Electronic Colloquium on Computational Complexity, Report No. 118 (2022)

# Strong XOR Lemma for Communication with Bounded Rounds

Huacheng Yu*

**Abstract**

In this paper, we prove a strong XOR lemma for bounded-round two-player randomized communication. For a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, the $n$-fold XOR function $f^{\oplus n} : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ maps $n$ input pairs $(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ to the XOR of the $n$ output bits $f(X_1, Y_1) \oplus \cdots \oplus f(X_n, Y_n)$. We prove that if every $r$-round communication protocols that computes $f$ with probability $2/3$ uses at least $C$ bits of communication, then any $r$-round protocol that computes $f^{\oplus n}$ with probability $1/2 + \exp(-O(n))$ must use $n \cdot \left(r^{-O(r)} \cdot C - 1\right)$ bits. When $r$ is a constant and $C$ is sufficiently large, this is $\Omega(n \cdot C)$ bits. It matches the communication cost and the success probability of the trivial protocol that computes the $n$ bits $f(X_i, Y_i)$ independently and outputs their XOR, up to a constant factor in $n$.

A similar XOR lemma has been proved for $f$ whose communication lower bound can be obtained via bounding the discrepancy [Sha03]. By the equivalence between the discrepancy and the correlation with 2-bit communication protocols [VW08], our new XOR lemma implies the previous result.

*Department of Computer Science, Princeton University. yuhch123@gmail.com. Supported by Simons Junior Faculty Award.

ISSN 1433-8092

# Contents

# 1 Introduction

In computational complexity, XOR lemmas study the relation between the complexity of a $\{0,1\}$-valued function $f(x)$ and the complexity of the $n$-fold XOR function $f^{\oplus n}$ where

$$f^{\oplus n}(x_1,\ldots,x_n) = f(x_1) \oplus \cdots \oplus f(x_n)$$

and $\oplus$ is the XOR. A classic example is Yao's XOR lemma for circuits [Yao82], which states if $f$ cannot be computed with probability $2/3$ on a random input by size-$s$ circuits, then $f^{\oplus n}$ cannot be computed with probability $1/2 + \exp(-\Omega(n))$ on a random input by size-$s'$ circuits for some $s' < s$ (and small $n$). Such lemmas can be used to create very hard functions in a blackbox way, which can only be computed barely better than random guessing, from functions that are "just" hard to compute with constant probability. This approach of hardness amplification has been used in one-way functions [Yao82, Lev87], pseudorandom generators [Imp95, IW97], and more recently, streaming lower bounds [AN21, CKP+21].

In general, suppose computing a function $f$ with probability $2/3$ requires resource $s$ in some model of computation (e.g., circuit size, running time, query complexity, communication cost, etc). Then the trivial way to compute $f^{\oplus n}$ is to compute each $f(x_i)$ using resource $s$ independently, and output their XOR. It uses resource $n \cdot s$ in total, and each instance is correct with probability $2/3$, hence, their XOR is correct with probability $1/2 + \exp(-\Theta(n))$: For two *independent* random bits $b_1, b_2$, if $\Pr[b_1 = 0] = 1/2 + \alpha_1/2$ and $\Pr[b_2 = 0] = 1/2 + \alpha_2/2$, then

$$\Pr[b_1 \oplus b_2 = 0] = (1/2 + \alpha_1/2)(1/2 + \alpha_2/2) + (1/2 - \alpha_1/2)(1/2 - \alpha_2/2) = 1/2 + \alpha_1\alpha_2/2;$$

let $b_i = 0$ if and only if $f(x_i)$ is computed correctly, applying the above calculation inductively gives the claimed probability. A *strong XOR lemma* asserts that to achieve $1/2 + \exp(-O(n))$ success probability, one must use $\Omega(n \cdot s)$ resource — the trivial solution is essentially optimal.

Now suppose that we are given $n \cdot s/2$ resource in total, and we want to compute $f^{\oplus n}$. If we try to solve the $n$ copies independently, then no matter how we distribute the resource among the $n$ copies, at least half of them will get no more than $s$. The function $f^{\oplus}$ is computed correctly with probability at most $1/2 + \exp(-\Omega(n))$. Of course, since all $n$ inputs $(x_1,\ldots,x_n)$ are given together, we can potentially process them jointly. This may correlate the $n$ copies, and in particular, it may correlate the correctness of computing each $f(x_i)$. Hence, one difficulty in proving the strong XOR lemma from the technical point of view is that in the above calculation of the probability of XOR of two independent bits, the linear terms perfectly cancel only because $b_1$ and $b_2$ are independent; when they are not independent, we may get a linear term remaining, and do not reduce the probability bias as desired. In computational models where one cannot expect the independence between the copies throughout the computation, a success probability lower bound of $1/2 + \exp(-\Omega(n))$ (hence, a strong XOR lemma) is generally difficult to prove.

In this paper, we prove a strong XOR lemma for the two-player randomized communication complexity with bounded rounds: Alice and Bob receive $X$ and $Y$ respectively, they alternatively send a total of $r$ messages to each other with the goal of computing $f(X,Y)$. For $f^{\oplus n}$, Alice receives $(X_1,\ldots,X_n)$ and Bob receives $(Y_1,\ldots,Y_n)$, and they wish to compute $f(X_1,Y_1) \oplus \cdots \oplus f(X_n,Y_n)$ after $r$ rounds of communication. Each player has half of the inputs for all copies, and can send messages that arbitrarily depend on them, which can nontrivially correlate the $n$ instances. Nevertheless, we show that one cannot do much better than simply solving all $n$ copies in parallel.

Let $\mathbf{R}_p^{(r)}(f)$ be the minimum number of bits of communication needed in $r$ messages in order to compute $f(X,Y)$ correctly with probability $p$. We prove the following theorem.

**Theorem 1.** *For any $\{0,1\}$-valued function $f$, we have*

$$\mathbf{R}^{(r)}_{1/2+2^{-n}}(f^{\oplus n}) \geq n \cdot \left(r^{-O(r)} \cdot \mathbf{R}^{(r)}_{2/3}(f) - 1\right).$$

In particular, when $r$ is a constant, it implies that $\mathbf{R}^{(r)}_{1/2+2^{-n}}(f^{\oplus n}) \geq \Omega\left(n \cdot \left(\mathbf{R}^{(r)}_{2/3}(f) - O(1)\right)\right)$.[1] To the best of our knowledge, such an XOR lemma was not known even for one-way communication and without the factor of $n$.

As pointed in [BBCR13], the "$-O(1)$" term is needed. This is because for $f(X,Y) = X \oplus Y$, we have $\mathbf{R}^{(r)}_{2/3}(f) = 2$. On the other hand, $f^{\oplus n}$ can also be computed with 2 bits of communication by simply (locally) computing $\bigoplus_{i=1}^n X_i$ and $\bigoplus_{i=1}^n Y_i$ and exchanging the values.

We obtain Theorem 1 via the following *distributional* strong XOR lemma. Let $\mathrm{suc}_\mu(f; C_A, C_B, r)$ be the maximum success probability of an $r$-round protocol $\pi$ computing $f(X,Y)$ where

- Alice sends at most $C_A$ bits in every odd round,
- Bob sends at most $C_B$ bits in every even round, and
- $(X,Y)$ is sampled from $\mu$.

**Theorem 2.** *Let $c > 0$ be a sufficiently large constant. Fix $\alpha \in (0, r^{-cr})$ and $C_A, C_B \geq 2c\log(r/\alpha)$. Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, and $\mu$ be a distribution over $\mathcal{X} \times \mathcal{Y}$. Suppose $f$ satisfies*

$$\mathrm{suc}_\mu(f; C_A, C_B, r) \leq 1/2 + \alpha/2,$$

*then for any integer $n \geq 2$, we have*

$$\mathrm{suc}_{\mu^n}(f^{\oplus n}; 2^{-8}r^{-1}n \cdot C_A, 2^{-8}r^{-1}n \cdot C_B, r) \leq \frac{1}{2} + \frac{\alpha^{2^{-12}n}}{2}.$$

This distributional strong XOR lemma states that for any fixed input distribution $\mu$ and function $f$, to compute $f^{\oplus n}$ when the $n$ inputs are sampled independently from $\mu$, either the advantage is exponentially small in $\Omega(n)$, or one of the players need to communicate at least $\Omega(n/r)$ times more than one copy. This also gives a strong XOR lemma in the asymmetric communication, where we separately count how many bits Alice and Bob send.

It is worth noting that Shaltiel [Sha03] proved a similar strong XOR lemma for functions whose communication lower bound can be obtained via bounding the *discrepancy*. By the equivalence between the discrepancy and the correlation with 2-bit protocols [VW08], Theorem 2 implies their result. See Appendix A for a more detailed argument.

Note that a simple argument shows that Theorem 2 implies Theorem 1 (see also Section 4). Therefore, we will focus on the distributional version, and assume that the $n$ input pairs are sampled independently from some distribution $\mu$.

Our proof of the distributional version is inspired by the *information complexity* [CSWY01]. We define a new complexity measure for protocols, the *$\chi^2$-cost*, which is related to the *internal information cost* [BJKS04, BBCR13]. Roughly speaking, it replaces the KL-divergence in the internal information cost with the $\chi^2$-divergence, which can be viewed as the "exponential" version of KL. This provides better concentration, which is needed in our argument. Throughout the proof, we will also work with distributions that are "close to" communication protocols, i.e., the speaker's message may slightly depend on the receiver's input. Such distributions have also been studied in the proof of direct product theorems [JPY12, BRWY13a, BRWY13b]. We will provide more details in Section 2.

[1] Observe that since $\mathbf{R}^{(r)}_{0.51}(f) \leq \mathbf{R}^{(r)}_{0.99}(f) \leq O(\mathbf{R}^{(r)}_{0.51}(f))$, the constant $2/3$ does not matter as long as it is in $(1/2, 1)$. Our proof will also show that the base in $2^{-n}$ can be any constant.

### 1.1 Related work

As we mentioned earlier, Shaltiel [Sha03] proved a strong XOR lemma for functions whose communication lower bound can be obtained via bounding the discrepancy. Sherstov [She11] extended this bound to generalized discrepancy and quantum communication complexity.

Barak, Braverman, Chen and Rao [BBCR13] obtained an XOR lemma for the information complexity and then an XOR lemma for communication (with worst parameters) via information compression. However, their XOR lemma does not give exponentially small advantage. They proved that if $f$ is hard to compute with information cost $C$, then $f^{\oplus n}$ is hard to compute with information cost $O(n \cdot C)$. In fact, the starting point of our proof is an alternative view of their argument, which we will outline in Section 2.1.

Viola and Wigderson [VW08] proved a strong XOR lemma for multi-player $c$-bit communication for small $c$. As pointed out in their paper, it implies the XOR lemma by Shaltiel [Sha03]. XOR lemmas have also been proved in circuit complexity [Yao82, Lev87, Imp95, IW97, GNW11], query complexity [Sha03, She11, Dru12, BKLS20], streaming [AN21] and for low degree polynomials [VW08].

Direct product and direct sum theorems, which are results of similar types, have also been studied in the literature. They ask to return the outputs of all $n$ copies instead of their XOR. Direct sum theorems state that the problem cannot be solved with the same probability unless $\Omega(n)$ times more resource is used, while direct product theorems state that the problem can only be solved with probability *exponentially small* in $\Omega(n)$ unless $\Omega(n)$ times more resource is used. The direct sum theorem for information complexity is known [CSWY01, BJKS04, BBCR13]. A direct sum theorem for communication complexity with suboptimal parameters can be obtained via information compression [BBCR13]. A direct sum theorem for bounded-round communication has been proved [BR11], and we use a similar argument in one component of the proof (see Section 2.6 and Section 7). Direct product theorems for communication complexity (with suboptimal parameters via information compression), bounded-round communication and from information complexity to communication complexity have also been studied [JPY12, BRWY13b, BW15].

## 2 Technical Overview

### 2.1 An alternative view of [BBCR13]

The starting point of our proof is an alternative view of the XOR lemma in [BBCR13] for *information complexity*, which does not give an exponentially small advantage. Running a protocol on an input pair sampled from some fixed input distribution defines a joint distribution over the input pairs and the transcripts. Information complexity studies that in this joint distribution, how much information the transcript reveals about the inputs. The (internal) information cost is defined as

$$I(X; \mathbf{M} \mid Y, M_0) + I(Y; \mathbf{M} \mid X, M_0),$$

where $\mathbf{M} = (M_0, M_1, \ldots, M_r)$ is the transcript and $M_0$ is the public random bits.[2] We assume that Alice sends all the odd $M_i$ and Bob sends all the even $M_i$. The internal information cost of a protocol is always at most its communication cost. It is also known that for bounded-round communication, the internal information complexity is roughly equal to the communication complexity [BR11] (up to some additive error probability).

[2] In the usual definition, the public random string is not part of the transcript. We add it for simplicity of notations. This does not change the values of the mutual information terms as it is already in the condition.


(a) Protocol $\pi_{<k}$ (b) Protocol $\pi_k$

Figure 1: Decomposition of $\pi$: ■ is the inputs, ▒ is sampled publicly, □ is sampled privately.

For the XOR lemma for information complexity, we consider input pair $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ sampled from $\mu^n$. Suppose there is a protocol $\pi$ computing $f^{\oplus n}$ with information cost $I$, we want to show that $f$ can be computed with information cost $\approx I/n$.

To this end, we show that $\pi$ can be "decomposed" into a protocol $\pi_{<n}$ computing $f^{\oplus n-1}$ with information cost $I_1$ and a protocol $\pi_n$ computing $f$ with information cost $I_2$ such that $I_1 + I_2 \approx I$, as follows (see also Figure 1).

- For $\pi_{<n}$, given $n-1$ input pairs, the players view them as $X_{<n}$ and $Y_{<n}$ as part of the inputs for $\pi$, where $X_{<n}$ denotes $(X_1, \ldots, X_{n-1})$ and $Y_{<n}$ denotes $(Y_1, \ldots, Y_{n-1})$; then the players publicly sample $X_n \sim \mu_X$, and Bob privately samples $Y_n$ conditioned on $X_n$; the players run $\pi$ to compute $f^{\oplus n}(X, Y)$; Bob sends one extra bit indicating $f(X_n, Y_n)$.
- For $\pi_n$, given one input pair, the players view it as $X_n$ and $Y_n$; then the players publicly sample $Y_{<n} \sim \mu_Y^{n-1}$, and Alice privately samples $X_{<n}$ conditioned on $Y_{<n}$; the players run $\pi$ to compute $f^{\oplus n}(X, Y)$; Alice sends one extra bit indicating $\oplus_{i=1}^{n-1} f(X_i, Y_i)$.

If $\pi$ computes $f^{\oplus n}$ correctly, then the two protocols compute $f^{\oplus n-1}$ and $f$ correctly respectively. For the information cost of $\pi_{<n}$ (if we exclude the last bit indicating $f(X_n, Y_n)$), the first term is equal to $I(X_{<n}; \mathbf{M} \mid Y_{<n}, X_n, M_0)$, since $X_n$ is sampled using public random bits. It is also equal to $I(X_{<n}; \mathbf{M} \mid Y, X_n, M_0)$ due to the *rectangle property* of communication protocols. For the information cost of $\pi_n$ (if we exclude the last bit), the first term is equal to $I(X_n; \mathbf{M} \mid Y, M_0)$ since $Y_{<n}$ is sampled using public random bits. Therefore, the first terms sum up to exactly $I(X; \mathbf{M} \mid Y, M_0)$, the first term in the information cost of $\pi$, by the chain rule of mutual information. Similarly, the second terms sum up to $I(Y; \mathbf{M} \mid X, M_0)$, the second term in the information cost of $\pi$.

Hence, including the last bits in the protocols, we have $I_1 + I_2 \leq I + O(1)$. Thus, by repeatedly applying this argument, we obtain a protocol for $f$ with information cost $I/n + O(1)$, as desired. Note that in this decomposition, the players *do not* need to sample the private parts explicitly. As long as they can send the messages from the same distribution (e.g., by directly sampling the messages conditioned on the previous messages and their own inputs), the information costs and correctness are not affected.

The original paper proves the same result by explicitly writing out the protocol for $f$ obtained after applying the above decomposition $i$ times for a random $i \in [n]$, and proving the expected cost is as claimed. The two proofs are essentially equivalent for this statement.[3] However, as we will see later, our new view is more flexible, allowing for more sophisticated manipulations when doing the decomposition.

## 2.2 Obtaining exponentially small advantage

The above decomposition preserves the success probability. However, if we start from a protocol for $f^{\oplus n}$ with exponentially small advantage, then we will not be able to obtain a protocol for $f$ with success proba-

[3]The orginal proof embeds the input to $f$ into a random coordinate $i$ of $f^{\oplus n}$, and samples $X_{>i}$ and $Y_{<i}$ using public random bits.


(a) Bob's view in $\pi_{<k}$ (b) Alice's view in $\pi_k$

Figure 2: Decomposition of $\pi$: ■ is unknown inputs, ■ is known.

bility $2/3$, which is required in order to prove the strong XOR lemma.[4]

Let $\mathrm{adv}(f(X,Y) \mid \mathbf{R})$ denote the advantage for $f(X,Y)$ conditioned on $\mathbf{R}$, which is defined as

$$|2\Pr[f(X,Y)=1 \mid \mathbf{R}]-1|,$$

i.e., the advantage is $\alpha$ if the conditional probability is either $1/2+\alpha/2$ or $1/2-\alpha/2$.

Now let us take a closer look at the two protocols $\pi_{<n}$ and $\pi_n$ (see Figure 2). For $\pi_{<n}$, in *Bob's view* at the end of the communication, he knows his input $Y_{<n}$, the publicly sampled $X_n$ and the transcript $\mathbf{M}$. Hence, he is able to predict $f^{\oplus n-1}(X_{<n},Y_{<n})$ with advantage $\mathrm{adv}(f^{\oplus n-1}(X_{<n},Y_{<n}) \mid X_n,Y_{<n},\mathbf{M})$. By letting Bob send one extra bit indicating his prediction, the advantage of the protocol achieves the same. For $\pi_n$, in *Alice's view* at the end of the communication, she knows her input $X_n$, the publicly sampled $Y_{<n}$ and the transcript $\mathbf{M}$. Hence, she is able to predict $f(X_n,Y_n)$ with advantage $\mathrm{adv}(f(X_n,Y_n) \mid X_n,Y_{<n},\mathbf{M})$. By letting Alice send one extra bit indicating her prediction, the advantage of the protocol achieves the same.

