemmas, the proof of Lemma 7.2 is as follows. Using Lemmas 6.3 and 6.4 (where we first apply Lemma 6.4 on the transcript W of the first t rounds of the protocol Π for the game G, and then apply Lemma 6.3 for every $w \in \text{supp}(W)$ on the game $G_{XY|W=w}$ and feasible transcript V_1, B, E, B', F_1, F_2), there exists $\{\delta_z\}_{z\in\text{supp}(Z)}$, $\delta_z \in [0,1]$, such that $\mathbf{E}_Z[\delta_Z] = \delta$, and

$$\mathsf{CC}(\Pi) \geq t + \mathop{\mathbf{E}}_{Z}\left[\mathsf{CC}_{\epsilon,\delta_Z}\left(G_{XY|Z}\right)\right].$$

Consider the game $G_{XY|Z}$. Since the first non-root vertex on the correct path, V_1, is already known (as we conditioned on its value), it holds that

$$\mathsf{CC}_{\epsilon,\delta_Z}\left(G_{X,Y|Z}\right) = \mathsf{CC}_{\epsilon,\delta_Z}\left(G_{(Y,X|Z)_{T_{V_1}}}\right).$$

In particular,

$$\mathsf{CC}(\Pi) \geq t + \mathop{\mathbf{E}}_{Z}\left[\mathsf{CC}_{\epsilon,\delta_Z}\left(G_{(Y,X|Z)_{T_{V_1}}}\right)\right].$$

Denote by $\mathcal{A}$ the event that the game $G_{(Y,X|Z)_{T_{V_1}}}$ is nice. If $\mathcal{A}$ occurs, we can apply Lemma 7.2 recursively, as the depth of the new game is $d-1$. If $\bar{\mathcal{A}}$ occurs, we can apply Lemma 7.3, as the depth of the new game is $d-1$. Informally, Lemma 8.2 shows that the probability that Lemma 7.3 is applied is small, and Lemma 8.4 bounds the losses in the bound that occur because of applying Lemma 7.3 rather than Lemma 7.2. Lemma 8.3 ensures that the recursive bound obtained by applying Lemma 7.2 is sufficient. Informally, the term $0.25\log(k)$ in Lemma 8.3 represents the "losses" of the players after communicating the first t bits.

Formally, we get

$$\begin{aligned}
\mathsf{CC}(\Pi) \quad \geq \quad & t + (d-1)\cdot(k + 0.1\log(k))\cdot(1 - 2\mathop{\mathbf{E}}_{Z}[\delta_Z]) \\
& - \Pr_{Z}[\mathcal{A}]\cdot 100k - \Pr_{Z}[\bar{\mathcal{A}}]\cdot 1000k \\
& - \Pr_{Z}\left[\mathcal{A}\right]\left(k - \mathop{\mathbf{E}}_{Z|\mathcal{A}}\left[\kappa(G_{(Y,X|Z)_{T_{V_1}}})\right]\right) \\
& - \Pr_{Z}\left[\bar{\mathcal{A}}\right]\cdot 100\mathop{\mathbf{E}}_{Z|\bar{\mathcal{A}}}\left[I(G_{(Y,X|Z)_{T_{V_1}}})\right].
\end{aligned}$$

By Lemma 8.2, $\Pr_Z[\mathcal{A}] \geq 1 - 2^{-0.005k}$. By Lemma 8.3 it holds that

$$\mathop{\mathbf{E}}_{Z|\mathcal{A}}\left[\kappa(G_{(Y,X|Z)_{T_{V_1}}})\right] \geq (0.75k + 0.25\log(k)) - 2^{-0.005k}k \geq 0.75k + 0.24\log(k).$$

By Lemma 8.4 it holds that

$$\begin{aligned}
&\Pr_Z\left[\bar{\mathcal{A}}\right]\cdot\mathop{\mathbf{E}}_{Z|\bar{\mathcal{A}}}\left[I(G_{(Y,X|Z)_{T_{V_1}}})\right]\leq\\
&\Pr_Z\left[\bar{\mathcal{A}}\right]\left(80k+\sum_{h\in\mathbb{N},h\geq 80k}(h+1)\cdot\Pr_Z\left[I(G_{(Y,X|Z)_{T_{V_1}}})>h\,\middle|\,\bar{\mathcal{A}}\right]\right)\leq\\
&2^{-0.005k}\cdot 80k+\sum_{h\in\mathbb{N},h\geq 80k}(h+1)\cdot\Pr_Z\left[I(G_{(Y,X|Z)_{T_{V_1}}})>h\right]\leq\\
&1+\sum_{h\in\mathbb{N},h\geq 80k}(h+1)\cdot 2\cdot 2^{-0.5(h-40k)}<2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathsf{CC}(\Pi) &\geq (\kappa(G)+0.25k-1)+(d-1)\cdot(k+0.1\log(k))\cdot(1-2\delta)\\
&\quad -100k-2^{-0.005k}1000k\\
&\quad -(0.25k-0.24\log(k))-100\cdot 2\\
&\geq \kappa(G)+d\cdot(k+0.1\log(k))\cdot(1-2\delta)-(k+0.1\log(k))-100k\\
&\quad +0.24\log(k)-202\\
&\geq d\cdot(k+0.1\log(k))\cdot(1-2\delta)-(k-\kappa(G))-100k+0.14\log(k)-202\\
&\geq d\cdot(k+0.1\log(k))\cdot(1-2\delta)-(k-\kappa(G))-100k.
\end{aligned}$$

This concludes the proof of Lemma 7.2.

8.2 Bounding "Bad" Events (Proof of Lemma 8.2)

In this subsection we prove Lemma 8.2. Denote

$$\begin{aligned}
\mathcal{S} &= \left\{(v_1,b,w,e,b',f_1,f_2)\in\mathrm{supp}(V_1,B,W,E,B',F_1,F_2)|v_1\in\bar{\mathcal{B}_1},b\in\bar{\mathcal{B}_2}\right\}\\
&= \mathrm{supp}\left(P_{V_1,B,W,E,B',F_1,F_2|V_1\in\bar{\mathcal{B}_1},B\in\bar{\mathcal{B}_2}}\right).
\end{aligned}$$

Recall that we already defined the two "bad" sets $\mathcal{B}_1$ and $\mathcal{B}_2$. We define the additional "bad" sets $\mathcal{B}_3,\ldots,\mathcal{B}_9$, each is a subset of tuples $(v_1,b,w,e,b',f_1,f_2)\in\mathcal{S}$.

- Denote by $\mathcal{B}_3\subseteq\mathcal{S}$ the set of all tuples $z=(v_1,b,w,e,b',f_1,f_2)$ such that

$$I_1\left(G_{(Y,X|Z=z)_{T_{v_1}}}\right)=\mathbf{I}\left((Y|Z=z)_{T_{v_1}}\right)>10k.$$

- Denote by $\mathcal{B}_4 \subseteq \mathcal{S}$ the set of all tuples $z = (v_1, b, w, e, b', f_1, f_2)$ such that

$$I_2\left(G_{(Y,X|Z=z)_{T_{v_1}}}\right) = \mathbf{I}\left((X|Z=z)_{T_{v_1}}\right) > 20k.$$

- Denote by $\mathcal{B}_5 \subseteq \mathcal{S}$ the set of all tuples $z = (v_1, b, w, e, b', f_1, f_2)$ such that

$$\kappa\left(G_{(Y,X|Z=z)_{T_{v_1}}}\right) = \mathbf{H}_\infty\left(\left((Y|Z=z)_{T_{v_1}}\right)_{v_1}\right) < 0.5k.$$

- Denote by $\mathcal{B}_6 \subseteq \mathcal{S}$ the set of all tuples $z = (v_1, b, w, e, b', f_1, f_2)$ such that $f_1 = r + 1$.
- Denote by $\mathcal{B}_7 \subseteq \mathcal{S}$ the set of all tuples $z = (v_1, b, w, e, b', f_1, f_2)$ such that $f_2 = r + 1$.
- Denote by $\mathcal{B}_8 \subseteq \mathcal{S}$ the set of all tuples $z = (v_1, b, w, e, b', f_1, f_2)$ such that $f_1 = -r$.
- Denote by $\mathcal{B}_9 \subseteq \mathcal{S}$ the set of all tuples $z = (v_1, b, w, e, b', f_1, f_2)$ such that $f_2 = -r$.

Proof of Lemma 8.2. Using Lemma 5.3 and the definition of a nice game, it holds that the game $G_{(Y,X|Z)_{T_{V_1}}}$ is nice unless one of the following "bad" events occurs: $V_1 \in \mathcal{B}_1$, or $B \in \mathcal{B}_2$, or $Z \in \mathcal{B}_i$ for some $i \in \{3, \ldots, 9\}$. (The flatness conditions hold by Lemma 5.3 because $P_{(Y|V_1,B,W,E,B',F_1,F_2)_{T_{V_1}}} = P_{(Y|V_1,B,W,E,B',F_1)_{T_{V_1}}}$ and $P_{(X|V_1,B,W,E,B',F_1,F_2)_{T_{V_1}}} = P_{(X|V_1,B,W,E,B',F_2)_{T_{V_1}}}$).

The assertion follows from the following claims (stated and proved below), as each claim bounds one of these "bad" events. The needed claims are 8.5, 8.6, 8.7, 8.8, 8.9 (part 1), 8.10, 8.13, 8.14, 8.15, and 8.16. □

The rest of this subsection is devoted to proving the claims used by the proof of Lemma 8.2. Each of the claims bounds the probability of obtaining a different set $\mathcal{B}_i$.

Claim 8.5. $\Pr_{V_1}[V_1 \in \mathcal{B}_1] < 2^{-0.2k}$.

Proof. Using Lemma 6.7, and the fact that G is nice, it holds that

$$\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}\left((Y|V_1)_{T_{V_1}}\right)\right] \leq 2^{-\mathbf{H}_\infty(X_0)}\mathbf{I}(Y) \leq 2^{-0.5k} \cdot 20k.$$

The lemma follows by Markov's inequality. □

Claim 8.6. $\Pr_B[B \in \mathcal{B}_2] < 2^{-0.02k}$.