Now an important observation is that $X_{<n}$ and $Y_n$ are independent conditioned on $(X_n,Y_{<n},\mathbf{M})$, by the rectangle property of communication protocols. Hence, $f^{\oplus n-1}(X_{<n},Y_{<n})$ and $f(X_n,Y_n)$ are also independent conditioned on $(X_n,Y_{<n},\mathbf{M})$. Since $f^{\oplus n}(X,Y)=f^{\oplus n-1}(X_{<n},Y_{<n})\oplus f(X_n,Y_n)$, by the probability of XOR of two independent bits, we have

$$\begin{aligned}\mathrm{adv}(f^{\oplus n}(X,Y) \mid X_n,Y_{<n},\mathbf{M}) = \mathrm{adv}(f^{\oplus n-1}(X_{<n},Y_{<n}) \mid X_n,Y_{<n},\mathbf{M}) \\ \times \mathrm{adv}(f(X_n,Y_n) \mid X_n,Y_{<n},\mathbf{M}).\end{aligned} \tag{1}$$

This suggests the following strategy for the decomposition:

- if the information cost of $\pi_n$ is large, then the information cost of $\pi_{<n}$ must be much smaller than that of $\pi$;
- if the information cost of $\pi_n$ is small and its advantage for $f$ is large, then we have obtained a good protocol for $f$;
- if the information cost of $\pi_n$ is small and its advantage for $f$ is small, then by (1), the advantage of $\pi_{<n}$ must be larger than that of $\pi$ by some factor.

Hence, in each decomposition, if we don't already obtain a good protocol for $f$, then when decrementing $n$ to $n-1$, we must either significantly decrease the information cost, or increase the advantage by a multiplicative factor. If we start with a protocol with a low cost and a mild-exponentially small advantage for $f^{\oplus n}$, then we must obtain a good protocol for $f$ by applying this decomposition iteratively.

[4] In fact, this is inherent for information complexity, since the strong XOR lemma for *information complexity* does not hold. This is because the information complexity is an average measure, and it is at most the *expected* communication. A protocol can choose to compute *all* $f(X_i,Y_i)$ with probability $\varepsilon$ and output a random bit otherwise, which achieves success probability $1/2+\varepsilon$ and expected communication $\varepsilon n$ times the one-copy cost. However, the reader is encouraged to continue reading this subsection pretending that they do not know about this counterexample.

It turns out that the main difficulty in applying the above strategy is to formalize the last bullet point. Note that the expected advantage of $\pi_n$ (after Alice sending the one extra bit indicating her prediction) is $\mathbb{E}\left[\mathrm{adv}(f(X_n, Y_n) \mid X_n, Y_{<n}, \mathbf{M})\right]$, the expected advantage of $\pi_{<n}$ (after Bob sending the one extra bit indicating his prediction) is $\mathbb{E}\left[\mathrm{adv}(f^{\oplus n-1}(X_{<n}, Y_{<n}) \mid X_n, Y_{<n}, \mathbf{M})\right]$, and the expected advantage of $\pi$ is $\mathbb{E}\left[\mathrm{adv}(f^{\oplus n}(X, Y) \mid \mathbf{M})\right]$, which is at most $\mathbb{E}\left[\mathrm{adv}(f^{\oplus n}(X, Y) \mid X_n, Y_{<n}, \mathbf{M})\right]$.

When we say that the advantage of $\pi_n$ for $f$ is small in the last bullet point, we can only guarantee that this expectation is small. Equation (1), which is a pointwise equality, does not directly give any useful bounds on the expectations. For example, it is possible that both $\mathbb{E}\left[\mathrm{adv}(f(X_n, Y_n) \mid X_n, Y_{<n}, \mathbf{M})\right]$ and $\mathbb{E}\left[\mathrm{adv}(f^{\oplus n-1}(X_{<n}, Y_{<n}) \mid X_n, Y_{<n}, \mathbf{M})\right]$ are very small, but $\mathrm{adv}(f(X_n, Y_n) \mid X_n, Y_{<n}, \mathbf{M})$ and $\mathrm{adv}(f^{\oplus n-1}(X_{<n}, Y_{<n}) \mid X_n, Y_{<n}, \mathbf{M})$ are always equal to zero or one at the same time, both concentrated on a small probability set. Then we have $\mathbb{E}[\mathrm{adv}(f^{\oplus n}(X, Y) \mid X_n, Y_{<n}, \mathbf{M})] = \mathbb{E}[\mathrm{adv}(f^{\oplus n-1}(X_{<n}, Y_{<n}) \mid X_n, Y_{<n}, \mathbf{M})]$, the advantage may not increase at all. In this case, the advantage $\mathrm{adv}(f^{\oplus n}(X, Y) \mid X_n, Y_{<n}, \mathbf{M})$ is also concentrated on the same small probability set.

On the other hand, observe that if $\mathrm{adv}(f^{\oplus n}(X, Y) \mid X_n, Y_{<n}, \mathbf{M})$ takes roughly the same value (say, $\varepsilon$) most of the time, then we do obtain an advantage increase:

$$\begin{aligned}
&\mathbb{E}[\mathrm{adv}(f^{\oplus n-1}(X_{<n}, Y_{<n}) \mid X_n, Y_{<n}, \mathbf{M})] \\
&\quad = \mathbb{E}[\varepsilon / \mathrm{adv}(f(X_n, Y_n) \mid X_n, Y_{<n}, \mathbf{M})] \\
&\quad \geq \varepsilon / \mathbb{E}[\mathrm{adv}(f(X_n, Y_n) \mid X_n, Y_{<n}, \mathbf{M})]
\end{aligned}$$

by the convexity of $1/x$.

This motivates us to consider the following two extreme cases:

1. $\mathrm{adv}(f^{\oplus n}(X, Y) \mid X_n, Y_{<n}, \mathbf{M})$ is roughly uniformly distributed among all $(X_n, Y_{<n}, \mathbf{M})$;
2. $\mathrm{adv}(f^{\oplus n}(X, Y) \mid X_n, Y_{<n}, \mathbf{M})$ is concentrated on a tiny fraction of the triples $(X_n, Y_{<n}, \mathbf{M})$.

Basically following what we just argued, the above strategy directly applies in the first case. The second case is related to the *direct product theorems*, where we also want to analyze protocols that is correct with exponentially small probability. This is because one possible strategy for the players is to compute all $f(X_i, Y_i)$ correctly with some probability $\varepsilon$ and output a random bit otherwise. We must at least show that in this case, $\varepsilon \leq \exp(-\Omega(n))$.

## 2.3 Generalized protocols

For the second case above, we follow one strategy for direct product theorems [BRWY13b]. When the advantage $\mathrm{adv}(f^{\oplus n}(X, Y) \mid X_n, Y_{<n}, \mathbf{M})$ is concentrated on a small set $U$ of triples $(X_n, Y_{<n}, \mathbf{M})$, we restrict our attention to $U$ by *conditioning* $\pi$ on $U$. However, this immediately creates two issues.

The first issue is that although $\pi \mid U$ is a well-defined distribution, it is not necessarily a *protocol*, since conditioning on an arbitrary event may break the independence between a message and the receiver's input, e.g., $M_1$ may no longer be independent of $Y$conditioned on $X$.[5]

This issue was also encountered in the direct product theorem proofs. Instead of studying standard protocols, we focus on *generalized protocols*, where we allow each message to depend on both player's

[5]Conditioning on an event also distorts the input distribution, which needs to be handled. But for simplicity, we omit it in the overview.

inputs, and we wish to restrict the correlation between the odd $M_i$ and Bob's input and the correlation between the even $M_i$ and Alice's input. In the previous work, it bounds

$$\theta(\pi) := \sum_{\text{odd } i} I(M_i; Y \mid X, M_{<i}) + \sum_{\text{even } i} I(M_i; X \mid Y, M_{<i}),$$

the mutual information between the message and the receiver's input.

Intuitively, the $\theta$-value measures how close to a standard protocol a generalized protocol is. It turns out that the $\theta$-value of a standard protocol conditioned on a not-too-small probability event is small; on the other hand, when the $\theta$-value is small, it is statistically close to a standard protocol. Furthermore, an important feature of $\theta(\pi)$ is that the decomposition of $\pi$ into $\pi_{<n}$ and $\pi_n$ also satisfies that $\theta(\pi) = \theta(\pi_{<n}) + \theta(\pi_n)$. Hence, when doing the decomposition, we can hope to obtain a generalized protocol for $f$ that is very close to a standard protocol.

The second issue is that conditioning on a small probability event $U$ could greatly increase the information cost, from $I$ to $\Omega(I/\Pr[U])$. Since $I$ is close to the communication cost, such a multiplicative loss in each step of decomposition is unaffordable. Such a loss occurs because the mutual information is an average measure (an expectation), which does not provide any concentration (also recall the counterexample in footnote 4 where the communication cost and the advantage are both concentrated on an $\varepsilon$-probability event, when we condition on this event, both the expected communication cost and the advantage increase by a factor of $1/\varepsilon$). More specifically, consider the first term in the information cost, $I(X; \mathbf{M} \mid Y)$ (omit the public random bits for now). For standard protocols, it is equal to

$$\sum_{x,y,\mathbf{m}} \pi(x, y, \mathbf{m}) \cdot \log\left(\frac{\pi(x \mid \mathbf{m}, y)}{\pi(x \mid y)}\right) = \mathop{\mathbb{E}}_{\pi}\left[\log\left(\frac{\pi(X \mid \mathbf{M}, Y)}{\pi(X \mid Y)}\right)\right] = \mathop{\mathbb{E}}_{\pi}\left[\log\left(\frac{\pi(X \mid \mathbf{M}, Y)}{\mu(X \mid Y)}\right)\right].$$

If we only have a bound on this expectation, then inevitably its value can greatly increase after conditioning on a small probability event, not to say that the logarithm inside the expectation is *not* nonnegative, so it can get worse than what Markov's inequality gives.

We also note that the argument in the previous subsections crucially uses the *rectangle property* of the communication protocols, which does not necessarily hold for generalized protocols. This turns out not to be a real issue, since throughout the argument, we will maintain the rectangle property *at all leaves*, which is sufficient for the argument to go through (see also Section 2.6).

## 2.4 $\theta$-cost and $\chi^2$-costs

Our novel solution to the second issue above is to focus on the "exponential version" of the information cost, i.e., for the first term,

$$\chi^2_{\mu,A}(\pi) := \mathop{\mathbb{E}}_{\pi}\left[\frac{\pi(X \mid \mathbf{M}, Y)}{\mu(X \mid Y)}\right],$$

which we call the $\chi^2$-cost by Alice. The $\chi^2$-cost by Bob, $\chi^2_{\mu,B}(\pi)$, is defined similarly for the second term in the information cost (see Definition 15).

This notion of the cost has the following benefits.

- For a (deterministic) standard protocol with $C$ bits of communication, $\chi^2_{\mu,A}(\pi) \leq 2^C$. Hence, it corresponds to the exponential of the communication cost.

- When conditioning on a small probability event $U$, we can essentially ensure that it increases by a factor of $O(1/\pi(U))$ (Lemma 37 gives a more generalized statement). Effectively, this only adds $\log(1/\pi(U))$ to the communication cost, which becomes affordable.

Note that the mutual information is the expected KL-divergence, and the $\chi^2$-cost is the expected $\chi^2$-divergence (plus one). Similarly, we also define an "exponential version" of $\theta(\pi)$, which we call the $\theta$-cost of $\pi$ (see Definition 12). It also ensures that the value does not increase significantly when conditioning on a small probability event.

On the other hand, going from mutual information to its "exponential version" loses many of its good properties, most importantly, the chain rule. The next crucial observation is that the chain rule for mutual information in fact holds *pointwisely*, which enables us to work with the $\chi^2$-costs.

More specifically, let $X, Y, Z$ be three random variables with joint distribution $\pi$, the chain rules says $I(X; Y, Z) = I(X; Y) + I(X; Z \mid Y)$. By writing the mutual information as an expectation, this is

$$\mathbb{E}\left[\log\left(\frac{\pi(Y, Z \mid X)}{\pi(Y, Z)}\right)\right] = \mathbb{E}\left[\log\left(\frac{\pi(Y \mid X)}{\pi(Y)}\right)\right] + \mathbb{E}\left[\log\left(\frac{\pi(Z \mid X, Y)}{\pi(Z \mid Y)}\right)\right].$$

This equality holds pointwisely in the sense that for any concrete values $(x, y, z)$, the equality holds for the logarithms inside the expectation

$$\log\left(\frac{\pi(y, z \mid x)}{\pi(y, z)}\right) = \log\left(\frac{\pi(y \mid x)}{\pi(y)}\right) + \log\left(\frac{\pi(z \mid x, y)}{\pi(z \mid y)}\right)$$

by the definition of conditional probability.

Therefore, the "exponential version" also holds pointwisely:

$$\frac{\pi(y, z \mid x)}{\pi(y, z)} = \frac{\pi(y \mid x)}{\pi(y)} \cdot \frac{\pi(z \mid x, y)}{\pi(z \mid y)}.$$

This is what we use in replacement of the chain rule for mutual information. See the next subsection for more details.

## 2.5 Proof outline

We now give an outline of the proof of the following statement: Given an $r$-round standard protocol $\pi$ for $f^{\oplus n}$ with communication cost $o(n \cdot C)$ that succeeds with advantage $\alpha^{o(n)}$ on the inputs sampled from $\mu^n$, we can obtain an $r$-round generalized protocol $\rho$ for $f$ with $\chi^2$-costs $\approx 2^C$, $\theta$-cost $\approx 1/\alpha$ and advantage $\approx \alpha$. We will then discuss how to convert such a generalized protocol to a standard protocol with low communication cost in the next subsection.

We first show that $\pi$ is also a generalized protocol with $\chi^2$-cost $2^{o(nC)}$ and $\theta$-cost 1 (in the proof of Lemma 25). Next, we decompose $\pi$ into $\pi_{<n}$ for $f^{\oplus n-1}$ and $\pi_n$ for $f$, and prove that the product of the $\theta$-cost [resp. $\chi^2$-costs] of $\pi_{<n}$ and $\pi_n$ is that of $\pi$ pointwisely (Section 5). Now if the advantage of $\pi$ is not roughly evenly distributed, we will identify an event $U$ such that the advantage conditioned on $U$ is much higher than the average advantage, and more importantly, the advantage within $U$ becomes roughly evenly distributed (not concentrated on any small probability event in $U$) (Section 6.1). Conditioning on $U$ increases the advantage while also increases the $\theta$-cost and $\chi^2$-costs, it turns out that they all increase by about the same factor. Next, we partition the sample space of $\pi$ into $S_{\text{high-cost}}, S_{\text{low-cost}}$ and $S_{\text{low-prob}}$ such that

- in $S_{\text{high-cost}}$, $\pi_n$ has high $\theta$-cost or high $\chi^2$-cost (excluding some corner cases), say $\geq 1/\alpha$ for $\theta$-cost or $\geq 2^C$ for $\chi^2$-cost,
- in $S_{\text{low-cost}}$, $\pi_n$ has low $\theta$-cost and low $\chi^2$-cost (also excluding some corner cases),
- $S_{\text{low-prob}}$ is the rest, which will happen with very low probability.

Since the advantage is not concentrated on any small probability in $U$, then (at least) one of $S_{\text{high-cost}}$ or $S_{\text{low-cost}}$ will have advantage about as high as the advantage of $U$. If $S_{\text{high-cost}}$ has the advantage as high as $U$, then we prove that by the pointwise equality for the costs, $\pi_{<n} \mid S_{\text{high-cost}}$ must have a much smaller cost than $\pi \mid U$, while they have roughly the same advantage (Section 6.2). If $S_{\text{low-cost}}$ has the advantage as high as $U$, then if $\pi_n \mid S_{\text{low-cost}}$ has high advantage, then we obtain a desired generalized protocol for $f$ with low costs and high advantage; otherwise we prove that $\pi_{<n} \mid S_{\text{low-cost}}$ has a much higher advantage than $\pi \mid U$ (as the advantage of $\pi$ is roughly evenly distribution within $U$), while they have roughly the same costs (Section 6.3).

To summarize the above argument, if we don't already find a desired generalized protocol for $f$, then when decrementing $n$ to $n-1$, we first condition on an event $U$, increasing costs and advantage simultaneously by about the same (while arbitrary) factor, then either we reduce the $\theta$-cost by a factor of $\geq 1/\alpha$, or we reduce the $\chi^2$-costs by a factor of $\geq 2^C$, or we increase the advantage by a factor of $\geq 1/\alpha$. Since we start with $\chi^2$-costs $2^{o(nC)}$, $\theta$-cost 1 and advantage $\alpha^{o(n)}$, we cannot repeat this for $n$ steps without finding a desired protocol for $f$. More formally, we will measure the progress by using a potential function that depends on the costs and advantage of the current protocol, and show that each time we decrement from $k$ to $k-1$, how much the potential must decrease (Section 4).

### 2.6 Convert a generalized protocol to a standard protocol

Finally, we need to show that the existence of a good generalized protocol implies the existence of a good standard protocol. We prove that if an $r$-round generalized protocol $\rho$ has $\chi^2$-costs $2^C$, $\theta$-cost $1/\alpha$ and advantage $\alpha$, then there is an $r$-round standard protocol $\tau$ with communication cost $\approx C$ and advantage $\approx \alpha^3$. Together with what we summarized in the last subsection, we obtain the strong XOR lemma for $r$-round communication.

[BR11] converts a *standard* protocol $\rho$ with constant rounds to a standard protocol with communication matching the internal information cost of $\rho$. Using a similar argument, we can convert $\rho$ to a standard protocol with communication $\approx C$. By the convexity of $2^x$, $\chi^2$-cost of $2^C$ implies internal information cost of at most $C$. It turns out that the (almost) same argument applies in our case, for generalized protocol $\rho$.

Then the next crucial observation is that we can ensure the generalized protocol $\rho$ that we obtain from the arguments in the previous subsection has the *rectangle property* with respect to $\mu$. Roughly speaking, it means that for all transcripts $\mathbf{M}$, if we look at the ratio of the probabilities $\frac{\rho(X,Y|\mathbf{M})}{\mu(X,Y)}$, it is a product function of $X$ and $Y$, i.e., it is equal to $g_A(X) \cdot g_B(Y)$ for some functions $g_A, g_B$ that may depend on $\mathbf{M}$. Note that a standard protocol has the rectangle property, since each message depends only on either $X$ or $Y$, and the same property holds even conditioned on any prefix of the transcript $M_{<i}$. A generalized protocol may not have this property in general, but we can ensure that the protocol we obtain has this product structure conditioned on any *complete* transcript $\mathbf{M}$.

After generating a transcript $\mathbf{M}$ using [BR11], the rectangle property allows the players to locally "readjust" the probabilities (via rejection sampling) so that after the readjustment, the probability of a triple $(X, Y, \mathbf{M})$ is proportional to the "right" probability $\rho(X, Y, \mathbf{M})$, which in turn, gives the advantage proportional to that of $\rho$.

The probability that is sacrificed in the rejection sampling depends on how far $\rho$ is from a standard protocol, i.e., the $\theta$-cost of $\rho$. It turns out that the above argument gives an overall advantage of at least $\alpha^2$ divided by the $\theta$-cost of $\rho$. See Section 7 for the formal proof.

# 3 Notations and Definitions for Generalized Protocols

## 3.1 Notations and standard probabilities

Throughout the paper, all logarithms have base 2. We use $[n]$ to denote the set $\{1, \ldots, n\}$. Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a binary-valued function. We use $f^{\oplus n}$ to denote the function $f^{\oplus n} : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ such that

$$f^{\oplus n}(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = \bigoplus_{i=1}^{n} f(X_i, Y_i),$$

where $\oplus$ is the XOR operation.

Let $X$ be a vector $(X_1, \ldots, X_n)$. We denote the prefix $(X_1, \ldots, X_i)$ by $X_{\leq i}$. Similarly, $X_{<i}$ denotes $(X_1, \ldots, X_{i-1})$ and $X_{>i}$ denotes $(X_{i+1}, \ldots, X_n)$. For vectors $X$ where we start the index from 0, $X_{\leq i}$ denotes $(X_0, \ldots, X_i)$.

Let $\pi$ be a distribution over triples $(X, Y, \mathbf{M}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{M}$, where $\mathbf{M} = (M_0, \ldots, M_r)$. For an event $W \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{M}$, we use $\pi(W)$ to denote its probability. For a random variable $\mathbf{M}$, we use $\pi(\mathbf{M})$ to denote the probability of $\mathbf{M}$ in distribution $\pi$, which by itself is a random variable that depends on the value of $\mathbf{M}$. It is similar for multiple variables, e.g., $\pi(X, M_{<i})$ denotes the probability of $(X, M_{<i})$.

Let $S$ be a set of possible values of several variables, say, $S$ is a set of possible values of $(X, M_{<i})$. We use $\pi(S)$ to denote the probability that $(X, M_{<i}) \in S$, i.e., $\pi(S) = \Pr_\pi[(X, M_{<i}) \in S] = \pi(\{(X, Y, \mathbf{M}) : (X, M_{<i}) \in S\})$. When there is no ambiguity, we may abuse the notation, and use $S$ to denote the event that $(X, M_{<i}) \in S$, which is the set $\{(X, Y, \mathbf{M}) : (X, M_{<i}) \in S\}$, e.g., if $T$ is a set of possible values of $(Y, M_{<j})$, then $S \cap T$ is the event that $(X, M_{<i}) \in S \wedge (Y, M_{<j}) \in T$, which is the set $\{(X, Y, \mathbf{M}) : (X, M_{<i}) \in S \wedge (Y, M_{<j}) \in T\}$.

The $\chi^2$-divergence of two distributions is defined as follows.

**Definition 3** ($\chi^2$-divergence)**.** Let $P$ and $Q$ be two distributions over a sample space $\mathcal{X}$. The $\chi^2$-divergence from $Q$ to $P$ is

$$\mathbf{D}_{\chi^2}(P \,\|\, Q) = \sum_{x \in \mathcal{X}} \frac{P(x)^2}{Q(x)} - 1 = \mathop{\mathbb{E}}_{x \sim P}\left[\frac{P(x)}{Q(x)}\right] - 1.$$

The KL-divergence of two distributions is defined as follows.

**Definition 4** (KL-divergence)**.** Let $P$ and $Q$ be two distributions over a sample space $\mathcal{X}$. The KL-divergence from $Q$ to $P$ is

$$\mathbf{D}_{\mathrm{KL}}(P \,\|\, Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right) = \mathop{\mathbb{E}}_{x \sim P}\left[\log\left(\frac{P(x)}{Q(x)}\right)\right].$$

A simple calculation gives the following proposition.

**Proposition 5.** *Let $R_1, R_2 \in \{0,1\}$ be two* independent *random variables such that $\Pr[R_1 = 0] = \frac{1}{2} + \frac{\sigma_1}{2}$ and $\Pr[R_2 = 0] = \frac{1}{2} + \frac{\sigma_2}{2}$. Then $\Pr[R_1 \oplus R_2 = 0] = \frac{1}{2} + \frac{\sigma_1\sigma_2}{2}$.*

### 3.2 Generalized communication protocols

For most standard communication protocols discussed in this paper, we pair it with an input distribution, and study the joint distribution.

**Definition 6** (standard protocols)**.** An $r$-round *standard protocol* $\pi$ for input distribution $\mu$ is a distribution over triples

$$(X, Y, \mathbf{M}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{M},$$

where the transcript $\mathbf{M} = (M_0, \ldots, M_r)$, and each $M_i$ is chosen from a prefix-free set of strings that only depends on $M_{<i}$. Moreover, $(X, Y) \sim \mu$; the public random string $M_0$ is independent of $(X, Y)$; for odd $i \geq 1$, $M_i$ (a message by Alice) is independent of $Y$ conditioned on $X$ and $M_{<i}$; for even $i \geq 1$, $M_i$ (a message by Bob) is independent of $X$ conditioned on $Y$ and $M_{<i}$. The output of $\pi$ is a function of $\mathbf{M}$.

Now we define $\mathrm{suc}_\mu(f; C_A, C_B, r)$ to be the maximum success probability of a protocol computing $f$ under the communication cost constraints.

**Definition 7.** Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a function, and $\mu$ be a distribution over $\mathcal{X} \times \mathcal{Y}$. For $C_A, C_B, r \geq 1$, let

$$\mathrm{suc}_\mu(f; C_A, C_B, r)$$

be the supremum over all $r$-round *standard* protocols $\pi$ where Alice sends at most $C_A$ bits and Bob sends at most $C_B$ bits in a round, the probability that the output of $\pi$ is equal to $f(X, Y)$ when $(X, Y)$ is sampled from $\mu$.

*Remark.* Without loss of generality, we may assume that $M_r \in \{0, 1\}$. This is because the output of the protocol is a function of $\mathbf{M}$. Instead of $M_r$, we could always let Bob send the output, which has only one bit.

Next, we define generalized protocols.

**Definition 8** (generalized protocols)**.** An $r$-round *generalized protocol* $\pi$ is a distribution over triples

$$(X, Y, \mathbf{M}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{M},$$

where $\mathbf{M} = (M_0, M_1, \ldots, M_r)$, and each $M_i$ is chosen from a prefix-free set of strings that only depends on $M_{<i}$.

In this paper, we also only consider generalized protocols with $M_r \in \{0, 1\}$.

Clearly, a standard protocol is also a generalized protocol. One still should think $M_0$ as the public random bits, and think $M_i$ as a message sent by Alice if $i$ is odd, and sent by Bob if $i$ is even. The messages and public random bits are allowed to be arbitrarily correlated with both players' inputs.

We do not explicitly define the output of a generalized protocol in this paper. When we study the correctness of a generalized protocol when computing some function $f$, we characterize it using the *advantage*.

**Definition 9** (advantage)**.** Let $\pi$ be a generalized protocol, and $f$ be a binary-valued random variable (e.g., $f = f(X, Y)$ for $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$). The advantage of $\pi$ for $f$ conditioned on $W$ is

$$\mathrm{adv}_\pi(f \mid W) := |2\pi(f = 0 \mid W) - 1| = |2\pi(f = 1 \mid W) - 1|\,.$$

We may omit the subscript $\pi$ when there is no ambiguity.

Note that fixing $\pi$ and $f$, $\mathrm{adv}_\pi(f \mid W)$ is a function of $W$. For a standard protocol with input distribution $\mu$, the larges probability that the output can equal $f$ when the transcript is $\mathbf{M}$ is

$$\frac{1}{2} + \frac{1}{2} \cdot \mathrm{adv}_\pi(f(X,Y) \mid \mathbf{M}).$$

Thus, the overall success probability is always at most

$$\frac{1}{2} + \frac{1}{2} \cdot \mathop{\mathbb{E}}_{\mathbf{M}\sim\pi} \left[\mathrm{adv}_\pi(f(X,Y) \mid \mathbf{M})\right].$$

For general protocols, we will also use $\mathbb{E}_{\mathbf{M}\sim\pi}\left[\mathrm{adv}_\pi(f(X,Y) \mid \mathbf{M})\right]$ to characterize the success probability.

The (conditional) advantage is superadditive when weighted by the probability of the condition.

**Lemma 10.** *Let $W_1, W_2$ be* disjoint *events and $\mathbf{R}$ be a set of random variables, then*

$$\begin{aligned}
&\pi(W_1 \cup W_2) \cdot \mathop{\mathbb{E}}_{\pi|W_1\cup W_2} \left[\mathrm{adv}(f(X,Y) \mid \mathbf{R}, W_1 \cup W_2)\right] \\
&\quad \le \pi(W_1) \cdot \mathop{\mathbb{E}}_{\pi|W_1} \left[\mathrm{adv}(f(X,Y) \mid \mathbf{R}, W_1)\right] + \pi(W_2) \cdot \mathop{\mathbb{E}}_{\pi|W_2} \left[\mathrm{adv}(f(X,Y) \mid \mathbf{R}, W_2)\right].
\end{aligned}$$

*Proof.* By definition, we have

$$\begin{aligned}
&\pi(W_1) \cdot \mathop{\mathbb{E}}_{\pi|W_1} \left[\mathrm{adv}(f(X,Y) \mid \mathbf{R}, W_1)\right] + \pi(W_2) \cdot \mathop{\mathbb{E}}_{\pi|W_2} \left[\mathrm{adv}(f(X,Y) \mid \mathbf{R}, W_2)\right] \\
&\quad = \pi(W_1) \cdot \sum_{\mathbf{R}} \pi(\mathbf{R} \mid W_1) \cdot |2\pi(f(X,Y) = 1 \mid \mathbf{R}, W_1) - 1| \\
&\qquad + \pi(W_2) \cdot \sum_{\mathbf{R}} \pi(\mathbf{R} \mid W_2) \cdot |2\pi(f(X,Y) = 1 \mid \mathbf{R}, W_2) - 1| \\
&\quad = \sum_{\mathbf{R}} |2\pi(f(X,Y) = 1, \mathbf{R}, W_1) - \pi(\mathbf{R}, W_1)| + \sum_{\mathbf{R}} |2\pi(f(X,Y) = 1, \mathbf{R}, W_2) - \pi(\mathbf{R}, W_2)|
\end{aligned}$$

which by the fact that $W_1$ and $W_2$ are disjoint, is

$$\begin{aligned}
&\ge \sum_{\mathbf{R}} |2\pi(f(X,Y) = 1, \mathbf{R}, W_1 \cup W_2) - \pi(\mathbf{R}, W_1 \cup W_2)| \\
&= \pi(W_1 \cup W_2) \cdot \mathop{\mathbb{E}}_{\pi|W_1\cup W_2} \left[\mathrm{adv}(f(X,Y) \mid \mathbf{R}, W_1 \cup W_2)\right].
\end{aligned}$$

□

The following proposition states that knowing more could only increase the expected advantage.

**Proposition 11.** *Let $\mathbf{R}_1, \mathbf{R}_2$ be two random variables, then*

$$\mathop{\mathbb{E}}_{\mathbf{R}_1\sim\pi} \left[\mathrm{adv}(f \mid \mathbf{R}_1)\right] \le \mathop{\mathbb{E}}_{\mathbf{R}_1,\mathbf{R}_2\sim\pi} \left[\mathrm{adv}(f \mid \mathbf{R}_1, \mathbf{R}_2)\right].$$

*Proof.* We have

$$\mathop{\mathbb{E}}_{\mathbf{R}_1,\mathbf{R}_2\sim\pi} \left[\mathrm{adv}(f \mid \mathbf{R}_1, \mathbf{R}_2)\right]$$

$$
\begin{aligned}
&= \sum_{\mathbf{R}_1,\mathbf{R}_2} \pi(\mathbf{R}_1,\mathbf{R}_2)\cdot|2\pi(f=1\mid\mathbf{R}_1,\mathbf{R}_2)-1| \\
&= \sum_{\mathbf{R}_1}\pi(\mathbf{R}_1)\sum_{\mathbf{R}_2}\pi(\mathbf{R}_2\mid\mathbf{R}_1)\cdot|2\pi(f=1\mid\mathbf{R}_1,\mathbf{R}_2)-1| \\
&\geq \sum_{\mathbf{R}_1}\pi(\mathbf{R}_1)\cdot\left|\sum_{\mathbf{R}_2}\pi(\mathbf{R}_2\mid\mathbf{R}_1)\cdot 2\pi(f=1\mid\mathbf{R}_1,\mathbf{R}_2)-\sum_{\mathbf{R}_2}\pi(\mathbf{R}_2\mid\mathbf{R}_1)\right| \\
&= \sum_{\mathbf{R}_1}\pi(\mathbf{R}_1)\cdot|2\pi(f=1\mid\mathbf{R}_1)-1| \\
&= \mathop{\mathbb{E}}_{\mathbf{R}_1\sim\pi}\left[\mathrm{adv}(f\mid\mathbf{R}_1)\right].
\end{aligned}
$$

□

### 3.3 The $\theta$-cost and $\chi^2$-costs

In a standard protocol, Alice's message must be independent of Bob's input conditioned on Alice's input and the previous messages, and vice versa, while we allow arbitrary correlation in a generalized protocol. The $\theta$-*cost* of a generalized protocol measures this correlation.

**Definition 12** ($\theta$-cost)**.** The $\theta$-cost of $\pi$ with respect to $\mu$ at $(X,Y,\mathbf{M})$ is

$$
\begin{aligned}
\theta_\mu(\pi\mathbin{@}X,Y,\mathbf{M}) &:= \frac{\pi(X,Y\mid M_0)}{\mu(X,Y)}\cdot\prod_{\text{odd }i\in[r]}\frac{\pi(M_i\mid X,Y,M_{<i})}{\pi(M_i\mid X,M_{<i})}\cdot\prod_{\text{even }i\in[r]}\frac{\pi(M_i\mid X,Y,M_{<i})}{\pi(M_i\mid Y,M_{<i})} \\
&= \frac{\pi(X,Y,\mathbf{M})}{\pi(M_0)\cdot\mu(X,Y)\cdot\prod_{\text{odd }i\in[r]}\pi(M_i\mid X,M_{<i})\cdot\prod_{\text{even }i\in[r]}\pi(M_i\mid Y,M_{<i})}.
\end{aligned}
$$

The $\theta$-cost of $\pi$ with respect to $\mu$ is

$$
\theta_\mu(\pi) := \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi}\left[\theta_\mu(\pi\mathbin{@}X,Y,\mathbf{M})\right].
$$

For an event $W$, the $\theta$-cost of $\pi$ respect to $\mu$ conditioned on $W$ is

$$
\theta_\mu(\pi\mid W) := \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi|W}\left[\theta_\mu(\pi\mathbin{@}X,Y,\mathbf{M})\right].
$$

*Remark.* We emphasize that $\theta_\mu(\pi\mid W)$ is different from $\theta_\mu(\pi_W)$ for $\pi_W$ being the distribution of $\pi$ conditioned on $W$. According to the definitions, although $(X,Y,\mathbf{M})$ is sampled from $\pi\mid W$ in both cases, the quantity inside the expectation is different. For $\theta_\mu(\pi\mid W)$, we still measure the $\theta$-cost at $(X,Y,\mathbf{M})$ according to distribution $\pi$, while for $\theta_\mu(\pi_W)$, we measure the $\theta$-cost at $(X,Y,\mathbf{M})$ according to $\pi_W$.

*Remark.* Let $\tau$ be the protocol obtained by "making $\pi$ standard." That is, $\tau(X,Y)$ is equal to $\mu(X,Y)$; $\tau(M_0)$ is $\pi(M_0)$, independent of $(X,Y)$. Each odd $M_i$ is sampled according to $\pi(M_i\mid X,M_{<i})$ independent of $Y$, and each even $M_i$ is sampled according to $\pi(M_i\mid Y,M_{<i})$ independent of $X$. Then $\tau$ is a standard protocol such that

$$
\tau(X,Y,\mathbf{M}) = \pi(M_0)\cdot\mu(X,Y)\cdot\prod_{\text{odd }i\in[r]}\pi(M_i\mid X,M_{<i})\cdot\prod_{\text{even }i\in[r]}\pi(M_i\mid Y,M_{<i}).
$$

The $\theta$-cost of $\pi$ is simply the $\chi^2$-divergence from $\tau$ to $\pi$ plus one.

By the above connection to $\chi^2$-divergence, we have the following proposition.

**Proposition 13.** *For any protocol $\pi$, we have*

$$\mathop{\mathbb{E}}_{\pi}\left[\theta_\mu(\pi \mathbin{@} X, Y, \mathbf{M})^{-1}\right] = 1.$$

*Proof.* Let $\tau$ be the protocol by making $\pi$ standard as described in the remark above. Then we have

$$\theta_\mu(\pi \mathbin{@} X, Y, \mathbf{M}) = \frac{\pi(X, Y, \mathbf{M})}{\tau(X, Y, \mathbf{M})}.$$

Therefore,

$$\mathop{\mathbb{E}}_{\pi}\left[\theta_\mu(\pi \mathbin{@} X, Y, \mathbf{M})^{-1}\right] = \sum_{(X,Y,\mathbf{M})} \pi(X, Y, \mathbf{M}) \cdot \frac{\tau(X, Y, \mathbf{M})}{\pi(X, Y, \mathbf{M})} = 1.$$

□

By standard bounds on conditional probabilities, we have the following bound on the conditional cost.

**Proposition 14.** *For events $W_1, W_2$, we have*

$$\theta_\mu(\pi \mid W_1 \cap W_2) \le \frac{\theta_\mu(\pi \mid W_1)}{\pi(W_2 \mid W_1)}.$$

*Proof.* Since $\theta_\mu(\pi \mathbin{@} X, Y, \mathbf{M})$ is nonnegative, we have

$$\begin{aligned}
\theta_\mu(\pi \mid W_1 \cap W_2) &= \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi|W_1\cap W_2}\left[\theta_\mu(\pi \mathbin{@} X, Y, \mathbf{M})\right] \\
&= \sum_{(X,Y,\mathbf{M})} \pi(X, Y, \mathbf{M} \mid W_1 \cap W_2) \cdot \theta_\mu(\pi \mathbin{@} X, Y, \mathbf{M}) \\
&\le \sum_{(X,Y,\mathbf{M})} \frac{\pi(X, Y, \mathbf{M} \mid W_1)}{\pi(W_2 \mid W_1)} \cdot \theta_\mu(\pi \mathbin{@} X, Y, \mathbf{M}) \\
&= \frac{\theta_\mu(\pi \mid W_1)}{\pi(W_2 \mid W_1)}.
\end{aligned}$$

□

Next, the $\chi^2$-*cost* measures the "communication cost" of a protocol: how different Alice's input becomes in Bob's view at the end of the communication compared to that in the input distribution.