Proof. For $v_1 \in \mathcal{B}_1$ the bit $B(v_2|v_1)$ is never 1. Thus, for $v_1 \in \mathcal{B}_1$,

$$\Pr_B[B \in \mathcal{B}_2 | V_1 = v_1] = 0.$$

For $v_1 \in \bar{\mathcal{B}}_1$, we have $\mathbf{I}\left(Y_{T_{v_1}}\right) = \mathbf{I}\left((Y|V_1 = v_1)_{T_{v_1}}\right) < 2^{-0.25k}$ (see Remark 6.5). Since, conditioned on $V_1 = v_1$, the value of V_2 contains the exact same information as $\left(Y_{T_{v_1}}\right)_{v_1}$, for every $v_1 \in \bar{\mathcal{B}}_1$ we have by the super-additivity of information that $\mathbf{I}\left(V_2|V_1 = v_1\right) \leq \mathbf{I}\left(Y_{T_{v_1}}|V_1 = v_1\right) = \mathbf{I}\left(Y_{T_{v_1}}\right) < 2^{-0.25k}$. Hence by Lemma 5.12, applied for the distribution $\mu = P_{V_2|V_1=v_1}$, for any $v_1 \in \bar{\mathcal{B}}_1$ we have

$$\Pr_B\left[B \in \mathcal{B}_2|V_1 = v_1\right] < 2^{-0.02k}.$$

□

Claim 8.7. $\Pr_Z\left[Z \in \mathcal{B}_3\right] < 2^{-0.02k}$.

Proof. Fix $v_1 \in \bar{\mathcal{B}}_1$. Thus, $\mathbf{I}\left((Y|V_1 = v_1)_{T_{v_1}}\right) < 2^{-0.25k}$. Using Lemma 6.1 we can write

$$P_{(Y|V_1=v_1)_{T_{v_1}}} = \sum_{b,w,e,b',f_1,f_2} \Pr\left[Z = (v_1, b, w, e, b', f_1, f_2)|V_1 = v_1\right] \cdot P_{(Y|Z=(v_1,b,w,e,b',f_1,f_2))_{T_{v_1}}}.$$

Using Lemma 5.13 with $\mu = P_{(Y|V_1=v_1)_{T_{v_1}}}$, $\mu_z = P_{(Y|Z=z)_{T_{v_1}}}$, $h = k$, and $c < 8k$, it holds that

$$\Pr_Z\left[Z \in \mathcal{B}_3|V_1 = v_1\right] = \sum_{z=(v_1,b,w,e,b',f_1,f_2)\in\mathcal{B}_3} \Pr\left[Z = z|V_1 = v_1\right] < 2^{-k} + 5\left(2^{-0.25k}\right)^{0.1} < 2^{-0.02k}.$$

Since this is true for every $v_1 \in \bar{\mathcal{B}}_1$, and since

$$\Pr_Z\left[Z \in \mathcal{B}_3|V_1 \in \mathcal{B}_1\right] = 0,$$

the claim follows. □

Claim 8.8. $\Pr_Z\left[Z \in \mathcal{B}_4\right] < 2^{-k}$.

Proof. Using the supper-additivity of information (Proposition 4.2), for every $z = (v_1, b, w, e, b', f_1, f_2) \in \mathrm{supp}(Z)$, it holds that

$$\mathbf{I}\left((X|Z = z)_{T_{v_1}}\right) \leq \mathbf{I}(X|Z = z).$$

Hence, for every $z \in \mathcal{B}_4$, it holds that $\mathbf{I}(X|Z = z) > 20k$.

Using Lemma 6.1 we can write

$$P_X = \sum_{z\in\mathrm{supp}(Z)} \Pr_Z[Z = z] \cdot P_{X|Z=z}.$$

Since G is nice, P_X is $(0.01k)$-flat and $\mathbf{I}(X) \leq 10k$. Thus, we can use Lemma 5.10 with $\mu = P_X$, $\mu_z = P_{X|Z=z}$, $h = k$, and $c < 8k$, and get

$$\Pr_Z[Z \in \mathcal{B}_4] = \sum_{z \in \mathcal{B}_4} \Pr_Z[Z = z] < 2^{-k}.$$

□

Claim 8.9. *If $t_2 \leq 0.4k$ then*

1. $\Pr_Z[Z \in \mathcal{B}_5] < 2^{-0.05k}$.
2. $\mathbf{E}_Z\left[\kappa\left(G_{(Y,X|Z)_{T_{V_1}}}\right)\right] > k - t_2 - 30$.

Proof. Fix $v_1 \in \bar{\mathcal{B}}_1$ and $b \in \bar{\mathcal{B}}_2$. Consider the random variable $(V_2|V_1 = v_1, B = b)$. Let $v_2 \in \mathrm{supp}(P_{V_2|V_1=v_1,B=b})$. Thus, $B(v_2|v_1) = b$. Hence,

$$\Pr_{V_2}[V_2 = v_2|V_1 = v_1] \leq 2 \cdot 2^{-k}.$$

In addition, in the proof of Claim 8.6, we proved that for every $v_1 \in \bar{\mathcal{B}}_1$ it holds that

$$\Pr[B \neq b|V_1 = v_1] = \Pr[B \in \mathcal{B}_2|V_1 = v_1] < 2^{-0.02k}.$$

For every three events A_1, A_2, A_3, it holds that

$$\Pr[A_1|A_2, A_3] = \frac{\Pr[A_1|A_2] \cdot \Pr[A_3|A_1, A_2]}{\Pr[A_3|A_2]}.$$

Therefore,

$$\begin{aligned}\Pr[V_2 = v_2|V_1 = v_1, B = b] &= \frac{\Pr[V_2 = v_2|V_1 = v_1] \cdot \Pr[B = b|V_2 = v_2, V_1 = v_1]}{\Pr[B = b|V_1 = v_1]} \\ &< \frac{(2 \cdot 2^{-k}) \cdot 1}{(1 - 2^{-0.02k})} < 4 \cdot 2^{-k}.\end{aligned}$$

Hence, $\mathbf{H}_\infty(V_2|V_1 = v_1, B = b) > k - 2$.

Given $V_1 = v_1$, the random variables V_2 and Y_{v_1} contain the same information. Therefore,

$$\begin{aligned}\mathbf{H}_\infty\left(\left((Y|V_1 = v_1, B = b)_{T_{v_1}}\right)_{v_1}\right) &= \mathbf{H}_\infty\left((Y|V_1 = v_1, B = b)_{v_1}\right) \\ &= \mathbf{H}_\infty(V_2|V_1 = v_1, B = b) > k - 2.\end{aligned}$$

Using Lemma 6.2 with the game $G_{(YX|V_1=v_1,B=b)_{T_{v_1}}}$, feasible transcript (W, E, B', F_1, F_2), and

parameters $h = 0.05k$ and $m_1 = t_2 + 1 + \lceil \log(2r+2) \rceil < 0.4k + 20$, it holds that

$$\Pr_Z \left[Z \in \mathcal{B}_5 | V_1 = v_1, B = b \right] = \sum_{z=(v_1,b,w,e,b',f_1,f_2) \in \mathcal{B}_5} \Pr \left[Z = z | V_1 = v_1, B = b \right] < 2^{-0.05k}.$$

Since this is true for every $v_1 \in \bar{\mathcal{B}}_1$ and $b \in \bar{\mathcal{B}}_2$, and since

$$\Pr_Z \left[Z \in \mathcal{B}_5 | (V_1 \in \mathcal{B}_1) \vee (B \in \mathcal{B}_2) \right] = 0,$$

the first part of the claim follows.

Using Lemma 6.2 with the same game, feasible transcript, and parameters as before, we also get that for every $v_1 \in \bar{\mathcal{B}}_1$ and $b \in \bar{\mathcal{B}}_2$

$$\begin{aligned} &\mathop{\mathbf{E}}_{Z|V_1=v_1,B=b} \left[\mathbf{H}_\infty \left(\left((Y|Z)_{T_{v_1}} \right)_{v_1} \right) \right] = \\ &\mathop{\mathbf{E}}_{W,E,B',F_1,F_2|V_1=v_1,B=b} \left[\mathbf{H}_\infty \left(\left((Y|V_1 = v_1, B = b, W, E, B', F_1, F_2)_{T_{v_1}} \right)_{v_1} \right) \right] > \\ &(k-2) - (t_2 + 20) - 4 = k - t_2 - 26. \end{aligned}$$

Since $\Pr \left[V_1 \in \bar{\mathcal{B}}_1, B \in \bar{\mathcal{B}}_2 \right] > 1 - 2^{-0.2k} - 2^{-0.02k}$, it holds that

$$\mathop{\mathbf{E}}_Z \left[\kappa \left(G_{(Y,X|Z)_{T_{V_1}}} \right) \right] = \mathop{\mathbf{E}}_Z \left[\mathbf{H}_\infty \left(\left((Y|Z)_{T_{V_1}} \right)_{V_1} \right) \right] > k - t_2 - 27.$$

□

Claim 8.10. *If $t_2 > 0.4k$ then $\Pr_Z \left[Z \in \mathcal{B}_5 \right] < 2^{-0.01k}$.*

Proof. Since $t_2 > 0.4k$, it holds that $t_1 = t - t_2 < \kappa(G) - 0.15k$. Denote by $\mathcal{B}'_1 \subseteq \mathrm{supp}(V_1, W)$ the set of all pairs (v_1, w) such that $\mathbf{I} \left((Y|V_1 = v_1, W = w)_{T_{v_1}} \right) \geq 2^{-0.05k}$.

Claim 8.11. $\Pr_{V_1,W} \left[(V_1, W) \in \mathcal{B}'_1 \right] < 2^{-0.03k}$.

Proof. Let $\mathcal{B} \subseteq \{0,1\}^t$ be the set of strings w such that $\mathbf{H}_\infty(X_0|W = w) < 0.1k$. Using Lemma 6.2, it holds that $\Pr_W \left[W \in \mathcal{B} \right] < 2^{-0.05k}$.

Let $\mathcal{B}' \subseteq \{0,1\}^t$ be the set of strings w such that $\mathbf{I}(Y|W = w) > 25k$. Since G is nice it holds that $\mathbf{I}(Y) \leq 20k$ and that P_Y is $(0.01k)$-flat. Using Lemma 5.10, it holds that $\Pr_W \left[W \in \mathcal{B}' \right] < 2^{-k}$.