**Definition 15** ($\chi^2$-cost)**.** Let $\pi$ be a generalized protocol, and $\mu$ be a distribution over the inputs. The $\chi^2$-cost of $\pi$ by Alice with respect to $\mu$ at $(X, Y, \mathbf{M})$ is

$$\chi^2_{\mu,A}(\pi \mathbin{@} X, Y, \mathbf{M}) := \frac{\pi(X \mid \mathbf{M}, Y)}{\mu(X \mid Y)};$$

the $\chi^2$-cost of $\pi$ by Bob with respect to $\mu$ at $(X, Y, \mathbf{M})$ is

$$\chi^2_{\mu,B}(\pi \mathbin{@} X, Y, \mathbf{M}) := \frac{\pi(Y \mid \mathbf{M}, X)}{\mu(Y \mid X)}.$$

The $\chi^2$-costs of $\pi$ with respect to $\mu$ are

$$\chi^2_{\mu,A}(\pi) := \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi}\left[\chi^2_{\mu,A}(\pi \mathbin{@} X,Y,\mathbf{M})\right],$$
$$\chi^2_{\mu,B}(\pi) := \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi}\left[\chi^2_{\mu,B}(\pi \mathbin{@} X,Y,\mathbf{M})\right].$$

For an event $W$, the $\chi^2$-costs of $\pi$ with respect to $\mu$ conditioned on $W$ are

$$\chi^2_{\mu,A}(\pi \mid W) := \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi|W}\left[\chi^2_{\mu,A}(\pi \mathbin{@} X,Y,\mathbf{M})\right],$$
$$\chi^2_{\mu,B}(\pi \mid W) := \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi|W}\left[\chi^2_{\mu,B}(\pi \mathbin{@} X,Y,\mathbf{M})\right].$$

*Remark.* Similar to the $\theta$-cost, $\chi^2_{\mu,A}(\pi \mid W)$ is also different from $\chi^2_{\mu,A}(\pi_W)$. The $\chi^2$-cost of $\pi$ by Alice is the expected $\chi^2$-divergence from $\mu_{X|Y}$ to $\pi_{X|Y,\mathbf{M}}$ plus one. Similarly, the $\chi^2$-cost of $\pi$ by Bob is the expected $\chi^2$-divergence from $\mu_{Y|X}$ to $\pi_{Y|X,\mathbf{M}}$ plus one. Observe that for standard protocols, if we measure the expected KL-divergence instead of the $\chi^2$-divergence, then we obtain the internal information costs:

$$\sum_{\text{odd } i} I(X; M_i \mid Y, M_{<i}) \qquad \text{and} \qquad \sum_{\text{even } i} I(Y; M_i \mid X, M_{<i}).$$

Similar proofs to Proposition 13 and Proposition 14 give us the following two propositions.

**Proposition 16.** *For any protocol $\pi$, we have*

$$\mathop{\mathbb{E}}_{\pi}\left[\chi^2_{\mu,A}(\pi \mathbin{@} X,Y,\mathbf{M})^{-1}\right] = 1,$$

*and*

$$\mathop{\mathbb{E}}_{\pi}\left[\chi^2_{\mu,B}(\pi \mathbin{@} X,Y,\mathbf{M})^{-1}\right] = 1.$$

**Proposition 17.** *For any events $W_1, W_2$, we have*

$$\chi^2_{\mu,A}(\pi \mid W_1 \cap W_2) \le \frac{\chi^2_{\mu,A}(\pi \mid W_1)}{\pi(W_2 \mid W_1)},$$

*and*

$$\chi^2_{\mu,B}(\pi \mid W_1 \cap W_2) \le \frac{\chi^2_{\mu,B}(\pi \mid W_1)}{\pi(W_2 \mid W_1)}.$$

### 3.4 Rectangle properties in generalized protocols

We will maintain the *rectangle property* for the generalized protocols throughout the proof.

**Definition 18** (Rectangle property)**.** A generalized protocol $\pi$ has the *rectangle property* with respect to $\mu$, if there exists nonnegative functions $g_1 : \mathcal{X} \times \mathcal{M} \to \mathbb{R}, g_2 : \mathcal{Y} \times \mathcal{M} \to \mathbb{R}$ such that

$$\pi(X,Y,\mathbf{M}) = \mu(X,Y) \cdot g_1(X,\mathbf{M}) \cdot g_2(Y,\mathbf{M}).$$

Let $W$ be an event, $(\pi \mid W)$ has the rectangle property with respect to $\mu$ if there exists nonnegative functions $g_1 : \mathcal{X} \times \mathcal{M} \to \mathbb{R}, g_2 : \mathcal{Y} \times \mathcal{M} \to \mathbb{R}$ such that

$$\pi(X,Y,\mathbf{M} \mid W) = \mu(X,Y) \cdot g_1(X,\mathbf{M}) \cdot g_2(Y,\mathbf{M}).$$

Equivalently, $\pi$ has the rectangle property if for every transcript $\mathbf{M}$, the posterior distribution $\pi_{X,Y|\mathbf{M}}$ is equal to $\mu$ rescaled by some product function with one factor depending only on $X$ and another factor depending only on $Y$. Note that this property holds for any standard protocol, since each message $M_i$ conditioned on $M_{<i}$ only depends on one of $X$ and $Y$. Hence, for standard protocols, we have such product structure even conditioned on any prefix $M_{<i}$.

When decomposing a protocol for $k$ instances, we need the following definition of the *partial rectangle property*.

**Definition 19** (Partial rectangle property)**.** Let $\pi$ be a generalized protocol such that $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$. $\pi$ satisfies the *partial rectangle property* with respect to $\mu^k$ if there exists nonnegative functions $g_1, g_2, g_3$ such that

$$\pi(X, Y, \mathbf{M}) = \mu^k(X, Y) \cdot g_1(X, \mathbf{M}) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}).$$

Let $W$ be an event, $(\pi \mid W)$ has the partial rectangle property with respect to $\mu^k$ if there exists nonnegative functions $g_1, g_2, g_3$ such that

$$\pi(X, Y, \mathbf{M} \mid W) = \mu^k(X, Y) \cdot g_1(X, \mathbf{M}) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}).$$

**Proposition 20.** *If $\pi$ has the partial rectangle property, then $X_{<k}$ and $Y_k$ are independent conditioned on $X_k, Y_{<k}, \mathbf{M}$; If $\pi \mid W$ has the partial rectangle property, then $X_{<k}$ and $Y_k$ are independent conditioned on $X_k, Y_{<k}, \mathbf{M}, W$.*

*Proof.* If $\pi$ has the partial rectangle property, then

$$\begin{aligned}
&\pi(X_{<k}, Y_k \mid X_k, Y_{<k}, \mathbf{M}) \\
&\quad = \mu^k(X, Y) \cdot g_1(X, \mathbf{M}) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}) \cdot \pi(X_k, Y_{<k}, \mathbf{M})^{-1} \\
&\quad = \left(\mu^{k-1}(X_{<k}, Y_{<k}) \cdot g_1(X, \mathbf{M})\right) \cdot \left(\mu(X_k, Y_k) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}) \cdot \pi(X_k, Y_{<k}, \mathbf{M})^{-1}\right).
\end{aligned}$$

Note that given $(X_k, Y_{<k}, \mathbf{M})$, the first factor only depends on $X_{<k}$, and the second factor only depends on $Y_k$. By normalizing the two factors, we obtain that

$$\pi(X_{<k}, Y_k \mid X_k, Y_{<k}, \mathbf{M}) = \pi(X_{<k} \mid X_k, Y_{<k}, \mathbf{M}) \cdot \pi(Y_k \mid X_k, Y_{<k}, \mathbf{M}).$$

The proof for $\pi \mid W$ is almost identical. We omit the details. □

For a protocol $\pi$, we define the follow sets that are related to the rectangle property and the partial rectangle property.

**Definition 21.** Let $\mathcal{U}_{X,M}(\pi)$ be the set consisting of all possible pairs $(X, \mathbf{M})$. Let $\mathcal{U}_{Y,M}(\pi)$ be the set consisting of all possible pairs $(Y, \mathbf{M})$. Let $\mathcal{U}_{X_k,Y_{<k},M}(\pi)$ be the set consisting of all possible triples $(X_k, Y_{<k}, \mathbf{M})$.

Let $\mathcal{S}_{\mathrm{rec}}(\pi)$ be the collection of all possible events $S$ such that there exist $S_{X,M} \subseteq \mathcal{U}_{X,M}, S_{Y,M} \subseteq \mathcal{U}_{Y,M}$, and

$$S = \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \in S_{X,M} \wedge (Y, \mathbf{M}) \in S_{Y,M}\}.$$

Let $\mathcal{S}_{\mathrm{pt}}(\pi)$ be the collection of all possible events $S$ such that there exist $S_{X,M} \subseteq \mathcal{U}_{X,M}, S_{Y,M} \subseteq \mathcal{U}_{Y,M}, S_{X_k,Y_{<k},M} \subseteq \mathcal{U}_{X_k,Y_{<k},M}$, and

$$S = \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \in S_{X,M} \wedge (Y, \mathbf{M}) \in S_{Y,M} \wedge (X_k, Y_{<k}, \mathbf{M}) \in S_{X_k,Y_{<k},M}\}.$$

We may omit $\pi$ and use $\mathcal{S}_{\mathrm{rec}}, \mathcal{S}_{\mathrm{pt}}$ when there is no ambiguity.

Intuitively, $\mathcal{S}_{\mathrm{rec}}(\pi)$ is the collection of events conditioned on which, $\pi$ remains to have the rectangle property. Similarly, $\mathcal{S}_{\mathrm{pt}}(\pi)$ is the collection of events conditioned on which, $\pi$ remains to have the partial rectangle property.

**Proposition 22.** *We have the following properties about $\mathcal{S}_{\mathrm{rec}}$ and $\mathcal{S}_{\mathrm{pt}}$:*

*(i) if $\pi$ has the rectangle property, then for any $S \in \mathcal{S}_{\mathrm{rec}}$, $(\pi \mid S)$ has the rectangle property;*

*(ii) if $\pi$ has the partial rectangle property, then for any $S \in \mathcal{S}_{\mathrm{pt}}$, $(\pi \mid S)$ has the partial rectangle property;*

*(iii) $\mathcal{S}_{\mathrm{rec}} \subseteq \mathcal{S}_{\mathrm{pt}}$;*

*(iv) both $\mathcal{S}_{\mathrm{rec}}$ and $\mathcal{S}_{\mathrm{pt}}$ are closed under intersection.*

*Proof.* For (i), suppose $S = \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \in S_{X,M} \wedge (Y, \mathbf{M}) \in S_{Y,M}\}$. Then

$$\pi(X, Y, \mathbf{M} \mid S) = \pi(X, Y, \mathbf{M}) \cdot \mathbf{1}_{S_{X,M}}(X, \mathbf{M}) \cdot \mathbf{1}_{S_{Y,M}}(Y, \mathbf{M}) \cdot \pi(S)^{-1}.$$

Thus, if $\pi$ has the rectangle property, then $(\pi \mid S)$ has the rectangle property.

Similarly for (ii), suppose

$$S = \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \in S_{X,M} \wedge (Y, \mathbf{M}) \in S_{Y,M} \wedge (X_k, Y_{<k}, \mathbf{M}) \in S_{X_k, Y_{<k}, M}\}.$$

Then

$$\pi(X, Y, \mathbf{M} \mid S) = \pi(X, Y, \mathbf{M}) \cdot \mathbf{1}_{S_{X,M}}(X, \mathbf{M}) \cdot \mathbf{1}_{S_{Y,M}}(Y, \mathbf{M}) \cdot \mathbf{1}_{S_{X_k, Y_{<k}, M}}(X_k, Y_{<k}, \mathbf{M}) \cdot \pi(S)^{-1}.$$

Thus, if $\pi$ has the partial rectangle property, then $(\pi \mid S)$ has the partial rectangle property.

(iii) and (iv) follow from the definitions. □

# 4 Main Setup

In this section, we set up the main framework for proving our main theorems.

**Theorem 1 (restated).** *For any $\{0,1\}$-valued function $f$, we have*

$$\mathbf{R}^{(r)}_{1/2+2^{-n}}(f^{\oplus n}) \geq n \cdot \left(r^{-O(r)} \cdot \mathbf{R}^{(r)}_{2/3}(f) - 1\right).$$

**Theorem 2 (restated).** *Let $c > 0$ be a sufficiently large constant. Fix $\alpha \in (0, r^{-cr})$ and $C_A, C_B \geq 2c \log(r/\alpha)$. Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, and $\mu$ be a distribution over $\mathcal{X} \times \mathcal{Y}$. Suppose $f$ satisfies*

$$\mathrm{suc}_{\mu}(f; C_A, C_B, r) \leq 1/2 + \alpha/2,$$

*then for any integer $n \geq 2$, we have*

$$\mathrm{suc}_{\mu^n}(f^{\oplus n}; 2^{-8} r^{-1} n \cdot C_A, 2^{-8} r^{-1} n \cdot C_B, r) \leq \frac{1}{2} + \frac{\alpha^{2^{-12} n}}{2}.$$

We first show that Theorem 2 implies Theorem 1.

*Proof of Theorem 1.* Fix a function $f$, suppose there is an $r$-round protocol $\pi$ for $f^{\oplus n}$ with communication cost $T$ and success probability $1/2 + 2^{-n}$. Let $\alpha = r^{-2cr}$ for a sufficiently large $c$, then $\pi$ has success probability more than $1/2 + \alpha^{2^{-12}n}/2$. By setting $C_A = C_B = 2^8 r \cdot T/n + 2c \log(r/\alpha) = O(r \cdot T/n + r \log r)$, Theorem 2 implies $\mathrm{suc}_\mu(f; C_A, C_B, r) > 1/2 + \alpha/2$, i.e., for distribution $\mu$, there is an $r$-round protocol with communication cost at most $O(r \cdot T/n + r \log r)$ in each round and success probability at least $1/2 + r^{-O(r)}$.

Since this holds for *any* $\mu$, by Yao's minimax lemma, there is an $r$-round randomized protocol with $O(r \cdot T/n + r \log r)$ communication in each round and success probability at least $1/2 + r^{-O(r)}$ for all inputs. By simply running such a protocol $r^{O(r)}$ times in parallel and outputting the majority, we obtain an $r$-round protocol with $r^{O(r)} \cdot (T/n + 1)$ total communication and success probability $2/3$. Thus, we obtain $\mathbf{R}^{(r)}_{2/3}(f) \le r^{O(r)} \cdot (\mathbf{R}_{1/2+2^{-n}}(f^{\oplus n})/n + 1)$. Rearranging the terms gives Theorem 1. □

In the rest of the paper, we will focus on proving Theorem 2. Let us fix a sufficiently large constant $c > 0$, parameters $C_A, C_B, r, \alpha$, function $f$ and input distribution $\mu$ satisfying its premises. As mentioned in Section 2, we will first define a potential function based on the costs and advantage, and then show that the potential function value decreases as we decrement $n$.

**Definition 23** (Potential functions)**.** For an $r$-round generalized protocol $\pi$ for $f^{\oplus n}$ and an event $W$, we define the potential function $\phi_n(\pi \mid W)$ (and $\phi_n^{\mathrm{cost}}, \phi_n^{\mathrm{adv}}$) as follows:

$$\phi_n(\pi \mid W) = \underbrace{\log \theta_{\mu^n}(\pi \mid W) + \frac{\log(1/\alpha)}{C_A - c\log(r/\alpha)} \cdot \log \chi^2_{\mu,A}(\pi \mid W) + \frac{\log(1/\alpha)}{C_B - c\log(r/\alpha)} \cdot \log \chi^2_{\mu,B}(\pi \mid W)}_{\phi_n^{\mathrm{cost}}(\pi|W)}$$
$$+ \underbrace{32 \log \left( \mathop{\mathbb{E}}_{\pi|W} \left[ \mathrm{adv}_\pi(f^{\oplus n}(X,Y) \mid \mathbf{M}, W) \right]^{-1} \right)}_{\phi_n^{\mathrm{adv}}(\pi|W)}.$$

We also define $\phi_{n,\mathrm{pt}}(\pi \mid W)$ (and $\phi^{\mathrm{adv}}_{n,\mathrm{pt}}$) as follows:

$$\phi_{n,\mathrm{pt}}(\pi \mid W) = \log \theta_{\mu^n}(\pi \mid W) + \frac{\log(1/\alpha)}{C_A - c\log(r/\alpha)} \cdot \log \chi^2_{\mu,A}(\pi \mid W) + \frac{\log(1/\alpha)}{C_B - c\log(r/\alpha)} \cdot \log \chi^2_{\mu,B}(\pi \mid W)$$
$$+ \underbrace{32 \log \left( \mathop{\mathbb{E}}_{\pi|W} \left[ \mathrm{adv}_\pi(f^{\oplus n}(X,Y) \mid X_n, Y_{<n}, \mathbf{M}, W) \right]^{-1} \right)}_{\phi^{\mathrm{adv}}_{n,\mathrm{pt}}(\pi|W)}.$$

When $W$ is the whole sample space, we may simply write $\phi_n(\pi)$ or $\phi_{n,\mathrm{pt}}(\pi)$.

The first three terms in both potential functions $\phi_n^{\mathrm{cost}}$ are the (normalized) costs of $\pi$. They are small if $\pi$ has low $\theta$-cost and low $\chi^2$-costs. The last term in both potential functions depends on the expected advantage. $\phi_n$ uses the standard advantage, while $\phi_{n,\mathrm{pt}}$ uses the advantage conditioned not only on the transcript, but also $X_n$ and $Y_{<n}$. As we will see later, it is used when decomposing $\pi$. The last term is small if the protocol has high advantage. By Proposition 11, knowing more could only increase the expected advantage. Hence, $\phi_{n,\mathrm{pt}}(\pi)$ is always at most $\phi_n(\pi)$.

We have the following lower bound on the potential of $\pi$ conditioned on $W$. In particular, when $W$ is the whole sample space, the potential function is nonnegative.

**Lemma 24.** *For any $\pi$, event $W$ and any $n \geq 1$, we must have*

$$\phi_n(\pi \mid W) \geq -3\log(1/\pi(W)).$$

*Proof.* For the $\theta$-cost, by the convexity of $1/x$, we have

$$\begin{aligned}\theta_{\mu^n}(\pi \mid W)^{-1} &= \mathop{\mathbb{E}}_{\pi|W}\left[\theta_{\mu^n}(\pi \mathbin{@} X, Y, \mathbf{M})\right]^{-1} \\ &\leq \mathop{\mathbb{E}}_{\pi|W}\left[\theta_{\mu^n}(\pi \mathbin{@} X, Y, \mathbf{M})^{-1}\right]\end{aligned}$$

which by the fact that $\theta_{\mu^n}(\pi \mathbin{@} X, Y, \mathbf{M})$ is nonnegative, is

$$\leq \pi(W)^{-1} \cdot \mathop{\mathbb{E}}_{\pi}\left[\theta_{\mu^n}(\pi \mathbin{@} X, Y, \mathbf{M})^{-1}\right]$$

which by Proposition 13, is

$$= \pi(W)^{-1}.$$

Hence, $\log \theta_{\mu^n}(\pi \mid W) \geq -\log(1/\pi(W))$. Similarly, we also have $\log \chi^2_{\mu^n,A}(\pi \mid W) \geq -\log(1/\pi(W))$, and $\log \chi^2_{\mu^n,B}(\pi \mid W) \geq -\log(1/\pi(W))$. By the fact that $\log(1/\alpha) \leq C_A - c\log(r/\alpha)$ and $\log(1/\alpha) \leq C_B - c\log(r/\alpha)$, we have

$$\phi^{\mathrm{adv}}_n(\pi \mid W) \geq -3\log(1/\pi(W)).$$

The advantage is always at most 1. Therefore, the last term is nonnegative. Hence, $\phi_n(\pi \mid W) \geq -3\log(1/\pi(W))$. □

The following lemma shows an upper bound on the potential of a *deterministic* standard protocol $\pi$ computing $f^{\oplus n}$.

**Lemma 25.** *Let $\pi$ be a* deterministic *standard protocol where Alice sends at most $T_A$ bits in each (odd) round and Bob sends at most $T_B$ bits in each (even) round. If it computes $f^{\oplus n}$ with probability $\frac{1}{2} + \frac{\sigma}{2}$ under input distribution $\mu^n$, then*

$$\phi_n(\pi) \leq \lceil r/2 \rceil \cdot \frac{T_A \cdot \log(1/\alpha)}{C_A - c\log(r/\alpha)} + \lfloor r/2 \rfloor \cdot \frac{T_B \cdot \log(1/\alpha)}{C_B - c\log(r/\alpha)} + 32\log(1/\sigma).$$

*Proof.* By the property of a standard protocol, $\theta_{\mu^n}(\pi \mathbin{@} X, Y, \mathbf{M}) = 1$ for any $X, Y, \mathbf{M}$ in the support of $\pi$. Hence, $\log \theta_{\mu^n}(\pi) = 0$.

For the $\chi^2$-cost by Alice, we have

$$\begin{aligned}\chi^2_{\mu^n,A}(\pi) &= \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi}\left[\frac{\pi(X \mid \mathbf{M}, Y)}{\mu^n(X \mid Y)}\right] \\ &= \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi}\left[\frac{\pi(X \mid \mathbf{M}, Y)}{\pi(X \mid Y)}\right] \\ &= \sum_{(X,Y,\mathbf{M})} \frac{\pi(X, Y, \mathbf{M})\pi(X \mid \mathbf{M}, Y)}{\pi(X \mid Y)}\end{aligned}$$

$$= \sum_{(X,Y,\mathbf{M})} \pi(\mathbf{M} \mid X, Y)\pi(X \mid Y, \mathbf{M})\pi(Y).$$

Since $\pi$ is a *deterministic* standard protocol, $M_0$ is fixed. All even messages $(M_2, M_4, \ldots)$ are sent by Bob such that each $M_i$ is determined by $M_{<i}$ and $Y$. Therefore, $(M_2, M_4, \ldots)$ are determined by all odd messages $(M_1, M_3, \ldots)$ and $Y$. Denote $(M_2, M_4, \ldots)$ by $M_{\text{even}}$ and $(M_1, M_3, \ldots)$ by $M_{\text{odd}}$, we have

$$\begin{aligned}
&\sum_{(X,Y,\mathbf{M})} \pi(\mathbf{M} \mid X, Y)\pi(X \mid Y, \mathbf{M})\pi(Y) \\
&= \sum_{(X,Y,\mathbf{M})} \pi(M_{\text{even}} \mid X, Y, M_{\text{odd}})\pi(M_{\text{odd}} \mid X, Y)\pi(X \mid Y, M_{\text{odd}}, M_{\text{even}})\pi(Y) \\
&= \sum_{(X,Y,\mathbf{M})} \pi(M_{\text{even}} \mid X, Y, M_{\text{odd}})\pi(M_{\text{odd}} \mid X, Y)\pi(X \mid Y, M_{\text{odd}})\pi(Y) \\
&= \sum_{(X,Y,\mathbf{M})} \pi(M_{\text{odd}} \mid X, Y)\pi(X, M_{\text{even}} \mid Y, M_{\text{odd}})\pi(Y) \\
&\le \sum_{(X,Y,\mathbf{M})} \pi(X, M_{\text{even}} \mid Y, M_{\text{odd}})\pi(Y) \\
&= \sum_{(Y,M_{\text{odd}})} \pi(Y) \\
&= \sum_{M_{\text{odd}}} 1 \\
&\le 2^{\lceil r/2 \rceil T_A},
\end{aligned}$$

where the last inequality uses the fact that Alice's messages have at most $T_A$ bits in each (odd) round. Similarly, we have $\chi^2_\pi(\mu, B) \le 2^{\lfloor r/2 \rfloor T_B}$.

Finally, by the connection between advantage and success probability, $\mathbb{E}_\pi\left[\mathrm{adv}_\pi(f^{\oplus n}(X, Y) \mid \mathbf{M})\right] \ge \sigma$. Hence,

$$\phi_n(\pi) \le \lceil r/2 \rceil \cdot \frac{T_A \cdot \log(1/\alpha)}{C_A - c\log(r/\alpha)} + \lfloor r/2 \rfloor \cdot \frac{T_B \cdot \log(1/\alpha)}{C_B - c\log(r/\alpha)} + 32\log(1/\sigma).$$

□

In the rest of the paper, we will prove the following lemma, which shows that given a protocol for $f^{\oplus k}$, we can construct a protocol for $f^{\oplus k-1}$ with a lower potential.

**Lemma 26.** *For $k \ge 2$,*

***if** there is a generalized protocol $\pi$ for $f^{\oplus k}$ with the rectangle property with respect to $\mu^k$ and an event $V \in \mathcal{S}_{\text{rec}}(\pi)$ such that $\pi(V) \ge 2^{-12}$,*

***then** there is a generalized protocol $\pi_{\text{new}}$ for $f^{\oplus k-1}$ with the rectangle property with respect to $\mu^{k-1}$ and an event $V_{\text{new}} \in \mathcal{S}_{\text{rec}}(\pi_{\text{new}})$ such that $\pi_{\text{new}}(V_{\text{new}}) \ge 2^{-12}$, and*

$$\phi_{k-1}(\pi_{\text{new}} \mid V_{\text{new}}) \le \phi_k(\pi \mid V) - \frac{1}{16}\log(1/\alpha).$$

Our main theorem is a direct corollary of Lemma 24, 25 and 26.

*Proof of Theorem 2.* Since $C_A, C_B \geq 2c\log(r/\alpha)$, $C_A/2 \leq C_A - c\log(r/\alpha)$ and $C_B/2 \leq C_B - c\log(r/\alpha)$. Suppose there exists an $r$-round protocol $\pi^{(n)}$ where Alice sends at most

$$2^{-8}r^{-1}n \cdot C_A \leq 2^{-7}r^{-1}n(C_A - c\log(r/\alpha))$$

bits in each round and Bob sends at most

$$2^{-8}r^{-1}n \cdot C_B \leq 2^{-7}r^{-1}n(C_B - c\log(r/\alpha))$$

in each round, which computes $f^{\oplus n}$ correctly with probability $1/2 + \sigma/2$ when the input is sampled from $\mu^n$. By fixing the randomness, we may assume that $\pi^{(n)}$ is deterministic. Then by Lemma 25, we have

$$\phi_n(\pi^{(n)}) \leq 2^{-7}n\log(1/\alpha) + 32\log(1/\sigma).$$

Now we set $V^{(n)}$ to be the whole sample space of $\pi^{(n)}$. Clearly, $\pi^{(n)}$ and $V^{(n)}$ satisfy the premise of Lemma 26. By inductively applying Lemma 26 a total of $n-1$ times, we obtain a protocol $\pi^{(1)}$ for $f$ and event $V^{(1)}$ such that $\pi^{(1)}(V^{(1)}) \geq 2^{-12}$ and

$$\phi_1(\pi^{(1)} \mid V^{(1)}) \leq \phi_n(\pi^{(n)}) - \frac{n-1}{16} \cdot \log(1/\alpha).$$

On the other hand, Lemma 24 implies that the LHS is at least $-3\log(1/\pi^{(1)}(V^{(1)})) \geq -36$, implying that

$$\phi_n(\pi^{(n)}) \geq \frac{n-1}{16} \cdot \log(1/\alpha) - 36 \geq 2^{-6}n\log(1/\alpha),$$

since $n \geq 2$ and $\alpha < r^{-cr}$ for a sufficiently large $c$.

Combining the above upper and lower bounds on $\phi_n(\pi^{(n)})$, we obtain

$$2^{-7}n\log(1/\alpha) + 32\log(1/\sigma) \geq 2^{-6}n\log(1/\alpha),$$

implying that $\log(1/\sigma) \geq 2^{-12}n\log(1/\alpha)$, i.e.,

$$\sigma \leq \alpha^{2^{-12}n}.$$

This proves the theorem. □

# 5 Decomposition of Generalized Protocols

To prove Lemma 26, we will decompose a generalized protocol $\pi$ for $f^{\oplus k}$ into a protocol $\pi_{<k}$ for $f^{\oplus k-1}$ and a protocol $\pi_k$ for $f$ such that the costs of $\pi_{<k}$ and $\pi_k$ "add up" to the costs of $\pi$ pointwisely. For simplicity of notations, we will assume that $r$ *is even* from now on, the case of odd $r$ is similar.

## 5.1 Definition of $\pi_{<k}$ and $\pi_k$

Fix a generalized protocol $\pi$ with the rectangle property with respect to $\mu^k$. Let $(X, Y, \mathbf{M}) \sim \pi$. We view the following tuple as the $r$-round generalized protocol $\pi_{<k}$ on inputs $(X_{<k}, Y_{<k})$

$$(X_{<k}, Y_{<k}, (M_0 \circ X_k, M_1, M_2, \ldots, M_{r-1}, M_r)),$$

where we append $X_k$ to $M_0$. We view the following tuple as the $r$-round generalized protocol $\pi_k$ on inputs $(X_k, Y_k)$

$$(X_k, Y_k, (M_0, Y_{<k} \circ M_1, M_2, \ldots, M_{r-1}, M_r)),$$

where we prepend $Y_{<k}$ to $M_1$.

It is useful to think that $\pi_{<k}$, $\pi_k$ and $\pi$ are the *same* distribution over the *same* sample space, only their inputs and transcripts are defined in different ways. Therefore, we may use $\pi_{<k}(W), \pi_k(W), \pi(W)$ interchangeably when measuring the probability of an event $W$.

For simplicity of notations, we use $\mathbf{M}^{(\pi_{<k})}$ to denote $(M_0 \circ X_k, M_1, M_2, \ldots, M_r)$, the transcript of $\pi_{<k}$, and use $\mathbf{M}^{(\pi_k)}$ to denote $(M_0, Y_{<k} \circ M_1, M_2, \ldots, M_r)$, the transcript of $\pi_k$. $M_i^{(\pi_{<k})}$ and $M_i^{(\pi_k)}$ are defined similarly. Since $(X, Y, \mathbf{M})$ determines $(X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})})$, we define $\theta$-cost of $\pi_{<k}$ at $(X, Y, \mathbf{M})$ as

$$\theta_{\mu^{k-1}}(\pi_{<k} @ X, Y, \mathbf{M}) := \theta_{\mu^{k-1}}(\pi_{<k} @ X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})}),$$

where $(X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})})$ is the triple determined by $(X, Y, \mathbf{M})$. Note that this cost does not depend on $Y_k$ given the other parts of $(X, Y, \mathbf{M})$. The $\chi^2$-costs of $\pi_{<k}$ and the costs of $\pi_k$ at $(X, Y, \mathbf{M})$ are defined similarly.

In the remainder of this section, we will analyze $\pi_{<k}$ and $\pi_k$. First, we observe that the partial rectangle property of $\pi$ implies the rectangle properties of $\pi_{<k}$ and $\pi_k$.

**Proposition 27.** *Let $W$ be an event such that $\pi \mid W$ has the partial rectangle property with respect to $\mu^k$. Then $\pi_{<k} \mid W$ has the rectangle property with respect to $\mu^{k-1}$, and $\pi_k \mid W$ has the rectangle property with respect to $\mu$.*

*Proof.* Since $\pi \mid W$ has the partial rectangle property, there exists $g_1, g_2, g_3$ such that

$$\pi(X, Y, \mathbf{M} \mid W) = \mu^k(X, Y) \cdot g_1(X, \mathbf{M}) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}).$$

Thus,

$$\begin{aligned}
\pi_{<k}(X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})} \mid W) &= \pi(X, Y_{<k}, \mathbf{M} \mid W) \\
&= \sum_{Y_k} \mu^k(X, Y) \cdot g_1(X, \mathbf{M}) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}) \\
&= \mu^{k-1}(X_{<k}, Y_{<k}) \cdot g_1(X, \mathbf{M}) \cdot \left( \sum_{Y_k} \mu(X_k, Y_k) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}) \right).
\end{aligned}$$

Note that the second factor is a function of only $X_{<k}$ and $\mathbf{M}^{(\pi_{<k})}$, the third factor is a function of only $Y_{<k}$ and $\mathbf{M}^{(\pi_{<k})}$.

For $\pi_k$, we have

$$\begin{aligned}
\pi_k(X_k, Y_k, \mathbf{M}^{(\pi_k)} \mid W) &= \pi(X_k, Y, \mathbf{M} \mid W) \\
&= \sum_{X_{<k}} \mu^k(X, Y) \cdot g_1(X, \mathbf{M}) \cdot g_2(Y, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}) \\
&= \mu(X_k, Y_k) \cdot \left( \sum_{X_{<k}} g_1(X, \mathbf{M}) \cdot g_3(X_k, Y_{<k}, \mathbf{M}) \right) \cdot g_2(Y, \mathbf{M}).
\end{aligned}$$

The second factor depends only on $X_k$ and $\mathbf{M}^{(\pi_k)}$, and the third factor depends only on $Y_k$ and $\mathbf{M}^{(\pi_k)}$. This proves the lemma. □

Similarly, we have the following relation between $\mathcal{S}_{\mathrm{rec}}(\pi_{<k}), \mathcal{S}_{\mathrm{rec}}(\pi_k)$ and $\mathcal{S}_{\mathrm{pt}}(\pi)$.

**Proposition 28.** *We have $\mathcal{S}_{\mathrm{rec}}(\pi_{<k}) \subseteq \mathcal{S}_{\mathrm{pt}}(\pi)$ and $\mathcal{S}_{\mathrm{rec}}(\pi_k) \subseteq \mathcal{S}_{\mathrm{pt}}(\pi)$.*

*Proof.* Let $S \in \mathcal{S}_{\mathrm{rec}}(\pi_{<k})$ be an event such that

$$S = \{(X, Y, \mathbf{M}) : (X_{<k}, \mathbf{M}^{(\pi_{<k})}) \in S_{X,M} \wedge (Y_{<k}, \mathbf{M}^{(\pi_{<k})}) \in S_{Y,M}\},$$

for sets $S_{X,M} \in \mathcal{U}_{X,M}(\pi_{<k})$ and $S_{Y,M} \in \mathcal{U}_{Y,M}(\pi_{<k})$. Hence,

$$S = \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \in S_{X,M} \wedge (X_k, Y_{<k}, \mathbf{M}) \in S_{Y,M}\}$$

is a set in $\mathcal{S}_{\mathrm{pt}}(\pi)$.

The proof of $\mathcal{S}_{\mathrm{rec}}(\pi_k) \subseteq \mathcal{S}_{\mathrm{pt}}(\pi)$ is similar, and we omit the details. □

## 5.2 Decomposition of the costs

Below is the first main lemma of the decomposition, stating that the product of $\theta$-costs of $\pi_{<k}$ and $\pi_k$ is equal to that of $\pi$ pointwisely.

**Lemma 29.** *The product of the $\theta$-costs of $\pi_{<k}$ and $\pi_k$ at $(X, Y, \mathbf{M})$ is $\theta$-cost of $\pi$ at $(X, Y, \mathbf{M})$,*

$$\theta_{\mu^{k-1}}(\pi_{<k} @ X, Y, \mathbf{M}) \cdot \theta_{\mu}(\pi_k @ X, Y, \mathbf{M}) = \theta_{\mu^k}(\pi @ X, Y, \mathbf{M}).$$

*Proof.* By definition, we have

$$\begin{aligned}
&\theta_{\mu^{k-1}}(\pi_{<k} @ X, Y, \mathbf{M}) \\
&= \frac{\pi_{<k}(X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})} \mid M_0^{(\pi_{<k})})}{\mu^{k-1}(X_{<k}, Y_{<k})} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi_{<k}(M_i^{(\pi_{<k})} \mid X_{<k}, M_{<i}^{(\pi_{<k})})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi_{<k}(M_i^{(\pi_{<k})} \mid Y_{<k}, M_{<i}^{(\pi_{<k})})} \\
&= \frac{\pi(X, Y_{<k}, \mathbf{M} \mid X_k, M_0)}{\mu^{k-1}(X_{<k}, Y_{<k})} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid X_k, Y_{<k}, M_{<i})} \\
&= \frac{\pi(X_{<k}, Y_{<k}, \mathbf{M} \mid X_k, M_0)}{\mu^{k-1}(X_{<k}, Y_{<k})} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid X_k, Y_{<k}, M_{<i})}. \qquad (2)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\theta_{\mu}(\pi_k @ X, Y, \mathbf{M}) \\
&= \frac{\pi_k(X_k, Y_k, \mathbf{M}^{(\pi_k)} \mid M_0^{(\pi_k)})}{\mu(X_k, Y_k)} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi_k(M_i^{(\pi_k)} \mid X_k, M_{<i}^{(\pi_k)})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi_k(M_i^{(\pi_k)} \mid Y_k, M_{<i}^{(\pi_k)})} \\
&= \frac{\pi(X_k, Y, \mathbf{M} \mid M_0)}{\mu(X_k, Y_k)} \cdot \frac{1}{\pi(Y_{<k}, M_1 \mid X_k, M_0)} \cdot \prod_{\text{odd } i \in [3,r]} \frac{1}{\pi(M_i \mid X_k, Y_{<k}, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid Y, M_{<i})} \\
&= \frac{\pi(X_k, Y, \mathbf{M} \mid M_0)}{\mu(X_k, Y_k) \cdot \pi(Y_{<k} \mid X_k, M_0)} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X_k, Y_{<k}, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid Y, M_{<i})}. \qquad (3)
\end{aligned}$$

Combining Equation (2) and (3), we have

$$
\begin{aligned}
&\theta_{\mu^{k-1}}(\pi_{<k} \circledcirc X, Y, \mathbf{M}) \cdot \theta_{\mu}(\pi_k \circledcirc X, Y, \mathbf{M}) \\
&= \frac{\pi(X_{<k}, Y_{<k}, \mathbf{M} \mid X_k, M_0)}{\mu^{k-1}(X_{<k}, Y_{<k})} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid X_k, Y_{<k}, M_{<i})} \\
&\qquad \cdot \frac{\pi(X_k, Y, \mathbf{M} \mid M_0)}{\mu(X_k, Y_k) \cdot \pi(Y_{<k} \mid X_k, M_0)} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X_k, Y_{<k}, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid Y, M_{<i})} \\
&= \frac{\pi(X_{<k}, Y_{<k}, \mathbf{M} \mid X_k, M_0)\pi(X_k, Y, \mathbf{M} \mid M_0)}{\mu^k(X, Y)\pi(Y_{<k} \mid X_k, M_0)} \cdot \frac{1}{\pi(\mathbf{M} \mid X_k, Y_{<k}, M_0)} \\
&\qquad \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid Y, M_i)} \\
&= \frac{\pi(X_{<k} \mid X_k, Y_{<k}, \mathbf{M})\pi(X_k, Y, \mathbf{M} \mid M_0)}{\mu^k(X, Y)} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid Y, M_{<i})}.
\end{aligned}
$$

Then by the rectangle property of $\pi$ and Proposition 20, $Y_k$ is independent of $X_{<k}$ conditioned on $(X_k, Y_{<k}, \mathbf{M})$. It is equal to

$$
\begin{aligned}
&\frac{\pi(X_{<k} \mid X_k, Y, \mathbf{M})\pi(X_k, Y, \mathbf{M} \mid M_0)}{\mu^k(X, Y)} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid Y, M_{<i})} \\
&\quad = \frac{\pi(X, Y, \mathbf{M} \mid M_0)}{\mu^k(X, Y)} \cdot \prod_{\text{odd } i \in [r]} \frac{1}{\pi(M_i \mid X, M_{<i})} \cdot \prod_{\text{even } i \in [r]} \frac{1}{\pi(M_i \mid Y, M_{<i})} \\
&\quad = \theta_{\mu^k}(\pi \circledcirc X, Y, \mathbf{M}).
\end{aligned}
$$

This proves the lemma. □

The second main lemma of the decomposition states that the product of the $\chi^2$-costs of $\pi_{<k}$ and $\pi_k$ is also equal to that of $\pi$ pointwisely.