Fix $w \in \bar{\mathcal{B}} \cap \bar{\mathcal{B}}'$. Using Lemma 6.7, it holds that

$$\mathop{\mathbf{E}}_{V_1|W=w} \left[\mathbf{I} \left((Y|V_1, W = w)_{T_{V_1}} \right) \right] \leq 2^{-\mathbf{H}_\infty(X_0|W=w)} \mathbf{I}(Y|W = w) \leq 2^{-0.1k} \cdot 25k.$$

By Markov's inequality, for every $w \in \bar{\mathcal{B}} \cap \bar{\mathcal{B}}'$, it holds that $\Pr_{V_1}[(V_1, W) \in \mathcal{B}'_1 | W = w] < 2^{-0.04k}$. Hence $\Pr_{V_1,W}[(V_1, W) \in \mathcal{B}'_1] < 2^{-0.04k} + 2^{-0.05k} + 2^{-k} < 2^{-0.03k}$. □

Denote $\mathcal{B}'_2 = \{1\}$. Denote $\bar{\mathcal{B}}'_2 = \{0, 1\} \setminus \mathcal{B}'_2 = \{0\}$.

Claim 8.12. $\Pr_{B'}[B' \in \mathcal{B}'_2] < 2^{-0.011k}$.

Proof. For $(v_1, w) \in \bar{\mathcal{B}}'_1$, we have $\mathbf{I}\left((Y|V_1 = v_1, W = w)_{T_{v_1}}\right) < 2^{-0.05k}$. Since, conditioned on $V_1 = v_1$, the value of V_2 contains the exact same information as $\left(Y_{T_{v_1}}\right)_{v_1}$, for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ we have by the super-additivity of information that $\mathbf{I}(V_2|V_1 = v_1, W = w) \leq \mathbf{I}\left(Y_{T_{v_1}}|V_1 = v_1, W = w\right) < 2^{-0.05k}$. Hence by Lemma 5.12, applied for the distribution $\mu = P_{V_2|V_1=v_1,W=w}$, for any $(v_1, w) \in \bar{\mathcal{B}}'_1$ we have

$$\Pr_{B'}[B' \in \mathcal{B}'_2 | V_1 = v_1, W = w] < 2^{-0.012k}.$$

Using Claim 8.11 it holds that $\Pr_{B'}[B' \in \mathcal{B}'_2] < 2^{-0.012k} + 2^{-0.03k} < 2^{-0.011k}$. □

We continue with the proof of Claim 8.10. Fix $(v_1, w) \in \bar{\mathcal{B}}'_1$ and $b' \in \bar{\mathcal{B}}'_2$. Consider the random variable $(V_2|V_1 = v_1, W = w, B' = b')$. Let $v_2 \in \mathrm{supp}(P_{V_2|V_1=v_1,W=w,B'=b'})$. Thus, $B'(v_2|v_1, w) = b'$. Hence,

$$\Pr_{V_2}[V_2 = v_2 | V_1 = v_1, W = w] \leq 2^{0.01k} \cdot 2^{-k}.$$

In addition, in the proof of Claim 8.12, we proved that for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ it holds that

$$\Pr[B' \neq b' | V_1 = v_1, W = w] = \Pr[B' \in \mathcal{B}'_2 | V_1 = v_1, W = w] < 2^{-0.012k}.$$

For every three events A_1, A_2, A_3, it holds that

$$\Pr[A_1|A_2, A_3] = \frac{\Pr[A_1|A_2] \cdot \Pr[A_3|A_1, A_2]}{\Pr[A_3|A_2]}.$$

Therefore,

$$\begin{aligned}
&\Pr[V_2 = v_2 | V_1 = v_1, W = w, B' = b'] = \\
&\frac{\Pr[V_2 = v_2 | V_1 = v_1, W = w] \cdot \Pr[B' = b' | V_2 = v_2, V_1 = v_1, W = w]}{\Pr[B' = b' | V_1 = v_1, W = w]} < \\
&\frac{(2^{0.01k} \cdot 2^{-k}) \cdot 1}{(1 - 2^{-0.012k})} < 2^{0.02k} \cdot 2^{-k}.
\end{aligned}$$

Hence, $\mathbf{H}_\infty(V_2|V_1 = v_1, W = w, B' = b') > 0.98k$.

Given $V_1 = v_1$, the random variables V_2 and Y_{v_1} contain the same information. Therefore,

$$\begin{aligned}\mathbf{H}_\infty\left(\left((Y|V_1=v_1,W=w,B'=b')_{T_{v_1}}\right)_{v_1}\right) &= \mathbf{H}_\infty\left((Y|V_1=v_1,W=w,B'=b')_{v_1}\right)\\ &= \mathbf{H}_\infty\left(V_2|V_1=v_1,W=w,B'=b'\right) > 0.98k.\end{aligned}$$

Using Lemma 6.2 with the game $G_{(YX|V_1=v_1,W=w,B'=b')_{T_{v_1}}}$, feasible transcript (B,E,F_1,F_2), and parameters $h = 0.4k$ and $m_1 = 1 + \lceil \log(2r+2) \rceil < 20$, it holds that

$$\Pr_Z\left[Z \in \mathcal{B}_5 | V_1 = v_1, W = w, B' = b'\right] = \sum_{z=(v_1,b,w,e,b',f_1,f_2)\in\mathcal{B}_5} \Pr\left[Z = z | V_1 = v_1, W = w, B' = b'\right] < 2^{-0.4k}.$$

Since this is true for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ and $b' \in \bar{\mathcal{B}}'_2$, using Claims 8.11 and 8.12 it holds that $\Pr_Z\left[Z \in \mathcal{B}_5\right] < 2^{-0.4k} + 2^{-0.03k} + 2^{-0.011k} < 2^{-0.01k}$. □

Claim 8.13. $\mathcal{B}_6 \subseteq \mathcal{B}_3$.

Proof. Let $z = (v_1, b, w, e, b', f_1, f_2) \in \mathcal{B}_6$. We will show $z \in \mathcal{B}_3$. Consider the distribution $\mu = P_{(Y|V_1=v_1,B=b,W=w,E=e,B'=b')_{T_{v_1}}}$. Recall that we denote $a = 0.01k$, and $r = \frac{20k}{a} = 2000$, and that since $f_1 = r + 1 \neq 0$ it holds that $f_1 = f_{\mu,a,r}(Y_{T_{v_1}})$.

Using Lemma 5.4, and since F_2 is independent of Y given (V_1, B, W, E, B', F_1), it holds that

$$\mathbf{I}\left(P_{(Y|Z=z)_{T_{v_1}}}\right) = \mathbf{I}\left(P_{(Y|V_1=v_1,B=b,W=w,E=e,B'=b',F_1=f_1)_{T_{v_1}}}\right) = \mathbf{I}\left(\mu|_{F_1=r+1}\right) > ar = 20k.$$

Therefore, $z \in \mathcal{B}_3$, and the assertion follows. □

Claim 8.14. $\mathcal{B}_7 \subseteq \mathcal{B}_4$.

Proof. Similar to the proof of Claim 8.13. □

Claim 8.15. $\Pr_Z\left[Z \in \mathcal{B}_8\right] < 2^{-k}$.

Proof. Let $z = (v_1, b, w, e, b', f_1, f_2) \in \mathcal{B}_8$. Consider the distribution $\mu = P_{(Y|V_1=v_1,B=b,W=w,E=e,B'=b')_{T_{v_1}}}$. Recall that we denote $a = 0.01k$, and $r = \frac{20k}{a} = 2000$, and that since $f_1 = -r \neq 0$ it holds that $f_1 = f_{\mu,a,r}(Y_{T_{v_1}})$. Using Lemma 5.5 applied with μ, it holds that

$$\Pr_Z[Z = z] \leq \Pr_Z\left[F_1 = -r | V_1 = v_1, B = b, W = w, E = e, B' = b'\right] < 2^{-ar} = 2^{-20k}.$$

Thus, the assertion follows using the union bound as $|\mathcal{B}_8| \leq |\mathrm{supp}(Z)| \leq 2^{3k}$. □

Claim 8.16. $\Pr_Z\left[Z \in \mathcal{B}_9\right] < 2^{-k}$.

Proof. Similar to the proof of Claim 8.15. □

8.3 Bounding the Expected κ of the Obtained Game (Proof of Lemma 8.3)

In this subsection we prove Lemma 8.3.

Proof of Lemma 8.3. We consider different ranges of the parameter t_1. If $t_1 \geq \kappa(G) + 0.5\log(k)$ (and hence $t_2 \leq 0.25k - 0.5\log(k)$), then the assertion follows from Claim 8.9. If $t_1 \leq \kappa(G) - 100\log(k)$, then the assertion follows from Claim 8.17 (below).

We consider the case where $\kappa(G) - 100\log(k) < t_1 < \kappa(G) + 0.5\log(k)$ (and hence $t_2 < 0.25k + 100\log(k)$). By Claim 8.20 (below), it holds that

$$\mathop{\mathbf{E}}_{Z|E=0}\left[\kappa\left(G_{(Y,X|Z,E=0)_{T_{V_1}}}\right)\right] > k - t_2 - 30 > k - (0.25k + 100\log(k)) - 30 > 0.75k - 101\log(k).$$

By Claim 8.21 (below), it holds that

$$\mathop{\mathbf{E}}_{Z|E=1}\left[\kappa\left(G_{(Y,X|Z,E=1)_{T_{V_1}}}\right)\right] > 0.9k.$$

The event $E = 1$ occurs with probability

$$t_1 \cdot \epsilon \cdot (1-\epsilon)^{t_1 - 1} > 0.95 t_1 \epsilon \geq 0.95 \cdot 0.45k \cdot \tfrac{2000\log(k)}{k^2} > 700\tfrac{\log(k)}{k}.$$

Therefore,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{Z}\left[\kappa\left(G_{(Y,X|Z)_{T_{V_1}}}\right)\right] > \\
&\left(1 - \tfrac{700\log(k)}{k}\right)(0.75k - 101\log(k)) + \tfrac{700\log(k)}{k} \cdot 0.9k > \\
&0.75k - 101\log(k) - 525\log(k) + 630\log(k) = 0.75k + 4\log(k).
\end{aligned}$$

□

The rest of this subsection is devoted to proving the claims used by the proof of Lemma 8.3.

Claim 8.17. *If* $t_1 \leq \kappa(G) - 100\log(k)$ *then* $\mathbf{E}_Z\left[\kappa\left(G_{(Y,X|Z)_{T_{V_1}}}\right)\right] > 0.95k$.

Proof. Denote by $\mathcal{B}'_1 \subseteq \mathrm{supp}(V_1, W)$ the set of all pairs (v_1, w) such that $\mathbf{I}\left((Y|V_1 = v_1, W = w)_{T_{v_1}}\right) \geq k^{-50}$.

Claim 8.18. $\Pr_{V_1,W}\left[(V_1, W) \in \mathcal{B}'_1\right] < k^{-22}$.