**Lemma 30.** *The product of the $\chi^2$-costs of $\pi_{<k}$ and $\pi_k$ at $(X, Y, \mathbf{M})$ is the $\chi^2$-cost of $\pi$ at $(X, Y, \mathbf{M})$ by Alice and Bob respectively,*

$$
\begin{aligned}
\chi^2_{\mu^{k-1}, A}(\pi_{<k} \circledcirc X, Y, \mathbf{M}) \cdot \chi^2_{\mu, A}(\pi_k \circledcirc X, Y, \mathbf{M}) &= \chi^2_{\mu^k, A}(\pi \circledcirc X, Y, \mathbf{M}), \\
\chi^2_{\mu^{k-1}, B}(\pi_{<k} \circledcirc X, Y, \mathbf{M}) \cdot \chi^2_{\mu, B}(\pi_k \circledcirc X, Y, \mathbf{M}) &= \chi^2_{\mu^k, B}(\pi \circledcirc X, Y, \mathbf{M}).
\end{aligned}
$$

*Proof.* For the $\chi^2$-cost by Alice, by definition, we have

$$
\begin{aligned}
\chi^2_{\mu^{k-1}, A}(\pi_{<k} \circledcirc X, Y, \mathbf{M}) &= \frac{\pi_{<k}(X_{<k} \mid Y_{<k}, \mathbf{M}^{(\pi_{<k})})}{\mu^{k-1}(X_{<k} \mid Y_{<k})} \\
&= \frac{\pi(X_{<k} \mid X_k, Y_{<k}, \mathbf{M})}{\mu^{k-1}(X_{<k} \mid Y_{<k})},
\end{aligned}
$$

and

$$
\chi^2_{\mu, A}(\pi_k \circledcirc X, Y, \mathbf{M}) = \frac{\pi_k(X_k \mid Y_k, \mathbf{M}^{(\pi_k)})}{\mu(X_k \mid Y_k)}
$$

$$= \frac{\pi(X_k \mid Y, \mathbf{M})}{\mu(X_k \mid Y_k)}.$$

Hence, by partial rectangle property of $\pi$ and Proposition 20, their product is equal to

$$\begin{aligned}
&\chi^2_{\mu^{k-1},A}(\pi_{<k} \mathbin{@} X, Y, \mathbf{M}) \cdot \chi^2_{\mu,A}(\pi_k \mathbin{@} X, Y, \mathbf{M}) \\
&= \frac{\pi(X_{<k} \mid X_k, Y_{<k}, \mathbf{M})}{\mu^{k-1}(X_{<k} \mid Y_{<k})} \cdot \frac{\pi(X_k \mid Y, \mathbf{M})}{\mu(X_k \mid Y_k)} \\
&= \frac{\pi(X_{<k} \mid X_k, Y, \mathbf{M}) \cdot \pi(X_k \mid Y, \mathbf{M})}{\mu^k(X \mid Y)} \\
&= \frac{\pi(X \mid Y, \mathbf{M})}{\mu^k(X \mid Y)} \\
&= \chi^2_{\mu^k,A}(\pi \mathbin{@} X, Y, \mathbf{M}).
\end{aligned}$$

The $\chi^2$-cost for Bob is similar,

$$\begin{aligned}
&\chi^2_{\mu^{k-1},B}(\pi_{<k} \mathbin{@} X, Y, \mathbf{M}) \cdot \chi^2_{\mu,B}(\pi_k \mathbin{@} X, Y, \mathbf{M}) \\
&= \frac{\pi(Y_{<k} \mid X, \mathbf{M})}{\mu^{k-1}(Y_{<k} \mid X_{<k})} \cdot \frac{\pi(Y_k \mid X_k, Y_{<k}, \mathbf{M})}{\mu(Y_k \mid X_k)} \\
&= \frac{\pi(Y_{<k} \mid X, \mathbf{M}) \cdot \pi(Y_k \mid X, Y_{<k}, \mathbf{M})}{\mu^k(Y \mid X)} \\
&= \chi^2_{\mu^k,B}(\pi \mathbin{@} X, Y, \mathbf{M}).
\end{aligned}$$

This proves the lemma. □

# 6 Induction: Proof of Lemma 26

In this section, we will use the decomposition of $\pi$ to prove Lemma 26.

## 6.1 Identify event $U$

As we mentioned in Section 2, to obtain a new protocol for $f^{\oplus k-1}$ from $\pi$, we first identify an event $U$ such that the advantage of $\pi$ is not concentrated on any $S$ for $S \in \mathcal{S}_{\mathrm{pt}}(\pi)$ and $S \subseteq U$.

Let $U \in \mathcal{S}_{\mathrm{pt}}(\pi)$ and $U \subseteq V$ be an event that maximizes

$$\pi(U)^{1/2} \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right]. \tag{4}$$

Since $\mathcal{S}_{\mathrm{pt}}(\pi)$ is a discrete set, such $U$ exists. If there is a tie, we fix $U$ to be any maximizer. We first show that conditioning on $U$ reduces the potential function value, and the reduction is large when the probability $U$ is small.

**Lemma 31.** *$\pi \mid U$ has the* partial rectangle property *with respect to $\mu^k$, and*

$$\phi_{k,\mathrm{pt}}(\pi \mid U) \le \phi_k(\pi \mid V) - 13 \log(1/\pi(U \mid V)).$$

*Proof.* By definition, we have

$$\phi_{k,\mathrm{pt}}(\pi \mid U) = \log \theta_{\mu^k}(\pi \mid U) + \frac{\log(1/\alpha)}{C_A - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^k,A}(\pi \mid U) + \frac{\log(1/\alpha)}{C_B - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^k,B}(\pi \mid U) + 32 \log \left( \mathop{\mathbb{E}}_{\pi|U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U) \right]^{-1} \right).$$

By Proposition 14, Proposition 17 and the fact that $U \subseteq V$, we have

$$\begin{aligned} \log \theta_{\mu^k}(\pi \mid U) &\le \log \theta_{\mu^k}(\pi \mid V) + \log(1/\pi(U \mid V)) \\ \log \chi^2_{\mu^k,A}(\pi \mid U) &\le \log \chi^2_{\mu^k,A}(\pi \mid V) + \log(1/\pi(U \mid V)) \\ \log \chi^2_{\mu^k,B}(\pi \mid U) &\le \log \chi^2_{\mu^k,B}(\pi \mid V) + \log(1/\pi(U \mid V)). \end{aligned}$$

Then since $U$ is the maximizer of Equation (4) and $V \in \mathcal{S}_{\mathrm{rec}}(\pi) \subseteq \mathcal{S}_{\mathrm{pt}}(\pi)$,

$$\begin{aligned} &\mathop{\mathbb{E}}_{\pi|U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U) \right] \\ &\quad \ge \pi(U)^{-1/2} \cdot \pi(V)^{1/2} \cdot \mathop{\mathbb{E}}_{\pi|V} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, V) \right] \\ &\quad = \pi(U \mid V)^{-1/2} \cdot \mathop{\mathbb{E}}_{\pi|V} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, V) \right]. \end{aligned}$$

Since knowing less could only decrease the advantage (Proposition 11),

$$\mathop{\mathbb{E}}_{\pi|V} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, V) \right] \ge \mathop{\mathbb{E}}_{\pi|V} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid \mathbf{M}, V) \right].$$

Combining the inequalities and using the fact that $\log(1/\alpha) < C_A - c\log(r/a)$ and $\log(1/\alpha) < C_B - c\log(r/a)$, we have

$$\begin{aligned} \phi_{k,\mathrm{pt}}(\pi \mid U) &\le \log \theta_{\mu^k}(\pi \mid V) + \frac{\log(1/\alpha)}{C_A - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^k,A}(\pi \mid V) + \frac{\log(1/\alpha)}{C_B - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^k,B}(\pi \mid V) \\ &\quad + 32 \log \left( \mathop{\mathbb{E}}_{\pi|V} \left[ f^{\oplus k}(X,Y) \mid \mathbf{M}, V \right]^{-1} \right) + 3\log(1/\pi(U \mid V)) - 16\log(1/\pi(U \mid V)) \\ &= \phi_k(\pi \mid V) - 13\log(1/\pi(U \mid V)). \end{aligned}$$

This proves the lemma. □

We need the following proposition in the later proof.

**Proposition 32.** *We have the following:*

*(i) for any $S \in \mathcal{S}_{\mathrm{pt}}(\pi)$ and $S \subseteq U$, we have*

$$\begin{aligned} &\pi(S) \cdot \mathop{\mathbb{E}}_{\pi|S} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S) \right] \\ &\qquad \le \pi(S \mid U)^{1/2} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U) \right]. \end{aligned}$$

*(ii) for any $S \in \mathcal{S}_{\mathrm{pt}}(\pi)$ and $S \subseteq U$, if*

$$\pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right] \le s \cdot \pi(S) \cdot \mathop{\mathbb{E}}_{\pi|S}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S)\right],$$

*for some $s \ge 1$, then for any $t \le 32$, we have*

$$\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U) + t\log(1/\pi(U)) \ge \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid S) + t\log(1/\pi(S)) - 32\log s.$$

*Proof.* (i) Since $U$ is the maximizer of Equation (4), $S \in \mathcal{S}_{\mathrm{pt}}(\pi)$ and $S \subseteq U$, we have

$$\begin{aligned}
&\pi(S) \cdot \mathop{\mathbb{E}}_{\pi|S}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S)\right] \\
&\le \pi(S)^{1/2} \cdot \pi(U)^{1/2} \cdot \mathop{\mathbb{E}}_{\pi|U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right] \\
&= \pi(S \mid U)^{1/2} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right].
\end{aligned}$$

(ii) By taking the logarithm on both sides of the premise, we have

$$\begin{aligned}
&\log\left(\mathop{\mathbb{E}}_{\pi|U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right]^{-1}\right) + \log(1/\pi(U)) \\
&\ge \log\left(\mathop{\mathbb{E}}_{\pi|S}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S)\right]^{-1}\right) + \log(1/\pi(S)) - \log s,
\end{aligned}$$

i.e., (recall Definition 23)

$$\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U) + 32\log(1/\pi(U)) \ge \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid S) + 32\log(1/\pi(S)) - 32\log s.$$

Since $\pi(U) \ge \pi(S)$, for any $t \le 32$,

$$\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U) + t\log(1/\pi(U)) \ge \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid S) + t\log(1/\pi(S)) - 32\log s.$$

□

Now we will divide the set of all $(X_k, Y_{<k}, \mathbf{M})$ with nonzero probability under $\pi$ into subsets based on the costs and the advantages of $\pi_{<k}$ and $\pi_k$. Then we show that for each subset, there is a way to construct a generalized protocol for $f^{\oplus k-1}$ such that at least one of the protocols satisfies the requirements of Lemma 26. To analyze the costs of these protocols, which we will construct later in this section, we need the following two lemmas.

**Lemma 33.** *Fix a set $S$ of triples $(X_k, Y_{<k}, \mathbf{M})$ and a parameter $\eta > 0$. If for all $(X_k, Y_{<k}, \mathbf{M}) \in S$,*

$$\mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U}\left[\theta_\mu(\pi_k @ X_k, Y_k, \mathbf{M}^{(\pi_k)})\right] \ge \eta,$$

*then we have*

$$\log \theta_{\mu^{k-1}}(\pi_{<k} \mid S \cap U) \le \log \theta_{\mu^k}(\pi \mid U) + \log(1/\pi(S \mid U)) - \log \eta,$$

*where we abused the notation to let $S$ also denote the set $\{(X, Y, \mathbf{M}) : (X_k, Y_{<k}, \mathbf{M}) \in S\}$.*

*Proof.* By Lemma 29, we have

$$\begin{aligned}
&\theta_{\mu^k}(\pi \mid S \cap U) \\
&= \mathop{\mathbb{E}}_{\pi}\left[\theta_{\mu^k}(\pi \mathbin{@} X, Y, \mathbf{M}) \mid S \cap U\right] \\
&= \mathop{\mathbb{E}}_{\pi}\left[\theta_{\mu^{k-1}}(\pi_{<k} \mathbin{@} X, Y, \mathbf{M}) \cdot \theta_{\mu}(\pi_k \mathbin{@} X, Y, \mathbf{M}) \mid S \cap U\right].
\end{aligned}$$

By the construction of $\pi_{<k}$ and $\pi_k$, $\theta_{\mu^{k-1}}(\pi_{<k} \mathbin{@} X, Y, \mathbf{M})$ is a function of $(X, Y_{<k}, \mathbf{M})$ and does not depend on $Y_k$, and $\theta_{\mu}(\pi_k \mathbin{@} X_k, Y, \mathbf{M})$ is a function of $(X_k, Y, \mathbf{M})$ does not depend on $X_{<k}$. Thus, it is equal to

$$\mathop{\mathbb{E}}_{\pi}\left[\theta_{\mu^{k-1}}(\pi_{<k} \mathbin{@} X, Y_{<k}, \mathbf{M}) \cdot \theta_{\mu}(\pi_k \mathbin{@} X_k, Y, \mathbf{M}) \mid S \cap U\right].$$

Since $\mathcal{S}_{\mathrm{pt}}$ is closed under intersection and $S \in \mathcal{S}_{\mathrm{pt}}$ by definition, we have that $S \cap U \in \mathcal{S}_{\mathrm{pt}}$. Hence, $\pi \mid S \cap U$ has the partial rectangle property by Proposition 22(ii). Then $X_{<k}$ and $Y_k$ are independent conditioned on $(X_k, Y_{<k}, \mathbf{M}, S \cap U)$ by Proposition 20. Hence, it is equal to

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi \mid S \cap U}\left[\mathop{\mathbb{E}}_{X_{<k} \sim \pi \mid X_k, Y_{<k}, \mathbf{M}, S \cap U}[\theta_{\mu^{k-1}}(\pi_{<k} \mathbin{@} X, Y_{<k}, \mathbf{M})] \cdot \mathop{\mathbb{E}}_{Y_k \sim \pi \mid X_k, Y_{<k}, \mathbf{M}, S \cap U}[\theta_{\mu}(\pi_k \mathbin{@} X_k, Y, \mathbf{M})]\right] \\
&= \mathop{\mathbb{E}}_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi \mid S \cap U}\left[\mathop{\mathbb{E}}_{X_{<k} \sim \pi \mid X_k, Y_{<k}, \mathbf{M}, S \cap U}[\theta_{\mu^{k-1}}(\pi_{<k} \mathbin{@} X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})})]\right. \\
&\qquad\qquad \left.\times \mathop{\mathbb{E}}_{Y_k \sim \pi \mid X_k, Y_{<k}, \mathbf{M}, S \cap U}[\theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})]\right].
\end{aligned}$$

Since $S$ is a set of triples $(X_k, Y_{<k}, \mathbf{M})$, $(\pi \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)$ is the same as $(\pi \mid X_k, Y_{<k}, \mathbf{M}, U)$ (for $(X_k, Y_{<k}, \mathbf{M}) \in S$). It is equal to

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi \mid S \cap U}\left[\mathop{\mathbb{E}}_{X_{<k} \sim \pi \mid X_k, Y_{<k}, \mathbf{M}, S \cap U}[\theta_{\mu^{k-1}}(\pi_{<k} \mathbin{@} X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})})]\right. \\
&\qquad\qquad \left.\times \mathop{\mathbb{E}}_{Y_k \sim \pi \mid X_k, Y_{<k}, \mathbf{M}, U}[\theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})]\right] \\
&\geq \mathop{\mathbb{E}}_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi \mid S \cap U}\left[\mathop{\mathbb{E}}_{X_{<k} \sim \pi \mid X_k, Y_{<k}, \mathbf{M}, S \cap U}[\theta_{\mu^{k-1}}(\pi_{<k} \mathbin{@} X_{<k}, Y_{<k}, \mathbf{M}^{(\pi_{<k})})] \cdot \eta\right] \\
&= \theta_{\mu^{k-1}}(\pi_{<k} \mid S \cap U) \cdot \eta.
\end{aligned}$$

Finally, by Proposition 14,

$$\theta_{\mu^k}(\pi \mid S \cap U) \leq \frac{\theta_{\mu^k}(\pi \mid U)}{\pi(S \mid U)}.$$

Hence, we have

$$\log \theta_{\mu^{k-1}}(\pi_{<k} \mid S \cap U) \leq \log \theta_{\mu^k}(\pi \mid U) + \log(1/\pi(S \mid U)) - \log \eta.$$

This proves the lemma. □

Next, by applying Lemma 30 and Proposition 17 instead of Lemma 29 and Proposition 14, the same proof gives the following lemma for the $\chi^2$-costs.

**Lemma 34.** *Fix a set $S$ of triples $(X_k, Y_{<k}, \mathbf{M})$ and a parameter $\eta > 0$. If for all $(X_k, Y_{<k}, \mathbf{M}) \in S$,*

$$\mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \chi^2_{\mu, A}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] \geq \eta,$$

*then we have*

$$\log \chi^2_{\mu^{k-1}, A}(\pi_{<k} \mid S \cap U) \leq \log \chi^2_{\mu^k, A}(\pi \mid U) + \log(1/\pi(S \mid U)) - \log \eta;$$

*similarly, if for all $(X_k, Y_{<k}, \mathbf{M}) \in S$,*

$$\mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \chi^2_{\mu, B}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] \geq \eta,$$

*then we have*

$$\log \chi^2_{\mu^{k-1}, B}(\pi_{<k} \mid S \cap U) \leq \log \chi^2_{\mu^k, B}(\pi \mid U) + \log(1/\pi(S \mid U)) - \log \eta.$$

We will also need the following lemma to relate the advantage for $f^{\oplus k-1}$ to the advantage for $f^{\oplus k}$.