Proof. Let $\mathcal{B} \subseteq \{0,1\}^t$ be the set of strings w such that $\mathbf{H}_\infty(X_0|W = w) < 75\log k$. Using Lemma 6.2, it holds that $\Pr_W[W \in \mathcal{B}] < k^{-25}$.

Let $\mathcal{B}' \subseteq \{0,1\}^t$ be the set of strings w such that $\mathbf{I}(Y|W=w) > 25k$. Since G is nice it holds that $\mathbf{I}(Y) \leq 20k$ and that P_Y is $(0.01k)$-flat. Using Lemma 5.10, it holds that $\Pr_W[W \in \mathcal{B}'] < 2^{-k}$.

Fix $w \in \bar{\mathcal{B}} \cap \bar{\mathcal{B}}'$. Using Lemma 6.7, it holds that

$$\mathop{\mathbf{E}}_{V_1|W=w}\left[\mathbf{I}\left((Y|V_1, W=w)_{T_{V_1}}\right)\right] \leq 2^{-\mathbf{H}_\infty(X_0|W=w)}\mathbf{I}(Y|W=w) \leq k^{-75} \cdot 25k.$$

By Markov's inequality, for every $w \in \bar{\mathcal{B}} \cap \bar{\mathcal{B}}'$, it holds that $\Pr_{V_1}[(V_1, W) \in \mathcal{B}'_1 | W=w] < k^{-23}$. Hence $\Pr_{V_1,W}[(V_1, W) \in \mathcal{B}'_1] < k^{-23} + k^{-25} + 2^{-k} < k^{-22}$. □

Denote $\mathcal{B}'_2 = \{1\}$. Denote $\bar{\mathcal{B}}'_2 = \{0,1\} \setminus \mathcal{B}'_2 = \{0\}$.

Claim 8.19. $\Pr_{B'}[B' \in \mathcal{B}'_2] < k^{-11}$.

Proof. For $(v_1, w) \in \bar{\mathcal{B}}'_1$, we have $\mathbf{I}\left((Y|V_1=v_1, W=w)_{T_{v_1}}\right) < k^{-50}$. Since, conditioned on $V_1 = v_1$, the value of V_2 contains the exact same information as $\left(Y_{T_{v_1}}\right)_{v_1}$, for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ we have by the super-additivity of information that $\mathbf{I}(V_2|V_1=v_1, W=w) \leq \mathbf{I}\left(Y_{T_{v_1}}|V_1=v_1, W=w\right) < k^{-50}$. Hence by Lemma 5.12, applied for the distribution $\mu = P_{V_2|V_1=v_1,W=w}$, for any $(v_1, w) \in \bar{\mathcal{B}}'_1$ we have

$$\Pr_{B'}[B' \in \mathcal{B}'_2 | V_1=v_1, W=w] < k^{-12}.$$

Using Claim 8.18 it holds that $\Pr_{B'}[B' \in \mathcal{B}'_2] < k^{-12} + k^{-22} < k^{-11}$. □

We continue with the proof of Claim 8.17. Fix $(v_1, w) \in \bar{\mathcal{B}}'_1$ and $b' \in \bar{\mathcal{B}}'_2$. Consider the random variable $(V_2|V_1=v_1, W=w, B'=b')$. Let $v_2 \in \mathrm{supp}(P_{V_2|V_1=v_1,W=w,B'=b'})$. Thus, $B'(v_2|v_1, w) = b'$. Hence,

$$\Pr_{V_2}[V_2 = v_2 | V_1=v_1, W=w] \leq 2^{0.01k} \cdot 2^{-k}.$$

In addition, in the proof of Claim 8.19, we proved that for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ it holds that

$$\Pr[B' \neq b' | V_1=v_1, W=w] = \Pr[B' \in \mathcal{B}'_2 | V_1=v_1, W=w] < k^{-12}.$$

For every three events A_1, A_2, A_3, it holds that

$$\Pr[A_1|A_2, A_3] = \frac{\Pr[A_1|A_2] \cdot \Pr[A_3|A_1, A_2]}{\Pr[A_3|A_2]}.$$

Therefore,

$$\begin{aligned}&\Pr[V_2 = v_2 | V_1 = v_1, W = w, B' = b'] = \\ &\frac{\Pr[V_2 = v_2 | V_1 = v_1, W = w] \cdot \Pr[B' = b' | V_2 = v_2, V_1 = v_1, W = w]}{\Pr[B' = b' | V_1 = v_1, W = w]} < \\ &\frac{(2^{0.01k} \cdot 2^{-k}) \cdot 1}{(1 - k^{-12})} < 2^{0.02k} \cdot 2^{-k}.\end{aligned}$$

Hence, $\mathbf{H}_\infty (V_2 | V_1 = v_1, W = w, B' = b') > 0.98k$.

Given $V_1 = v_1$, the random variables V_2 and Y_{v_1} contain the same information. Therefore,

$$\begin{aligned}\mathbf{H}_\infty \left(\left((Y | V_1 = v_1, W = w, B' = b')_{T_{v_1}} \right)_{v_1} \right) &= \mathbf{H}_\infty \left((Y | V_1 = v_1, W = w, B' = b')_{v_1} \right) \\ &= \mathbf{H}_\infty (V_2 | V_1 = v_1, W = w, B' = b') > 0.98k.\end{aligned}$$

Using Lemma 6.2 with the game $G_{(YX|V_1=v_1,W=w,B'=b')_{T_{v_1}}}$, feasible transcript (B, E, F_1, F_2), and parameter $m_1 = 1 + \lceil \log(2r+2) \rceil < 20$, it holds that for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ and $b' \in \bar{\mathcal{B}}'_2$

$$\begin{aligned}&\mathop{\mathbf{E}}_{Z|V_1=v_1,W=w,B'=b'} \left[\mathbf{H}_\infty \left(\left((Y|Z)_{T_{v_1}} \right)_{v_1} \right) \right] = \\ &\mathop{\mathbf{E}}_{B,E,F_1,F_2|V_1=v_1,W=w,B'=b'} \left[\mathbf{H}_\infty \left(\left((Y | V_1 = v_1, W = w, B' = b', B, E, F_1, F_2)_{T_{v_1}} \right)_{v_1} \right) \right] > \\ &0.98k - 20 - 4 > 0.97k.\end{aligned}$$

Since (using Claims 8.18 and 8.19) $\Pr\left[(V_1, W) \in \bar{\mathcal{B}}'_1, B' \in \bar{\mathcal{B}}'_2 \right] > 1 - k^{-11} - k^{-22}$, it holds that

$$\mathop{\mathbf{E}}_{Z} \left[\kappa \left(G_{(Y,X|Z)_{T_{V_1}}} \right) \right] = \mathop{\mathbf{E}}_{Z} \left[\mathbf{H}_\infty \left(\left((Y|Z)_{T_{V_1}} \right)_{V_1} \right) \right] > 0.95k.$$

□

Claim 8.20.

$$\mathop{\mathbf{E}}_{Z|E=0} \left[\kappa \left(G_{(Y,X|Z,E=0)_{T_{V_1}}} \right) \right] = \mathop{\mathbf{E}}_{V_1,B,W,B',F_1,F_2|E=0} \left[\kappa \left(G_{(Y,X|V_1,B,W,E=0,B',F_1,F_2)_{T_{V_1}}} \right) \right] > k - t_2 - 30.$$

Proof. Fix $v_1 \in \bar{\mathcal{B}}_1$ and $b \in \bar{\mathcal{B}}_2$. In the proof of Claim 8.9, we proved

$$\mathbf{H}_\infty \left(\left((Y | V_1 = v_1, B = b)_{T_{v_1}} \right)_{v_1} \right) > k - 2.$$

Since E is independent of (Y, V_1, B), we can condition on the event $E = 0$ and get

$$\mathbf{H}_\infty \left(\left((Y | V_1 = v_1, B = b, E = 0)_{T_{v_1}} \right)_{v_1} \right) > k - 2.$$

Using Lemma 6.2 with the game $G_{(YX|V_1=v_1,B=b,E=0)_{T_{v_1}}}$, feasible transcript (W, B', F_1, F_2), and parameter $m_1 = t_2 + 1 + \lceil \log(2r+2) \rceil < t_2 + 20$, it holds that for every $v_1 \in \bar{\mathcal{B}}_1$ and $b \in \bar{\mathcal{B}}_2$

$$\underset{Z|V_1=v_1,B=b,E=0}{\mathbf{E}} \left[\mathbf{H}_\infty \left(\left((Y|Z)_{T_{v_1}} \right)_{v_1} \right) \right] =$$
$$\underset{W,B',F_1,F_2|V_1=v_1,B=b,E=0}{\mathbf{E}} \left[\mathbf{H}_\infty \left(\left((Y|V_1=v_1,B=b,E=0,W,B',F_1,F_2)_{T_{v_1}} \right)_{v_1} \right) \right] >$$
$$(k-2) - (t_2+20) - 4 = k - t_2 - 26.$$

Since (using Claims 8.5 and 8.6)

$$\Pr\left[V_1 \in \bar{\mathcal{B}}_1, B \in \bar{\mathcal{B}}_2 | E=0\right] = \Pr\left[V_1 \in \bar{\mathcal{B}}_1, B \in \bar{\mathcal{B}}_2\right] > 1 - 2^{-0.2k} - 2^{-0.02k},$$

it holds that

$$\underset{Z|E=0}{\mathbf{E}} \left[\kappa \left(G_{(Y,X|Z,E=0)_{T_{V_1}}} \right) \right] = \underset{Z|E=0}{\mathbf{E}} \left[\mathbf{H}_\infty \left(\left((Y|Z,E=0)_{T_{V_1}} \right)_{V_1} \right) \right] > k - t_2 - 27.$$

□

Claim 8.21. *If* $\kappa(G) - 100\log(k) < t_1 < \kappa(G) + 0.5\log(k)$ *then*

$$\underset{Z|E=1}{\mathbf{E}} \left[\kappa \left(G_{(Y,X|Z,E=1)_{T_{V_1}}} \right) \right] = \underset{V_1,B,W,B',F_1,F_2|E=1}{\mathbf{E}} \left[\kappa \left(G_{(Y,X|V_1,B,W,E=1,B',F_1,F_2)_{T_{V_1}}} \right) \right] > 0.9k.$$

Proof. Let $\mathcal{B} \subseteq \{0,1\}^t$ be the set of strings w such that $\mathbf{H}_\infty(X_0|W=w,E=1) < 0.49\log(k)$.

Claim 8.22. $\Pr_W[W \in \mathcal{B} | E=1] < 2^{-8}$.

Proof. We need to prove

$$\Pr_W \left[\mathbf{H}_\infty(X_0|W,E=1) < 0.49\log(k) \mid E=1 \right] < 2^{-8}.$$

Let $w \in \{0,1\}^t$. Since (W,E) is a feasible transcript, it holds that conditioned on the event $W=w, E=1$, the distribution of the inputs X, Y remains a product distribution, and in particular, the variables X_0 and Y are independent. Therefore, for every $y \in \mathrm{supp}(P_{Y|W=w,E=1})$

$$\mathbf{H}_\infty(X_0|W=w,E=1) = \mathbf{H}_\infty(X_0|W=w,E=1,Y=y).$$

Thus, it suffices to prove

$$\Pr_{W,Y} \left[\mathbf{H}_\infty(X_0|W,E=1,Y) < 0.49\log(k) \mid E=1 \right] < 2^{-8}.$$

Denote by $\mathcal{A}$ the set of indices $i \in [t]$, such that the i^{th} bit in the protocol was sent by the first player. Note that $|\mathcal{A}| = t_1$. Denote $W_{\mathcal{A}} = \{W_i\}_{i\in\mathcal{A}}$ and $W_{\bar{\mathcal{A}}} = \{W_i\}_{i\in\bar{\mathcal{A}}}$. Denote by $N_{\bar{\mathcal{A}}}$ the noise vector of the channel in locations $\bar{\mathcal{A}}$. That is, $N_{\bar{\mathcal{A}}} = \{N_i\}_{i\in\bar{\mathcal{A}}}$, where $N_i \in \{0,1\}$ is 1 if and only if the i^{th} bit sent by the players was received incorrectly (due to the noise in the channel).

Note that $W_{\bar{\mathcal{A}}} \oplus N_{\bar{\mathcal{A}}}$ are the bits sent by the second player. Therefore, $W_{\bar{\mathcal{A}}} \oplus N_{\bar{\mathcal{A}}}$ is a deterministic function of Y and $W_{\mathcal{A}}$. Hence, conditioned on Y and $W_{\mathcal{A}}$, we have that $W_{\bar{\mathcal{A}}}$ uniquely determines $N_{\bar{\mathcal{A}}}$ and vice versa. Hence, we can replace the conditioning on $W_{\bar{\mathcal{A}}}$ by conditioning on $N_{\bar{\mathcal{A}}}$, as follows:

$$\Pr_{W,Y}\left[\mathbf{H}_\infty(X_0|W, E=1, Y) < 0.49\log(k) \mid E=1\right] =$$

$$\Pr_{W_{\mathcal{A}},W_{\bar{\mathcal{A}}},Y}\left[\mathbf{H}_\infty(X_0|W_{\mathcal{A}}, W_{\bar{\mathcal{A}}}, E=1, Y) < 0.49\log(k) \mid E=1\right] =$$

$$\Pr_{W_{\mathcal{A}},N_{\bar{\mathcal{A}}},Y}\left[\mathbf{H}_\infty(X_0|W_{\mathcal{A}}, N_{\bar{\mathcal{A}}}, E=1, Y) < 0.49\log(k) \mid E=1\right].$$

We will prove that for every $y \in \mathrm{supp}(Y)$ and $n_{\bar{\mathcal{A}}} \in \mathrm{supp}(N_{\bar{\mathcal{A}}})$,

$$\Pr_{W_{\mathcal{A}}}\left[\mathbf{H}_\infty(X_0|W_{\mathcal{A}}, N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E=1, Y=y) < 0.49\log(k) \mid N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E=1, Y=y\right]$$

$$< 2^{-8}. \tag{7}$$

Hence the claim follows. In the rest of the proof we prove Equation 7.

Fix $y \in \mathrm{supp}(Y)$ and $n_{\bar{\mathcal{A}}} \in \mathrm{supp}(N_{\bar{\mathcal{A}}})$. Since $X_0, Y, E, N_{\bar{\mathcal{A}}}$ are independent random variables,

$$\mathbf{H}_\infty(X_0|N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E=1, Y=y) = \mathbf{H}_\infty(X_0) = \kappa(G). \tag{8}$$

Assume that $E=1$. Denote by $L \in \mathcal{A}$ the location of the single noise applied to the bits sent by the first player. We will now argue that conditioned on the event $E=1$, for every fixed values of $N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, Y=y, X=x$ there are exactly t_1 possibilities for $W_{\mathcal{A}}$, each obtained with equal probability. The t_1 possibilities for $W_{\mathcal{A}}$ correspond to the t_1 possibilities for L. This is true because (assuming that $E=1$) W is determined by $X, Y, N_{\bar{\mathcal{A}}}, L$, and since two different possibilities $\ell < \ell'$ for L result in two different values of W (and hence of $W_{\mathcal{A}}$) that differ in coordinate $\ell \in \mathcal{A}$. Therefore, for every $x_0 \in \mathrm{supp}(X_0)$,

$$\begin{aligned}
&\mathbf{H}_\infty(W_{\mathcal{A}}|X_0 = x_0, N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E=1, Y=y) \geq \\
&\min_{x\in\mathrm{supp}(X|X_0=x_0)}\{\mathbf{H}_\infty(W_{\mathcal{A}}|X_0 = x_0, N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E=1, Y=y, X=x)\} \geq \\
&\min_{x\in\mathrm{supp}(X)}\{\mathbf{H}_\infty(W_{\mathcal{A}}|N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E=1, Y=y, X=x)\} = \log(t_1).
\end{aligned}$$

By the last equation and Equation 8 and since $t_1 > \kappa(G) - 100\log(k) > 0.25k$, for every $y \in \mathrm{supp}(Y)$ and $n_{\bar{\mathcal{A}}} \in \mathrm{supp}(N_{\bar{\mathcal{A}}})$,

$$\mathbf{H}_\infty((X_0, W_{\mathcal{A}})|N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E = 1, Y = y) \geq$$

$$\kappa(G) + \log(t_1) \geq \kappa(G) + \log(0.25k) = \kappa(G) + \log(k) - 2. \tag{9}$$

Fix $y \in \mathrm{supp}(Y)$ and $n_{\bar{\mathcal{A}}} \in \mathrm{supp}(N_{\bar{\mathcal{A}}})$. Let $\mathcal{B}' = \mathcal{B}'(n_{\bar{\mathcal{A}}}, y) \subseteq \{0,1\}^{\mathcal{A}}$ be the set of strings $w_{\mathcal{A}}$ such that

$$\mathbf{H}_\infty(X_0|W_{\mathcal{A}} = w_{\mathcal{A}}, N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E = 1, Y = y) < 0.5\log(k) - 10.$$

Denote

$$\beta = \Pr_{W_{\mathcal{A}}}[W_{\mathcal{A}} \in \mathcal{B}'|N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E = 1, Y = y].$$

Thus, to prove Equation 7 (and hence to prove the claim), it suffices to show that

$$\beta < 2^{-8}.$$

Since $|\mathcal{B}'| \leq 2^{t_1}$, assuming that $\mathcal{B}'$ is not empty, there exists $w_{\mathcal{A}} \in \mathcal{B}'$ such that

$$\Pr_{W_{\mathcal{A}}}[W_{\mathcal{A}} = w_{\mathcal{A}}|N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E = 1, Y = y] \geq \beta \cdot 2^{-t_1} \geq \beta \cdot 2^{-\kappa(G)-0.5\log(k)}.$$

Since $w_{\mathcal{A}} \in \mathcal{B}'$, there exists x_0, such that,

$$\Pr_{X_0}(X_0 = x_0|W_{\mathcal{A}} = w_{\mathcal{A}}, N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E = 1, Y = y) > 2^{-0.5\log(k)+10}.$$

Thus,

$$\Pr_{X_0, W_{\mathcal{A}}}((X_0 = x_0, W_{\mathcal{A}} = w_{\mathcal{A}})|N_{\bar{\mathcal{A}}} = n_{\bar{\mathcal{A}}}, E = 1, Y = y) > \beta \cdot 2^{-\kappa(G)-\log(k)+10}.$$

Hence, by Equation 9,

$$\beta \cdot 2^{-\kappa(G)-\log(k)+10} < 2^{-\kappa(G)-\log(k)+2}.$$

That is, $\beta < 2^{-8}$, and the assertion follows. □

Denote by $\mathcal{B}'_1 \subseteq \mathrm{supp}(P_{V_1,W|E=1})$ the set of all pairs (v_1, w) such that

$$\mathbf{I}\left((Y|V_1 = v_1, W = w, E = 1)_{T_{v_1}}\right) \geq k^{0.75}.$$

Claim 8.23. $\Pr_{V_1,W}[(V_1,W) \in \mathcal{B}'_1 | E=1] < 2^{-7}$.

Proof. Recall that we denote by $\mathcal{B} \subseteq \{0,1\}^t$ the set of strings w such that $\mathbf{H}_\infty(X_0|W=w,E=1) < 0.49\log(k)$. By Claim 8.22, $\Pr_W[W \in \mathcal{B}|E=1] < 2^{-8}$.

Let $\mathcal{B}' \subseteq \{0,1\}^t$ be the set of strings w such that $\mathbf{I}(Y|W=w,E=1) > 25k$. Since G is nice and since Y,E are independent, it holds that $\mathbf{I}(Y|E=1) = \mathbf{I}(Y) \leq 20k$ and that $P_{Y|E=1} = P_Y$ is $(0.01k)$-flat. Using Lemma 5.10, it holds that $\Pr_W[W \in \mathcal{B}'|E=1] < 2^{-k}$.

Fix $w \in \bar{\mathcal{B}} \cap \bar{\mathcal{B}}'$. Using Lemma 6.7, it holds that

$$\mathop{\mathbf{E}}_{V_1|W=w,E=1}\left[\mathbf{I}\left((Y|V_1,W=w,E=1)_{T_{V_1}}\right)\right] \leq 2^{-\mathbf{H}_\infty(X_0|W=w,E=1)}\mathbf{I}(Y|W=w,E=1)$$

$$\leq k^{-0.49} \cdot 25k < k^{0.52}.$$

By Markov's inequality, for every $w \in \bar{\mathcal{B}} \cap \bar{\mathcal{B}}'$, it holds that

$$\Pr_{V_1}[(V_1,W) \in \mathcal{B}'_1 | W=w,E=1] < k^{-0.23}.$$

Hence $\Pr_{V_1,W}[(V_1,W) \in \mathcal{B}'_1|E=1] < k^{-0.23} + 2^{-8} + 2^{-k} < 2^{-7}$. □

Denote $\mathcal{B}'_2 = \{1\}$. Denote $\bar{\mathcal{B}}'_2 = \{0,1\} \setminus \mathcal{B}'_2 = \{0\}$.