**Lemma 35.** *Fix a set $S \in \mathcal{S}_{\mathrm{pt}}(\pi)$ such that $\pi(S \cap U) > 0$. Suppose there exists $b \in \{0, 1\}$ such that for any $(X_k, Y_{<k}, \mathbf{M})$ with $\pi(X_k, Y_{<k}, \mathbf{M}, S \cap U) > 0$, we have*

$$\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \geq 1/2.$$

*Then we have*

$$\mathop{\mathbb{E}}_{\pi | S \cap U} \left[ \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, S \cap U) \right] \geq \mathop{\mathbb{E}}_{\pi | S \cap U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \right].$$

*Moreover, if we further have*

$$\mathop{\mathbb{E}}_{\pi | S \cap U} \left[ \mathrm{adv}_\pi(f(X_k, Y_k) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \right] \leq \eta,$$

*for*

$$\eta^{1/4} \leq \frac{1}{2} \cdot \frac{\pi(S \cap U)^{1/2} \cdot \mathbb{E}_{\pi | S \cap U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \right]}{\pi(U)^{1/2} \cdot \mathbb{E}_{\pi | U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, U) \right]},$$

*then we have*

$$\begin{aligned} &\mathop{\mathbb{E}}_{\pi | S \cap U} \left[ \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, S \cap U) \right] \\ &\quad \geq \frac{1}{2} \eta^{-1/2} \cdot \mathop{\mathbb{E}}_{\pi | S \cap U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \right]. \end{aligned}$$

The first condition $\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \geq 1/2$ is used to ensure that the expected advantage conditioned on $(\mathbf{M}^{(\pi_{<k})}, S \cap U)$ is the same as the expected advantage conditioned on $(Y_{<k}, \mathbf{M}^{(\pi_{<k})}, S \cap U)$.

*Proof.* We have

$$\begin{aligned}
&\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, S \cap U) \\
&= \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, \mathbf{M}, S \cap U) \\
&= \left|2\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, \mathbf{M}, S \cap U) - 1\right| \\
&= \left|2 \sum_{Y_{<k}} \pi(Y_{<k} \mid X_k, \mathbf{M}, S \cap U) \cdot \pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) - 1\right| \\
&= \left|\sum_{Y_{<k}} \pi(Y_{<k} \mid X_k, \mathbf{M}, S \cap U) \cdot (2\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) - 1)\right|.
\end{aligned}$$

By the assumption that $\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \geq 1/2$, the absolute value of the sum is equal to the sum of absolute values:

$$\begin{aligned}
&\left|\sum_{Y_{<k}} \pi(Y_{<k} \mid X_k, \mathbf{M}, S \cap U) \cdot (2\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) - 1)\right| \\
&\quad = \sum_{Y_{<k}} \pi(Y_{<k} \mid X_k, \mathbf{M}, S \cap U) \cdot \left|2\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) - 1\right| \\
&\quad = \mathop{\mathbb{E}}_{Y_{<k} \sim \pi \mid X_k, \mathbf{M}, S \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\right].
\end{aligned}$$

By taking the expectation over $(X_k, \mathbf{M})$ conditioned on $S \cap U$, we obtain

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\pi \mid S \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, S \cap U)\right] \\
&\qquad = \mathop{\mathbb{E}}_{\pi \mid S \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\right]. \qquad (5)
\end{aligned}$$

Since $S, U \in \mathcal{S}_{\mathrm{pt}}(\pi)$, we have $S \cap U \in \mathcal{S}_{\mathrm{pt}}(\pi)$. Therefore, $X_{<k}$ and $Y_k$ are independent conditioned on $(X_k, Y_{<k}, \mathbf{M}, S \cap U)$ by Proposition 22(ii) and Proposition 20. In particular, $f^{\oplus k-1}(X_{<k}, Y_{<k})$ and $f(X_k, Y_k)$ are independent conditioned on $(X_k, Y_{<k}, \mathbf{M}, S \cap U)$. By the fact that $f(X, Y) = f^{\oplus k-1}(X_{<k}, Y_{<k}) \oplus f(X_k, Y_k)$ and Proposition 5, we have

$$\begin{aligned}
&\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \\
&\quad = \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \cdot \mathrm{adv}_\pi(f(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \qquad (6) \\
&\quad \leq \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U).
\end{aligned}$$

Thus, the expected advantage is at least

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\pi \mid S \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, S \cap U)\right] \\
&\qquad \geq \mathop{\mathbb{E}}_{\pi \mid S \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\right].
\end{aligned}$$

This proves the first part of the lemma.

For the second part, let $T$ be the set of all triples $(X_k, Y_{<k}, \mathbf{M})$ such that

$$\pi(X_k, Y_{<k}, \mathbf{M}, S \cap U) > 0$$

and

$$\mathrm{adv}_\pi(f(X_k, Y_k) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) \geq \eta^{1/2}.$$

Then by Markov's inequality, if we have

$$\mathop{\mathbb{E}}_{\pi|S\cap U}[\mathrm{adv}_\pi(f(X_k, Y_k) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)] \leq \eta,$$

then $\pi(T \mid S \cap U) \leq \eta^{1/2}$.

Hence, for $(X_k, Y_{<k}, \mathbf{M}) \notin T$ and $\pi(X_k, Y_{<k}, \mathbf{M}, S \cap U) > 0$, Equation (6) implies that

$$\begin{aligned}&\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\\ &\quad\leq \eta^{1/2} \cdot \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U).\end{aligned}$$

Hence, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{\pi|S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\right]\\ &\quad\geq \sum_{(X_k, Y_{<k}, \mathbf{M})\notin T} \pi(X_k, Y_{<k}, \mathbf{M} \mid S \cap U) \cdot \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\\ &\quad\geq \eta^{-1/2} \cdot \sum_{(X_k, Y_{<k}, \mathbf{M})\notin T} \pi(X_k, Y_{<k}, \mathbf{M} \mid S \cap U) \cdot \mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\\ &\quad\geq \eta^{-1/2} \cdot \mathop{\mathbb{E}}_{\pi|S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\right]\\ &\qquad- \eta^{-1/2} \cdot \pi(T \mid S \cap U) \cdot \mathop{\mathbb{E}}_{\pi|T\cap S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\right].\end{aligned}$$

Next we show that (the absolute value of) the second term is at most half of the first term. First since $T$ is a set of $(X_k, Y_{<k}, \mathbf{M})$, we have

$$\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U) = \mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, T \cap S \cap U)$$

for any $(X_k, Y_{<k}, \mathbf{M}) \in T$. Hence, by the fact that $T \cap S \cap U \in \mathcal{S}_{\mathrm{pt}}(\pi)$ and $U$ maximizes Equation (4), we have that

$$\begin{aligned}&\pi(T \mid S \cap U) \cdot \mathop{\mathbb{E}}_{\pi|T\cap S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, S \cap U)\right]\\ &= \pi(S \cap U)^{-1/2} \cdot \pi(T \mid S \cap U)^{1/2} \cdot \left(\pi(T \cap S \cap U)^{1/2} \cdot \mathop{\mathbb{E}}_{\pi|T\cap S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, T \cap S \cap U)\right]\right)\\ &\leq \pi(S \cap U)^{-1/2} \cdot \eta^{1/4} \cdot \left(\pi(U)^{1/2} \cdot \mathop{\mathbb{E}}_{\pi|U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right]\right),\end{aligned}$$

which by the bound on $\eta$, is at most

$$\frac{1}{2} \cdot \mathop{\mathbb{E}}_{\pi|S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S\cap U)\right].$$

Thus, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{\pi|S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k},Y_{<k}) \mid X_k, Y_{<k}, \mathbf{M}, S\cap U)\right]\\ &\qquad\geq \frac{1}{2}\eta^{-1/2} \cdot \mathop{\mathbb{E}}_{\pi|S\cap U}\left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S\cap U)\right].\end{aligned}$$

Combining it with Equation (5) proves the lemma. □

## 6.2 High costs

We first consider all $(X_k, Y_{<k}, \mathbf{M})$ at which $\pi_k$ has high costs. We will show that it leads to significant lower $\phi^{\mathrm{cost}}_{k-1}$.

**High $\theta$-cost.** The first set of triples consists of all $(X_k, Y_{<k}, \mathbf{M})$ such that

$$\begin{aligned}\alpha^{-1/2} &\leq \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\theta_\mu(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right],\\ 2^{-2^{-5}(C_A - c\log(r/\alpha))}\cdot\pi(U) &\leq \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\chi^2_{\mu,A}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right],\\ 2^{-2^{-5}(C_B - c\log(r/\alpha))}\cdot\pi(U) &\leq \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\chi^2_{\mu,B}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right].\end{aligned}$$

This is the set of triples at which $\pi_k$ has high $\theta$-cost and not-too-low $\chi^2$-costs.

Note that $\theta_\mu(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})$, $\chi^2_{\mu,A}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})$ and $\chi^2_{\mu,B}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})$ are functions of $(X_k, Y, \mathbf{M})$, hence, they do *not* depend on $X_{<k}$. Denote this set of $(X_k, Y_{<k}, \mathbf{M})$ by $S_{\text{high-}\theta}$. We will also abuse the notation, and use $S_{\text{high-}\theta}$ to denote the set $\{(X, Y, \mathbf{M}) : (X_k, Y_{<k}, \mathbf{M}) \in S_{\text{high-}\theta}\}$, which can also be treated as an event.

By applying Lemma 33 and Lemma 34 to $S_{\text{high-}\theta}$ and the corresponding $\eta$, we obtain the following bounds on the costs of $\pi_{<k}$ conditioned on $S_{\text{high-}\theta\cap U}$:

$$\begin{aligned}\log\theta_{\mu^{k-1}}(\pi_{<k} \mid S_{\text{high-}\theta}\cap U) &\leq \log\theta_{\mu^k}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\theta} \mid U)) - \frac{1}{2}\log(1/\alpha),\\ \log\chi^2_{\mu^{k-1},A}(\pi_{<k} \mid S_{\text{high-}\theta}\cap U) &\leq \log\chi^2_{\mu^k,A}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\theta} \mid U))\\ &\qquad + 2^{-5}(C_A - c\log(r/\alpha)) + \log(1/\pi(U)),\\ \log\chi^2_{\mu^{k-1},B}(\pi_{<k} \mid S_{\text{high-}\theta}\cap U) &\leq \log\chi^2_{\mu^k,B}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\theta} \mid U))\\ &\qquad + 2^{-5}(C_B - c\log(r/\alpha)) + \log(1/\pi(U)).\end{aligned}$$

Thus, it implies that (recall Definition 23)

$$\phi^{\mathrm{cost}}_{k-1}(\pi_{<k} \mid S_{\text{high-}\theta}\cap U) \leq \phi^{\mathrm{cost}}_k(\pi \mid U) + 3\log(1/\pi(S_{\text{high-}\theta} \mid U)) + 2\log(1/\pi(U)) - \frac{1}{4}\log(1/\alpha), \quad (7)$$

where we used the assumption that $C_A - c\log(r/\alpha) > \log(1/\alpha), C_B - c\log(r/\alpha) > \log(1/\alpha)$.

**High $\chi^2$-cost by Alice.** The next set consists of all $(X_k, Y_{<k}, \mathbf{M})$ such that

$$\begin{aligned}
\alpha^{2^{-5}} \cdot \pi(U) &\le \mathop{\mathbb{E}}_{Y_k|X_k,Y_{<k},\mathbf{M}\sim\pi|U}\left[\theta_\mu(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right] < \alpha^{-1/2},\\
2^{C_A - c\log(r/\alpha)} &\le \mathop{\mathbb{E}}_{Y_k|X_k,Y_{<k},\mathbf{M}\sim\pi|U}\left[\chi^2_{\mu,A}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right],\\
2^{-2^{-5}(C_B - c\log(r/\alpha))} \cdot \pi(U) &\le \mathop{\mathbb{E}}_{Y_k|X_k,Y_{<k},\mathbf{M}\sim\pi|U}\left[\chi^2_{\mu,B}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right].
\end{aligned}$$

This is the set of triples at which $\pi_k$ has high $\chi^2$-cost by Alice and not-too-low $\theta$-cost and $\chi^2$-cost by Bob. Denote this set of $(X_k, Y_{<k}, \mathbf{M})$ by $S_{\text{high-}\chi^2\text{-A}}$. The upper bound on the $\theta$-cost ensures that it is disjoint from $S_{\text{high-}\theta}$. Similarly, we also abuse the notation to let $S_{\text{high-}\chi^2\text{-A}}$ also denote the set $\{(X, Y, \mathbf{M}) : (X_k, Y_{<k}, \mathbf{M}) \in S_{\text{high-}\chi^2\text{-A}}\}$.

By applying Lemma 33 and Lemma 34 to $S_{\text{high-}\chi^2\text{-A}}$ and the corresponding $\eta$, we have the following bounds:

$$\begin{aligned}
\log \theta_{\mu^{k-1}}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-A}} \cap U) &\le \log \theta_{\mu^k}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\chi^2\text{-A}} \mid U))\\
&\quad + 2^{-5}\log(1/\alpha) + \log(1/\pi(U)),\\
\log \chi^2_{\mu^{k-1},A}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-A}} \cap U) &\le \log \chi^2_{\mu^k,A}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\chi^2\text{-A}} \mid U))\\
&\quad - (C_A - c\log(r/\alpha)),\\
\log \chi^2_{\mu^{k-1},B}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-A}} \cap U) &\le \log \chi^2_{\mu^k,B}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\chi^2\text{-A}} \mid U))\\
&\quad + 2^{-5}(C_B - c\log(r/\alpha)) + \log(1/\pi(U)),
\end{aligned}$$

which also implies

$$\phi^{\text{cost}}_{k-1}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-A}} \cap U) \le \phi^{\text{cost}}_k(\pi \mid U) + 3\log(1/\pi(S_{\text{high-}\chi^2\text{-A}} \mid U)) + 2\log(1/\pi(U)) - \frac{1}{4}\log(1/\alpha). \quad (8)$$

**High $\chi^2$-cost by Bob.** The third case consists of all $(X_k, Y_{<k}, \mathbf{M})$ such that

$$\begin{aligned}
\alpha^{2^{-5}} \cdot \pi(U) &\le \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\theta_\mu(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right] < \alpha^{-1/2},\\
2^{-2^{-5}(C_A - c\log(r/\alpha))} \cdot \pi(U) &\le \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\chi^2_{\mu,A}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right] < 2^{C_A - c\cdot\log(r/\alpha)},\\
2^{C_B - c\log(r/\alpha)} &\le \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\chi^2_{\mu,B}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})\right].
\end{aligned}$$

This is the set of triples at which $\pi_k$ has high $\chi^2$-cost by Bob and not-too-low $\theta$-cost and $\chi^2$-cost by Alice. Denote this set of $(X_k, Y_{<k}, \mathbf{M})$ by $S_{\text{high-}\chi^2\text{-B}}$. It is disjoint from $S_{\text{high-}\theta}$ and $S_{\text{high-}\chi^2\text{-A}}$. Similarly, we also use $S_{\text{high-}\chi^2\text{-B}}$ to denote the set $\{(X, Y, \mathbf{M}) : (X_k, Y_{<k}, \mathbf{M}) \in S_{\text{high-}\chi^2\text{-B}}\}$.

By applying Lemma 33 and Lemma 34 to $S_{\text{high-}\chi^2\text{-B}}$ and the appropriate $\eta$, we have the following bounds:

$$\begin{aligned}
\log \theta_{\mu^{k-1}}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-B}} \cap U) &\le \log \theta_{\mu^k}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\chi^2\text{-B}} \mid U))\\
&\quad + 2^{-5}\log(1/\alpha) + \log(1/\pi(U)),\\
\log \chi^2_{\mu^{k-1},A}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-B}} \cap U) &\le \log \chi^2_{\mu^k,A}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\chi^2\text{-B}} \mid U))
\end{aligned}$$

$$+ 2^{-5}(C_A - c\log(r/\alpha)) + \log(1/\pi(U)),$$
$$\log \chi^2_{\mu^{k-1},B}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-B}} \cap U) \le \log \chi^2_{\mu^k,B}(\pi \mid U) + \log(1/\pi(S_{\text{high-}\chi^2\text{-B}} \mid U))$$
$$- (C_B - c\log(r/\alpha)),$$

which also implies that

$$\phi^{\text{cost}}_{k-1}(\pi_{<k} \mid S_{\text{high-}\chi^2\text{-B}} \cap U) \le \phi^{\text{cost}}_k(\pi \mid U) + 3\log(1/\pi(S_{\text{high-}\chi^2\text{-B}} \mid U)) + 2\log(1/\pi(U)) - \frac{1}{4}\log(1/\alpha). \quad (9)$$

Equation (7), (8) and (9) implies that for $\beta \in \{\text{high-}\theta, \text{high-}\chi^2\text{-A}, \text{high-}\chi^2\text{-B}\}$, we all have

$$\begin{aligned}\phi^{\text{cost}}_{k-1}(\pi_{<k} \mid S_\beta \cap U) &\le \phi^{\text{cost}}_k(\pi \mid U) + 3\log(1/\pi(S_\beta \mid U)) + 2\log(1/\pi(U)) - \frac{1}{4}\log(1/\alpha) \\ &\le \phi^{\text{cost}}_k(\pi \mid U) + 3\log(1/\pi(S_\beta \cap U)) - \frac{1}{4}\log(1/\alpha). \end{aligned} \quad (10)$$

The main lemma of this subsection is the following, stating that if the above three sets contribute a nontrival amount of total advantage in $U$ (weighted by the probability), then we can construct a protocol for $f^{\oplus k-1}$ satisfying the requirements of Lemma 26 (by conditioning $\pi_{<k}$ on a carefully chosen event).

**Lemma 36.** *Let $S_{\text{high-cost}}$ be the union $S_{\text{high-}\theta} \cup S_{\text{high-}\chi^2\text{-A}} \cup S_{\text{high-}\chi^2\text{-B}}$. If we have*

$$\begin{aligned}&\pi(S_{\text{high-cost}} \cap U) \cdot \mathop{\mathbb{E}}_{\pi|S_{\text{high-cost}} \cap U} \left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{high-cost}} \cap U)\right] \\ &\ge \frac{1}{3} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right],\end{aligned}$$

*then Lemma 26 holds.*

Protocol $\pi_{<k}$ and event $S_\beta \cap U$ may be one potential choice for $\pi_{\text{new}}$ and $V_{\text{new}}$ in Lemma 26. However, Lemma 26 requires the probability of $V_{\text{new}}$ to be $\Omega(1)$, which is not necessarily true for any $\beta$, since $U$ may have very small probability. On the other hand, we could also consider setting $\pi_{\text{new}}$ to the distribution of $\pi_{<k}$ conditioned on $S_\beta \cap U$ and $V_{\text{new}}$ to the entire sample space, but this protocol may have very large costs.