Claim 8.24. $\Pr_{B'}[B' \in \mathcal{B}'_2|E=1] < 2^{-6}$.

Proof. For $(v_1,w) \in \bar{\mathcal{B}}'_1$, we have $\mathbf{I}\left((Y|V_1=v_1,W=w,E=1)_{T_{v_1}}\right) < k^{0.75}$. Since, conditioned on $V_1=v_1$, the value of V_2 contains the exact same information as $\left(Y_{T_{v_1}}\right)_{v_1}$, for every $(v_1,w) \in \bar{\mathcal{B}}'_1$ we have by the super-additivity of information that

$$\mathbf{I}(V_2|V_1=v_1,W=w,E=1) \leq \mathbf{I}\left(Y_{T_{v_1}}|V_1=v_1,W=w,E=1\right) < k^{0.75}.$$

Recall that E is a deterministic function of W and X. Therefore (V_1,W,E) is a feasible transcript, where, conditioned on V_1,W, the variable E depends only on X and is independent of Y. Therefore $P_{V_2|V_1=v_1,W=w,E=1} = P_{V_2|V_1=v_1,W=w}$. In particular,

$$\mathbf{I}(V_2|V_1=v_1,W=w) < k^{0.75}.$$

Hence by Lemma 5.6, part 2, applied for the distribution $\mu = P_{V_2|V_1=v_1,W=w}$, with parameters $a=1$, $r=k$, and $m=10k^{0.24}$, for any $(v_1,w) \in \bar{\mathcal{B}}'_1$ we have

$$\Pr_{B'}[B' \in \mathcal{B}'_2|V_1=v_1,W=w,E=1] = \Pr_{B'}[B' \in \mathcal{B}'_2|V_1=v_1,W=w] < k^{-0.24}.$$

Using Claim 8.23 it holds that $\Pr_{B'}[B' \in \mathcal{B}'_2 | E=1] < k^{-0.24} + 2^{-7} < 2^{-6}$. $\square$

We continue with the proof of Claim 8.21. Fix $(v_1, w) \in \bar{\mathcal{B}}'_1$ and $b' \in \bar{\mathcal{B}}'_2$. Consider the random variable $(V_2 | V_1 = v_1, W = w, E = 1, B' = b')$. Let $v_2 \in \mathrm{supp}(P_{V_2|V_1=v_1,W=w,E=1,B'=b'})$. Thus, $B'(v_2|v_1, w) = b'$. Hence,

$$\Pr_{V_2}[V_2 = v_2 | V_1 = v_1, W = w, E = 1] = \Pr_{V_2}[V_2 = v_2 | V_1 = v_1, W = w] \leq 2^{0.01k} \cdot 2^{-k}.$$

In addition, in the proof of Claim 8.24, we proved that for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ it holds that

$$\Pr[B' \neq b' | V_1 = v_1, W = w, E = 1] = \Pr[B' \in \mathcal{B}'_2 | V_1 = v_1, W = w, E = 1] < k^{-0.24}.$$

For every three events A_1, A_2, A_3, it holds that

$$\Pr[A_1 | A_2, A_3] = \frac{\Pr[A_1 | A_2] \cdot \Pr[A_3 | A_1, A_2]}{\Pr[A_3 | A_2]}.$$

Therefore,

$$\begin{aligned}
&\Pr[V_2 = v_2 | V_1 = v_1, W = w, E = 1, B' = b'] = \\
&\frac{\Pr[V_2 = v_2 | V_1 = v_1, W = w, E = 1] \cdot \Pr[B' = b' | V_2 = v_2, V_1 = v_1, W = w, E = 1]}{\Pr[B' = b' | V_1 = v_1, W = w, E = 1]} < \\
&\frac{(2^{0.01k} \cdot 2^{-k}) \cdot 1}{(1 - k^{-0.24})} < 2^{0.02k} \cdot 2^{-k}.
\end{aligned}$$

Hence, $\mathbf{H}_\infty(V_2 | V_1 = v_1, W = w, E = 1, B' = b') > 0.98k$.

Given $V_1 = v_1$, the random variables V_2 and Y_{v_1} contain the same information. Therefore,

$$\begin{aligned}
&\mathbf{H}_\infty\left(\left((Y | V_1 = v_1, W = w, E = 1, B' = b')_{T_{v_1}}\right)_{v_1}\right) = \\
&\mathbf{H}_\infty\left((Y | V_1 = v_1, W = w, E = 1, B' = b')_{v_1}\right) = \\
&\mathbf{H}_\infty(V_2 | V_1 = v_1, W = w, E = 1, B' = b') > 0.98k.
\end{aligned}$$

Using Lemma 6.2 with the game $G_{(YX|V_1=v_1,W=w,E=1,B'=b')_{T_{v_1}}}$, feasible transcript (B, F_1, F_2), and parameter $m_1 = 1 + \lceil \log(2r+2) \rceil < 20$, it holds that for every $(v_1, w) \in \bar{\mathcal{B}}'_1$ and $b' \in \bar{\mathcal{B}}'_2$

$$\begin{aligned}
&\mathop{\mathbf{E}}_{Z|V_1=v_1,W=w,E=1,B'=b'}\left[\mathbf{H}_\infty\left(\left((Y|Z)_{T_{v_1}}\right)_{v_1}\right)\right] = \\
&\mathop{\mathbf{E}}_{B,F_1,F_2|V_1=v_1,W=w,E=1,B'=b'}\left[\mathbf{H}_\infty\left(\left((Y|V_1 = v_1, W = w, E = 1, B' = b', B, F_1, F_2)_{T_{v_1}}\right)_{v_1}\right)\right] > \\
&0.98k - 20 - 4 > 0.97k.
\end{aligned}$$

Since (using Claims 8.23 and 8.24) $\Pr\left[(V_1,W)\in\bar{\mathcal{B}}'_1, B'\in\bar{\mathcal{B}}'_2|E=1\right] > 1-2^{-7}-2^{-6}$, it holds that

$$\mathop{\mathbf{E}}_{Z|E=1}\left[\kappa\left(G_{(Y,X|Z,E=1)_{T_{V_1}}}\right)\right] = \mathop{\mathbf{E}}_{Z|E=1}\left[\mathbf{H}_\infty\left(\left((Y|Z,E=1)_{T_{V_1}}\right)_{V_1}\right)\right] > 0.9k.$$

□

8.4 Bounding the Information of the Obtained Game (Proof of Lemma 8.4)

In this subsection we prove Lemma 8.4.

Proof of Lemma 8.4. We first show that

$$\Pr_Z\left[I_2\left(G_{(Y,X|Z)_{T_{V_1}}}\right) > 15k+h\right] < 2^{-h}.$$

Using the supper-additivity of information (Proposition 4.2), for every $z\in\mathrm{supp}(Z)$ and $v_1\in\mathrm{supp}(V_1)$, it holds that

$$\mathbf{I}\left((X|Z=z)_{T_{v_1}}\right) \le \mathbf{I}(X|Z=z).$$

Using Lemma 6.1 we can write

$$P_X = \sum_{z\in\mathrm{supp}(Z)} \Pr_Z[Z=z]\cdot P_{X|Z=z}.$$

Since G is nice, P_X is $(0.01k)$-flat and $\mathbf{I}(X)\le 10k$. Thus, we can use Lemma 5.10 with $\mu=P_X$, $\mu_z=P_{X|Z=z}$, and $c<3k$, and get

$$\Pr_Z\left[I_2\left(G_{(Y,X|Z)_{T_{V_1}}}\right) > 15k+h\right] = \Pr_Z\left[\mathbf{I}\left((X|Z)_{T_{V_1}}\right) > 15k+h\right] \le$$
$$\Pr_Z\left[\mathbf{I}(X|Z) > 10k+3k+0.01k+h\right] < 2^{-h}.$$

Using the same argument, it also holds that

$$\Pr_Z\left[I_1\left(G_{(Y,X|Z)_{T_{V_1}}}\right) > 25k+h\right] < 2^{-h},$$

and the assertion follows. □

9 Communication Lower-Bound for Non-Nice Games (Proof of Lemma 7.3)

Let G^* be a (possibly not nice) game with parameters (k, d, P_{X^*}, P_{Y^*}), and underlying tree T^*. Let ϵ and δ be as specified by Lemma 7.3. We assume that $d \geq 100$ and $\delta \leq 0.5$, as otherwise the lemma holds trivially. Our goal is to bound $\mathsf{CC}_{\epsilon,\delta}(G^*)$ in the other cases.

We first "reveal" the value of V_1^*, the first non-root vertex on the correct path defined by the inputs X^*, Y^*, to the second player (the first player already knows this value). That is, we condition the game G^* on the value of V_1^*. We then reduce the game to the subtree $T^*_{V_1^*}$. That is, we consider the game $G_{(Y^*,X^*|V_1^*)_{T^*_{V_1^*}}} = G^*_{(Y^*,X^*|V_1^*)_{T^*_{V_1^*}}}$.

Let $v_1^* \in \mathrm{supp}(V_1^*)$. Denote by $G_{v_1^*}$ the game $G_{(Y^*,X^*|V_1^*=v_1^*)_{T^*_{v_1^*}}}$. Using Lemma 6.3, there exists $\left\{\delta_{v_1^*}\right\}_{v_1^* \in \mathrm{supp}(V_1^*)}$, $\delta_{v_1^*} \in [0,1]$, such that $\mathbf{E}_{V_1^*}[\delta_{V_1^*}] = \delta$, and

$$\mathsf{CC}_{\epsilon,\delta}(G^*) \geq \mathop{\mathbf{E}}_{V_1^*}\left[\mathsf{CC}_{\epsilon,\delta_{V_1^*}}\left(G_{X^*Y^*|V_1^*}\right)\right].$$

Consider the game $G_{X^*Y^*|V_1^*}$. Since the first non-root vertex on the correct path, V_1^*, is already known (as we conditioned on its value), it holds that

$$\mathsf{CC}_{\epsilon,\delta_{V_1^*}}\left(G_{X^*Y^*|V_1^*}\right) = \mathsf{CC}_{\epsilon,\delta_{V_1^*}}\left(G_{(Y^*,X^*|V_1^*)_{T^*_{V_1^*}}}\right) = \mathsf{CC}_{\epsilon,\delta_{V_1^*}}\left(G_{V_1^*}\right).$$

Hence,

$$\mathsf{CC}_{\epsilon,\delta}(G^*) \geq \mathop{\mathbf{E}}_{V_1^*}\left[\mathsf{CC}_{\epsilon,\delta_{V_1^*}}\left(G_{V_1^*}\right)\right].$$

Using Lemma 6.7, it holds that

$$\mathop{\mathbf{E}}_{V_1^*}\left[I\left(G_{V_1^*}\right)\right] \leq I(G^*).$$

Therefore, to prove the lemma, it suffices to show that for a fixed $v_1^* \in \mathrm{supp}(V_1^*)$, the game $G = G_{v_1^*}$ satisfies

$$\mathsf{CC}_{\epsilon,\delta}(G) \geq d \cdot (k + 0.1\log(k)) \cdot (1 - 2\delta) - 100I(G) - 1000k, \tag{10}$$

for any $\delta := \delta_{v_1^*} \in [0,1]$.

Fix $v_1^* \in \mathrm{supp}(V_1^*)$, and let $G = G_{v_1^*}$ and $\delta = \delta_{v_1^*}$. Denote the input variables of G by X and Y, and denote the underlying tree by T. Observe that G is of depth $d-1$. In the rest of the proof we prove Equation 10. We use our standard notation for the game G, and forget

about the game G^* altogether (the reason that we switched from G^* to G is that the depth of G is smaller, so we can use induction in Case 4 below). We consider the following cases.

Case 1: $\mathbf{I}(X_0) > 0.1k$. We "reveal" the value of V_1 to the second player, and consider the game $G_{(Y,X|V_1)_{T_{V_1}}}$. Using Lemma 6.7, it holds that

$$\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((X|V_1)_{T_{V_1}})\right] \leq \mathbf{I}(X) - \mathbf{I}(X_0) < \mathbf{I}(X) - 0.1k.$$

In addition, since V_1 gives no information about Y and by the super-additivity of information (Proposition 4.2),

$$\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((Y|V_1)_{T_{V_1}})\right] \leq \mathbf{I}(Y).$$

Therefore,

$$\mathop{\mathbf{E}}_{V_1}\left[I\left(G_{(Y,X|V_1)_{T_{V_1}}}\right)\right] \leq I(G) - 0.1k.$$

As above, using Lemma 6.3, there exists $\{\delta_{v_1}\}_{v_1 \in \mathrm{supp}(V_1)}$, $\delta_{v_1} \in [0,1]$, such that $\mathbf{E}_{V_1}[\delta_{V_1}] = \delta$, and

$$\mathsf{CC}_{\epsilon,\delta}(G) \geq \mathop{\mathbf{E}}_{V_1}\left[\mathsf{CC}_{\epsilon,\delta_{V_1}}\left(G_{(Y,X|V_1)_{T_{V_1}}}\right)\right].$$

Since the game $G_{(Y,X|V_1)_{T_{V_1}}}$ is of depth $d-2$, we can recursively apply Lemma 7.3 and get

$$\begin{aligned}
\mathsf{CC}_{\epsilon,\delta}(G) &\geq (d-2)\cdot(k+0.1\log(k))\cdot(1-2\mathop{\mathbf{E}}_{V_1}[\delta_{V_1}]) - 100\mathop{\mathbf{E}}_{V_1}\left[I\left(G_{(Y,X|V_1)_{T_{V_1}}}\right)\right] - 1000k \\
&\geq (d-2)\cdot(k+0.1\log(k))\cdot(1-2\delta) - 100(I(G)-0.1k) - 1000k \\
&\geq d\cdot(k+0.1\log(k))\cdot(1-2\delta) - 100I(G) - 1000k.
\end{aligned}$$

Case 2: $I_2(G) = \mathbf{I}(Y) > 0.2k$. It suffices to consider the case where $\mathbf{I}(X_0) \leq 0.1k$. Since $\mathbf{I}(X_0) \leq 0.1k$, it holds that $\mathbf{H}_\infty(X_0) \geq 1$. As in Case 1, in this case we also "reveal" the value of V_1 to the second player, and consider the game $G_{(Y,X|V_1)_{T_{V_1}}}$. Using Lemma 6.7, it holds that

$$\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((Y|V_1)_{T_{V_1}})\right] \leq 0.5\cdot\mathbf{I}(Y) \leq \mathbf{I}(Y) - 0.1k.$$

By Lemma 6.7, it also holds that

$$\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((X|V_1)_{T_{V_1}})\right] \leq \mathbf{I}(X).$$

Therefore,

$$\mathop{\mathbf{E}}_{V_1}\left[I\left(G_{(Y,X|V_1)_{T_{V_1}}}\right)\right] \leq I(G) - 0.1k.$$

The assertion follows as in Case 1.

Case 3: $I_1(G) = \mathbf{I}(X) > 0.2k$**.** It suffices to consider the case where $I_2(G) \leq 0.2k$. Since $I_2(G) \leq 0.2k$, for every $v_1 \in \text{supp}(V_1)$ it holds that $\mathbf{H}_\infty(V_2|V_1 = v_1) \geq 1$. In this case we "reveal" the value of V_1 to the second player, and "reveal" the value of V_2 to the first player. That is, we consider the game $G_{\left((X,Y|V_1)_{T_{V_1}}|V_2\right)_{T_{V_2}}} = G_{(X,Y|V_1,V_2)_{T_{V_2}}}$.

Using Lemma 6.7, it holds that

$$\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((Y|V_1)_{T_{V_1}})\right] \leq \mathbf{I}(Y).$$

$$\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((X|V_1)_{T_{V_1}})\right] \leq \mathbf{I}(X).$$

By applying Lemma 6.7 again on the game $G_{(Y,X|V_1)_{T_{V_1}}}$, it holds that

$$\begin{aligned}\mathop{\mathbf{E}}_{V_1}\mathop{\mathbf{E}}_{V_2}\left[\mathbf{I}\left(\left((X|V_1)_{T_{V_1}}|V_2\right)_{T_{V_2}}\right)\right] &\leq 2^{-\min_{v_1\in\text{supp}(V_1)}\{\mathbf{H}_\infty(V_2|V_1=v_1)\}}\cdot\mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((X|V_1)_{T_{V_1}})\right]\\ &\leq 0.5\cdot\mathbf{I}(X) \leq \mathbf{I}(X) - 0.1k.\end{aligned}$$

and

$$\mathop{\mathbf{E}}_{V_1}\mathop{\mathbf{E}}_{V_2}\left[\mathbf{I}\left(\left((Y|V_1)_{T_{V_1}}|V_2\right)_{T_{V_2}}\right)\right] \leq \mathop{\mathbf{E}}_{V_1}\left[\mathbf{I}((Y|V_1)_{T_{V_1}})\right] \leq \mathbf{I}(Y).$$

Therefore,

$$\mathop{\mathbf{E}}_{V_1,V_2}\left[I\left(G_{\left((X,Y|V_1)_{T_{V_1}}|V_2\right)_{T_{V_2}}}\right)\right] \leq I(G) - 0.1k.$$

The assertion follows as in Case 1.

Case 4: The above cases are not satisfied. We consider the case where $\mathbf{I}(X_0) \leq 0.1k$, and $\mathbf{I}(X), \mathbf{I}(Y) \leq 0.2k$. Let $a = 0.01k$, $r = 2000$, and $r_0 = 45$. Define the flattening values $F_1 = f_{P_X,a,r}(X)$, $F_2 = f_{P_Y,a,r}(Y)$, and $F_0 = f_{P_{(X_0|F_1,F_2)},a,r_0}(X_0)$. We "reveal" the value of F_1 to the second player (the first player already knows this value), and the value of F_2 to the first player (the second player already knows this value). We then "reveal" the value of F_0 to the second player (the first player already knows this value). That is, we condition the game G on the value of F, where $F = (F_1, F_2, F_0)$, and consider the game $G_{X,Y|F}$. Observe that $\mathbf{H}(F) < 100$, as F can be represented using $3\log(2r+2) < 100$ bits.

To complete the proof of lemma 7.3, we will use the following lemma that is proved in Subsection 9.1.

Lemma 9.1. *The game $G_{X,Y|F}$ is nice with probability at least 0.5 (over the selection of F).*

Equipped with Lemma 9.1, we can complete the proof of Lemma 7.3 as follows. As

before, by Lemma 6.3, there exists $\{\delta_f\}_{f\in\mathrm{supp}(F)}$, $\delta_f \in [0,1]$, such that $\mathbf{E}_F[\delta_F] = \delta$, and

$$\mathsf{CC}_{\epsilon,\delta}(G) \geq \mathop{\mathbf{E}}_{F}\left[\mathsf{CC}_{\epsilon,\delta_F}\left(G_{X,Y|F}\right)\right].$$

Consider the game $G_{X,Y|F}$. Denote by $\mathcal{A}$ the event that the game $G_{X,Y|F}$ is nice. If $\mathcal{A}$ occurs, we can apply Lemma 7.2 recursively, as the depth of the game is $d-1$. If $\bar{\mathcal{A}}$ occurs, we can apply Lemma 7.3, as the depth of the game is $d-1$. We get

$$\begin{aligned}\mathsf{CC}_{\epsilon,\delta}(G) \;\geq\; & (d-1)\cdot(k+0.1\log(k))\cdot(1-2\mathop{\mathbf{E}}_{F}[\delta_F]) - \Pr_{F}[\mathcal{A}]\cdot 100k - \Pr_{F}[\bar{\mathcal{A}}]\cdot 1000k \\ & - \Pr_{F}[\mathcal{A}]\left(k - \mathop{\mathbf{E}}_{F|\mathcal{A}}\left[\kappa(G_{X,Y|F})\right]\right) - \Pr_{F}\left[\bar{\mathcal{A}}\right]\cdot 100\mathop{\mathbf{E}}_{F|\bar{\mathcal{A}}}\left[I(G_{X,Y|F})\right].\end{aligned}$$

By Lemma 9.1, $\Pr_F[\mathcal{A}] \geq 0.5$. By the chain rule for the entropy function, it holds that

$$\begin{aligned}&\Pr_{F}\left[\bar{\mathcal{A}}\right]\mathop{\mathbf{E}}_{F|\bar{\mathcal{A}}}\left[I(G_{X,Y|F})\right] \leq \mathop{\mathbf{E}}_{F}\left[I(G_{X,Y|F})\right] = \mathop{\mathbf{E}}_{F}\left[\mathbf{I}(X|F)\right] + \mathop{\mathbf{E}}_{F}\left[\mathbf{I}(Y|F)\right] \\ &\leq (\mathbf{I}(X)+\mathbf{H}(F)) + (\mathbf{I}(Y)+\mathbf{H}(F)) \leq I(G)+200.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathsf{CC}_{\epsilon,\delta}(G) \;&\geq\; (d-1)\cdot(k+0.1\log(k))\cdot(1-2\delta) - 550k - k - 100(I(G)+200) \\ &\geq\; d\cdot(k+0.1\log(k))\cdot(1-2\delta) - 100I(G) - 1000k.\end{aligned}$$

This concludes the proof of Lemma 7.3.

9.1 Bounding "Bad" Events (Proof of Lemma 9.1)

In this subsection we prove Lemma 9.1. We define the following "bad" sets $\mathcal{B}_1,\ldots,\mathcal{B}_7$, each is a subsets of $\mathrm{supp}(F)$.

- Denote by $\mathcal{B}_1 \subseteq \mathrm{supp}(F)$ the set of all elements f such that

$$I_1\left(G_{X,Y|F=f}\right) = \mathbf{I}(X|F=f) > 10k.$$

- Denote by $\mathcal{B}_2 \subseteq \mathrm{supp}(F)$ the set of all elements f such that

$$I_2\left(G_{X,Y|F=f}\right) = \mathbf{I}(Y|F=f) > 20k.$$

- Denote by $\mathcal{B}_3 \subseteq \text{supp}(F)$ the set of all elements f such that

$$\kappa\left(G_{X,Y|F=f}\right) = \mathbf{H}_\infty(X_0|F=f) < 0.5k.$$

- Denote by $\mathcal{B}_4 \subseteq \text{supp}(F)$ the set of all elements $f = (f_1, f_2, f_0)$ such that $f_1 = r+1$.
- Denote by $\mathcal{B}_5 \subseteq \text{supp}(F)$ the set of all elements $f = (f_1, f_2, f_0)$ such that $f_2 = r+1$.
- Denote by $\mathcal{B}_6 \subseteq \text{supp}(F)$ the set of all elements $f = (f_1, f_2, f_0)$ such that $f_1 = -r$.
- Denote by $\mathcal{B}_7 \subseteq \text{supp}(F)$ the set of all elements $f = (f_1, f_2, f_0)$ such that $f_2 = -r$.

Proof of Lemma 9.1. We first claim that if $f = (f_1, f_2, f_0) \in \bar{\mathcal{B}}_4 \cap \bar{\mathcal{B}}_6$, then the distribution $P_{X|F=f}$ is $(0.01k)$-flat. The reason is the following. Using Lemma 5.3, the distribution $P_{X|F_1=f_1}$ is $(0.01k)$-flat. That is, for $x, x' \in \text{supp}(P_{X|F_1=f_1})$ it holds that

$$\frac{P_{X|F_1=f_1}(x)}{P_{X|F_1=f_1}(x')} \leq 2^{0.01k}.$$

Denote $\mathcal{S} = \text{supp}\left(P_{X|F=f}\right)$, and assume $\mathcal{S} \neq \phi$. Observe that $P_{X|F=f} = P_{X|F_1=f_1}|_{\mathcal{S}}$, that is, $P_{X|F=f}$ is the distribution $P_{X|F_1=f_1}$ restricted to the set $\mathcal{S}$. Therefore, for $x, x' \in \mathcal{S}$ it holds that

$$\frac{P_{X|F=f}(x)}{P_{X|F=f}(x')} = \frac{\left(\frac{P_{X|F_1=f_1}(x)}{P_{X|F_1=f_1}(\mathcal{S})}\right)}{\left(\frac{P_{X|F_1=f_1}(x')}{P_{X|F_1=f_1}(\mathcal{S})}\right)} = \frac{P_{X|F_1=f_1}(x)}{P_{X|F_1=f_1}(x')} \leq 2^{0.01k}.$$

Hence, $P_{X|F=f}$ is $(0.01k)$-flat.

Similarly, by Lemma 5.3, if $f \in \bar{\mathcal{B}}_5 \cap \bar{\mathcal{B}}_7$ then $P_{Y|F=f}$ is $(0.01k)$-flat.

The game $G_{X,Y|F}$ is nice unless one of the "bad" events $F \in \mathcal{B}_i$, for some $i \in \{1, \ldots, 7\}$, occurs. The assertion follows from the following claims (stated and proved below), as each claim bounds one of these "bad" events. The needed claims are 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, and 9.8. □

The rest of this subsection is devoted to proving the claims used by the proof of Lemma 9.1. Each of the claims bounds the probability of obtaining a different set $\mathcal{B}_i$.

Claim 9.2. $\Pr_F[F \in \mathcal{B}_1] < 0.025$.

Proof. By the chain rule for the entropy function, it holds that

$$\mathop{\mathbf{E}}_F[\mathbf{I}(X|F)] \leq \mathbf{I}(X) + \mathbf{H}(F) \leq 0.2k + 100 < 0.25k.$$

Using Markov's inequity, it holds that

$$\Pr_F[\mathbf{I}(X|F) > 10k] < 0.025.$$

□

Claim 9.3. $\Pr_F[F \in \mathcal{B}_2] < 0.025$.

Proof. Similar to the proof of Claim 9.2. □

Claim 9.4. $\Pr_F[F \in \mathcal{B}_3] < 0.25$.

Proof. By the chain rule for the entropy function, it holds that

$$\mathop{\mathbf{E}}_F[\mathbf{I}(X_0|F)] \leq \mathbf{I}(X_0) + \mathbf{H}(F) \leq 0.1k + 100 < 0.101k.$$

Using Markov's inequity, it holds that

$$\Pr_F[\mathbf{I}(X_0|F) > 0.45k] < 0.23.$$

Let $f = (f_1, f_2, -r_0) \in \mathrm{supp}(F)$. By Lemma 5.5 applied with $\mu = P_{X_0|F_1=f_1,F_2=f_2}$, it holds that

$$\Pr_F[F = f] \leq \Pr[F_0 = -r_0 | F_1 = f_1, F_2 = f_2] \leq 2^{-a \cdot r_0} = 2^{-0.45k}.$$

Using the union bound, and since $|\mathrm{supp}(F)| \leq (2r+2)^3 = 4002^3$, it holds that $\Pr_F[F_0 = -r_0] \leq 2^{-0.4k}$.

Denote by $\mathcal{B} \subseteq \mathrm{supp}(F)$ the set of all elements $f = (f_1, f_2, f_0)$ such that $f_0 = -r_0$ or $\mathbf{H}(X_0|F = f) < 0.55k$ (that is, $\mathbf{I}(X_0|F = f) > 0.45k$). Thus,

$$\Pr_F[F \in \mathcal{B}] < 0.23 + 2^{-0.4k} < 0.25.$$

Let $f = (f_1, f_2, f_0) \in \bar{\mathcal{B}}$. By Lemma 5.4 applied with $\mu = P_{X_0|F_1=f_1,F_2=f_2}$, if $f_0 = r_0 + 1$, then $\mathbf{I}(X_0|F = f) > a \cdot r_0 = 0.45k$. But, this is impossible as $f \in \bar{\mathcal{B}}$. Thus, $-r_0 + 1 \leq f_0 \leq r_0$. By Lemma 5.3, the distribution $P_{X_0|F=f}$ is $(0.01k)$-flat. Using Proposition 5.2, for very $f \in \bar{\mathcal{B}}$ it holds that

$$\mathbf{H}_\infty(P_{X_0|F=f}) \geq \mathbf{H}(P_{X_0|F=f}) - 0.01k \geq 0.55k - 0.01k = 0.54k,$$

and the assertion follows. □

Claim 9.5. $\Pr_F[F \in \mathcal{B}_4] < 0.025$.

Proof. By the chain rule for the entropy function, it holds that

$$\mathop{\mathbf{E}}_{F_1}[\mathbf{I}(X|F_1)] \leq \mathbf{I}(X) + \mathbf{H}(F_1) \leq 0.2k + 100 < 0.25k.$$

Using Markov's inequity, it holds that

$$\Pr_{F_1}[\mathbf{I}(X|F_1) > 10k] < 0.025.$$

By Lemma 5.4 applied with the distribution P_X, it holds that $\mathbf{I}\left(P_{X|F_1=r+1}\right) > ar = 20k$. Therefore, $\Pr_F[F \in \mathcal{B}_4] = \Pr_{F_1}[F_1 = r+1] < 0.025$. □

Claim 9.6. $\Pr_F[F \in \mathcal{B}_5] < 0.025$.

Proof. Similar to the proof of Claim 9.5. □

Claim 9.7. $\Pr_F[F \in \mathcal{B}_6] < 2^{-k}$.

Proof. By Lemma 5.5 applied with the distribution P_X, it holds that

$$\Pr_F[F \in \mathcal{B}_6] = \Pr_{F_1}[F_1 = -r] \leq 2^{-ar} = 2^{-20k}.$$

□

Claim 9.8. $\Pr_F[F \in \mathcal{B}_7] < 2^{-k}$.

Proof. Similar to the proof of Claim 9.7. □

References

[1] Zvika Brakerski and Yael Tauman Kalai. Efficient interactive coding against adversarial noise. In *FOCS*, pages 160–166, 2012.

[2] Zvika Brakerski and Moni Naor. Fast algorithms for interactive coding. In *SODA*, pages 443–456, 2013.

[3] Mark Braverman. Towards deterministic tree code constructions. In *ITCS*, pages 161–167, 2012.

[4] Mark Braverman and Anup Rao. Towards coding for maximum errors in interactive communication. In *STOC*, pages 159–166, 2011.

[5] Ran Gelles, Ankur Moitra, and Amit Sahai. Efficient and explicit coding for interactive communication. In *FOCS*, pages 768–777, 2011.

[6] Leonard J. Schulman. Communication on noisy channels: A coding theorem for computation. In *FOCS*, pages 724–733, 1992.

[7] Leonard J. Schulman. Deterministic coding for interactive communication. In *STOC*, pages 747–756, 1993.

[8] Leonard J. Schulman. Coding for interactive communication. *IEEE Transactions on Information Theory*, 42(6):1745–1756, 1996.

[9] C. E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:July 379–423, October 623–656, 1948.