To resolve this issue, we will use the following lemma, which turns $(\pi_{<k} \mid S_\beta \cap U)$ into a protocol $(\pi_{<k})_G$ with bounded costs for some event $G \approx S_\beta \cap U$. Moreover, by dividing $G$ into $G_0 \cup G_1$ according to whether the function value is more likely to be 0 or 1 conditioned on $Y$ and $\mathbf{M}$, the lemma guarantees that the costs conditioned on $G_b$ are also bounded (for $b = 0, 1$). This will allow us to apply Lemma 35 later to lower bound the advantage.

**Lemma 37.** *Fix any $\gamma \in (0, 1/2)$. Let $\rho$ be an $r$-round generalized protocol over $\mathcal{X} \times \mathcal{Y} \times \mathcal{M}$, $W$ be an event, $\nu$ be an input distribution and $h : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function of the inputs. Then there exists a* partition *of $W$ into three events $G, B_0, B_1$ and a* partition *of $\mathcal{Y} \times \mathcal{M}$ into $E_0, E_1$ such that the following holds:*

1. *all three events $G$, $B_0$, $B_1$ have the form $W \cap S$ for some $S \in \mathcal{S}_{\text{rec}}(\rho)$;*
2. *$\rho(B_0 \cup B_1 \mid W) \le \gamma$;*

*3. let $\rho_G$ be the protocol $(\rho \mid G)$, $G_0 = G \cap E_0$ and $G_1 = G \cap E_1$, then for $b = 0, 1$,*

$$\begin{aligned}
\log \theta_\nu(\rho_G \mid G_b) &\le \log \theta_\nu(\rho \mid W) + (r+1)\log\left((r+3)/\gamma\right) + \log\left(1/((1-\gamma)\rho(G_b))\right), \\
\log \chi^2_{\nu,A}(\rho_G \mid G_b) &\le \log \chi^2_{\nu,A}(\rho \mid W) + \log((r+3)/\gamma) + \log(1/((1-\gamma)\rho(G_b))), \\
\log \chi^2_{\nu,B}(\rho_G \mid G_b) &\le \log \chi^2_{\nu,B}(\rho \mid W) + \log((r+3)/\gamma) + \log(1/((1-\gamma)\rho(G_b)));
\end{aligned}$$

*4. for $b = 0, 1$, and all $(Y, \mathbf{M})$ such that $\rho(Y, \mathbf{M} \mid G_b) > 0$,*

$$\rho(h(X,Y) = b \mid Y, \mathbf{M}, G_b) \ge 1/2.$$

Note that we upper bound the costs of $\rho_G$ conditioned on $G_b$ by the costs of $\rho$ conditioned on $W$ (plus some small quantity). Thus, for $\rho = \pi_{<k}$ and $W = S_\beta \cap U$, the costs are bounded due to Equation (10). To focus on our main proof, we will defer the proof of Lemma 37 to Section 6.5. Now we use it to prove Lemma 36.

*Proof of Lemma 36.* We first fix some $\beta \in \{\text{high-}\theta, \text{high-}\chi^2\text{-A}, \text{high-}\chi^2\text{-B}\}$. By applying Lemma 37 to protocol $\rho = \pi_{<k}$, event $W = S_\beta \cap U$, input distribution $\nu = \mu^{k-1}$ and function $h = f^{\oplus k-1}$ for $\gamma = 2^{-12}$, we obtain sets $G_\beta, B_{\beta,0}, B_{\beta,1}, E_{\beta,0}, E_{\beta,1}$. Let $G_{\beta,0} = G_\beta \cap E_{\beta,0}$, $G_{\beta,1} = G_\beta \cap E_{\beta,1}$, and $(\pi_{<k})_{G_\beta}$ be the distribution $\pi_{<k}$ conditioned on $G_\beta$. The lemma guarantees that $G_\beta, B_{\beta,0}, B_{\beta,1}$ all have the form $S_\beta \cap U \cap S$ for some $S \in \mathcal{S}_{\mathrm{rec}}(\pi_{<k}) \subseteq \mathcal{S}_{\mathrm{pt}}(\pi)$ (Proposition 28). Since $S_\beta \cap U \in \mathcal{S}_{\mathrm{pt}}(\pi)$, we have that $G_\beta, B_{\beta,0}, B_{\beta,1} \in \mathcal{S}_{\mathrm{pt}}(\pi)$. Since $E_{\beta,0}, E_{\beta,1} \in \mathcal{S}_{\mathrm{rec}}(\pi_{<k})$, we also have $G_{\beta,0}, G_{\beta,1} \in \mathcal{S}_{\mathrm{pt}}(\pi)$.

For each $\beta$, since $G_\beta \in \mathcal{S}_{\mathrm{pt}}(\pi)$, Proposition 22(ii) implies that $(\pi \mid G_\beta)$ has the partial rectangle property with respect to $\mu^k$. Then Proposition 27 implies that $(\pi_{<k} \mid G_\beta)$, i.e., $(\pi_{<k})_{G_\beta}$, has the rectangle property with respect to $\mu^{k-1}$. For each $\beta, b$, since $E_{\beta,b} \in \mathcal{S}_{\mathrm{rec}}(\pi_{<k}) = \mathcal{S}_{\mathrm{rec}}((\pi_{<k})_{G_\beta})$, the protocol $(\pi_{<k})_{G_\beta}$ and the event $E_{\beta,b}$ are one candidate for $\pi_{\mathrm{new}}$ and $V_{\mathrm{new}}$ in Lemma 26. We will prove the following sufficient condition for them to satisfy the requirements of Lemma 26.

**Claim 38.** *If we have*

$$\begin{aligned}
&\pi(G_{\beta,b}) \cdot \mathop{\mathbb{E}}_{\pi|G_{\beta,b}} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\beta,b}))\right] \\
&\qquad \ge 2^{-6} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right],
\end{aligned} \tag{11}$$

*then $(\pi_{<k})_{G_\beta}$ and event $E_{\beta,b}$ satisfy the requirements of Lemma 26 for $\pi_{\mathrm{new}}$ and $V_{\mathrm{new}}$.*

Before proving the claim, we first show that it implies Lemma 36. If Equation (11) holds for any $\beta \in \{\text{high-}\theta, \text{high-}\chi^2\text{-A}, \text{high-}\chi^2\text{-B}\}$ and $b \in \{0, 1\}$, then the lemma holds. Otherwise we must have for every $\beta$ and $b$,

$$\begin{aligned}
&\pi(G_{\beta,b}) \cdot \mathop{\mathbb{E}}_{\pi|G_{\beta,b}} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\beta,b}))\right] \\
&\qquad < 2^{-6} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right].
\end{aligned}$$

On the other hand, since $B_{\beta,b} \in \mathcal{S}_{\mathrm{pt}}(\pi)$, $B_{\beta,b} \subseteq S_\beta \cap U$ and $\pi(B_{\beta,b} \mid U) \le \pi(B_{\beta,b} \mid S_\beta \cap U) \le 2^{-12}$, by Proposition 32(i), we also have

$$\begin{aligned}
&\pi(B_{\beta,b}) \cdot \mathop{\mathbb{E}}_{\pi|B_{\beta,b}} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, B_{\beta,b}))\right] \\
&\quad \le 2^{-6} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right].
\end{aligned}$$

Since $G_{\beta,0} \cup G_{\beta,1} \cup B_{\beta,0} \cup B_{\beta,1} = S_\beta \cap U$, $S_{\text{high-cost}} = S_{\text{high-}\theta} \cup S_{\text{high-}\chi^2\text{-A}} \cup S_{\text{high-}\chi^2\text{-B}}$, and all 12 sets are disjoint, by summing up the above inequalities for all $B_{\beta,b}$ and $G_{\beta,b}$ and applying Lemma 10, we have

$$\begin{aligned}
&\pi(S_{\text{high-cost}} \cap U) \cdot \mathop{\mathbb{E}}_{\pi|S_{\text{high-cost}} \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{high-cost}} \cap U))\right] \\
&\quad < \frac{1}{3} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right],
\end{aligned}$$

contradicting with the lemma premise.

Now it suffices to prove the claim. We first observe that by Proposition 32(i), Equation (11) also implies that $\pi\left(G_{\beta,b} \mid U\right)^{1/2} \ge 2^{-6}$, which in turn, implies that $\pi(E_{\beta,b} \mid G_\beta) = \pi(G_{\beta,b} \mid G_\beta) \ge 2^{-12}$, i.e., the probability of $E_{\beta,b}$ in the distribution $(\pi_{<k})_{G_\beta}$ is at least $2^{-12}$, as required by Lemma 26. In the following, we show that the bound on $\phi_{k-1}\left((\pi_{<k})_{G_\beta} \mid E_{\beta,b}\right) = \phi_{k-1}\left((\pi_{<k})_{G_\beta} \mid G_{\beta,b}\right)$ also holds.

**Bounding $\phi^{\mathrm{cost}}_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b})$.** We first bound its $\phi^{\mathrm{cost}}_{k-1}$ value. By Lemma 37 and the fact that $\log(1/\alpha) < C_A - c\log(r/\alpha)$ and $\log(1/\alpha) < C_B - c\log(r/\alpha)$, we have

$$\begin{aligned}
&\phi^{\mathrm{cost}}_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) \\
&= \log \theta_{\mu^{k-1}}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) + \frac{\log(1/\alpha)}{C_A - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^{k-1},A}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) \\
&\qquad + \frac{\log(1/\alpha)}{C_B - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^{k-1},B}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) \\
&\le \log \theta_{\mu^{k-1}}(\pi_{<k} \mid S_\beta \cap U) + \frac{\log(1/\alpha)}{C_A - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^{k-1},A}(\pi_{<k} \mid S_\beta \cap U) \\
&\qquad + \frac{\log(1/\alpha)}{C_B - c\log(r/\alpha)} \cdot \log \chi^2_{\mu^{k-1},B}(\pi_{<k} \mid S_\beta \cap U) \\
&\qquad + (r+3)\log\left((r+3)/\gamma\right) + 3\log(1/\left((1-\gamma)\pi_{<k}(G_{\beta,b})\right)) \\
&= \phi^{\mathrm{cost}}_{k-1}(\pi_{<k} \mid S_\beta \cap U) + (r+3)\log\left((r+3)/\gamma\right) + 3\log(1/\left((1-\gamma)\pi_{<k}(G_{\beta,b})\right))
\end{aligned}$$

which by Equation (10), is at most

$$\begin{aligned}
&\le \phi^{\mathrm{cost}}_k(\pi \mid U) + O(r\log r) + 3\log(1/\pi_{<k}(G_{\beta,b})) + 3\log(1/\pi(S_\beta \cap U)) - \frac{1}{4}\log(1/\alpha) \\
&\le \phi^{\mathrm{cost}}_k(\pi \mid U) + 6\log(1/\pi(G_{\beta,b})) - \frac{1}{8}\log(1/\alpha), \qquad (12)
\end{aligned}$$

where we used the assumption that $\log(1/\alpha) > cr\log r$ for a sufficiently large $c$, and the fact that $\pi(S_\beta \cap U) \ge \pi(G_{\beta,b})$, and the fact that $G_{\beta,b}$ can also be viewed as an event in $\pi$.

**Bounding $\phi^{\mathrm{adv}}_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b})$.** Next we bound its $\phi^{\mathrm{adv}}_{k-1}$ value. Lemma 37 also guarantees that for all $(X_k, Y_{<k}, \mathbf{M})$ such that $\pi(X_k, Y_{<k}, \mathbf{M} \mid G_{\beta,b}) > 0$ (recall that $\mathbf{M}^{(\pi_{<k})} = (X_k, \mathbf{M})$), we have

$$\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, G_{\beta,b}) \geq 1/2.$$

This allows us to bound its advantage by applying the first part of Lemma 35 for $S = G_{\beta,b} \in \mathcal{S}_{\mathrm{pt}}(\pi)$ (note that $G_{\beta,b} \subseteq U$). Thus, it implies

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\pi|G_{\beta,b}}\left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, G_{\beta,b})\right] \\
&\qquad \geq \mathop{\mathbb{E}}_{\pi|G_{\beta,b}}\left[\mathrm{adv}_\pi(f^{\oplus k}(X, Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\beta,b})\right].
\end{aligned}$$

Note that the LHS is exactly the expected advantage of protocol $(\pi_{<k})_{G_\beta}$ *conditioned on* $G_{\beta,b}$:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\pi|G_{\beta,b}}\left[\mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, G_{\beta,b})\right] \\
&\quad = \mathop{\mathbb{E}}_{(\pi_{<k})|G_\beta, G_{\beta,b}}\left[\mathrm{adv}_{\pi_{<k}}(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, G_\beta, G_{\beta,b})\right] \\
&\quad = \mathop{\mathbb{E}}_{(\pi_{<k})_{G_\beta}|G_{\beta,b}}\left[\mathrm{adv}_{(\pi_{<k})_{G_\beta}}(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{((\pi_{<k})_{G_\beta})}, G_{\beta,b})\right].
\end{aligned}$$

Thus, by definition, that is

$$\phi^{\mathrm{adv}}_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) \leq \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid G_{\beta,b}). \tag{13}$$

**Bounding $\phi_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b})$.** Now we sum up the two parts of the potential function. By Equation (12) and (13), we have

$$\begin{aligned}
&\phi_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) \\
&= \phi^{\mathrm{cost}}_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) + \phi^{\mathrm{adv}}_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) \\
&\leq \phi^{\mathrm{cost}}_{k}(\pi \mid U) + \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid G_{\beta,b}) + 6\log(1/\pi(G_{\beta,b})) - \frac{1}{8}\log(1/\alpha) \\
&= \phi_{k,\mathrm{pt}}(\pi \mid U) - \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U) + \left(\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid G_{\beta,b}) + 6\log(1/\pi(G_{\beta,b}))\right) - \frac{1}{8}\log(1/\alpha)
\end{aligned}$$

which by Lemma 31, is

$$\leq \phi_k(\pi \mid V) - 13\log(1/\pi(U \mid V)) - \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U) + \left(\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid G_{\beta,b}) + 6\log(1/\pi(G_{\beta,b}))\right) - \frac{1}{8}\log(1/\alpha)$$

which by the fact that $\log(1/\pi(U \mid V)) \geq 0$ and $\pi(V) \geq 2^{-12}$, is

$$\begin{aligned}
&\leq \phi_k(\pi \mid V) - 6(\log(1/\pi(U)) - \log(1/\pi(V))) - \phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U) + \left(\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid G_{\beta,b}) + 6\log(1/\pi(G_{\beta,b}))\right) \\
&\qquad - \frac{1}{8}\log(1/\alpha) \\
&\leq \phi_k(\pi \mid V) - \left(\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U) + 6\log(1/\pi(U))\right) + \left(\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid G_{\beta,b}) + 6\log(1/\pi(G_{\beta,b}))\right)
\end{aligned}$$

$$-\frac{1}{8}\log(1/\alpha)+72. \tag{14}$$

Finally, by Equation (11) and Proposition 32(ii) (for $s=2^6$ and $t=6$), we have

$$\left(\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid U)+6\log(1/\pi(U))\right) \geq \left(\phi^{\mathrm{adv}}_{k,\mathrm{pt}}(\pi \mid G_{\beta,b})+6\log(1/\pi(G_{\beta,b}))\right)-192.$$

Plugging it into Equation (14), we obtain

$$\phi_{k-1}((\pi_{<k})_{G_\beta} \mid G_{\beta,b}) \leq \phi_k(\pi \mid V)-\frac{1}{16}\log(1/\alpha),$$

since $\alpha \leq r^{-cr}$ for a sufficiently large constant $c$. Hence, $(\pi_{<k})_{G_\beta}$ and $E_{\beta,b}$ satisfy the requirements of Lemma 26. This proves the claim, completing the proof of Lemma 36. □

### 6.3 Low costs

Now we consider the case where $\pi_k$ has low costs. It consists of all $(X_k, Y_{<k}, \mathbf{M})$ such that

$$\begin{aligned}
\alpha^{2^{-5}}\cdot\pi(U) &\leq \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\theta_\mu(\pi_k @ X_k, Y_k, \mathbf{M}^{(\pi_k)})\right] < \alpha^{-1/2},\\
2^{-2^{-5}(C_A-c\log(r/\alpha))}\cdot\pi(U) &\leq \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\chi^2_{\mu,A}(\pi_k @ X_k, Y_k, \mathbf{M}^{(\pi_k)})\right] < 2^{C_A-c\log(r/\alpha)},\\
2^{-2^{-5}(C_B-c\log(r/\alpha))}\cdot\pi(U) &\leq \mathop{\mathbb{E}}_{Y_k\sim\pi|X_k,Y_{<k},\mathbf{M},U}\left[\chi^2_{\mu,B}(\pi_k @ X_k, Y_k, \mathbf{M}^{(\pi_k)})\right] < 2^{C_B-c\log(r/\alpha)}.
\end{aligned}$$

This is the set of triples at which the costs of $\pi_k$ are not high, nor too low. Denote this set of $(X_k, Y_{<k}, \mathbf{M})$ by $S_{\text{low-cost}}$. By definition, it is disjoint from $S_{\text{high-cost}}$. Similarly, we also use $S_{\text{low-cost}}$ to denote set $\{(X,Y,\mathbf{M}) : (X_k, Y_{<k}, \mathbf{M}) \in S_{\text{low-cost}}\}$.

By applying Lemma 33 and Lemma 34 to $S_{\text{low-cost}}$ with the appropriate $\eta$, we have the following bounds:

$$\begin{aligned}
\log\theta_{\mu^{k-1}}(\pi_{<k} \mid S_{\text{low-cost}}\cap U) \leq \log\theta_{\mu^k}(\pi \mid U) &+ \log(1/\pi(S_{\text{low-cost}} \mid U))\\
&+ 2^{-5}\log(1/\alpha)+\log(1/\pi(U)),\\
\log\chi^2_{\mu^{k-1},A}(\pi_{<k} \mid S_{\text{low-cost}}\cap U) \leq \log\chi^2_{\mu^k,A}(\pi \mid U) &+ \log(1/\pi(S_{\text{low-cost}} \mid U))\\
&+ 2^{-5}(C_A-c\log(r/\alpha))+\log(1/\pi(U)),\\
\log\chi^2_{\mu^{k-1},B}(\pi_{<k} \mid S_{\text{low-cost}}\cap U) \leq \log\chi^2_{\mu^k,B}(\pi \mid U) &+ \log(1/\pi(S_{\text{low-cost}} \mid U))\\
&+ 2^{-5}(C_B-c\log(r/\alpha))+\log(1/\pi(U)).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\phi^{\mathrm{cost}}_{k-1}(\pi_{<k} \mid S_{\text{low-cost}}\cap U) &\leq \phi^{\mathrm{cost}}_k(\pi \mid U)+3\log(1/\pi(S_{\text{low-cost}} \mid U))+3\log(1/\pi(U))+3\cdot 2^{-5}\log(1/\alpha)\\
&= \phi^{\mathrm{cost}}_k(\pi \mid U)+3\log(1/\pi(S_{\text{low-cost}}\cap U))+3\cdot 2^{-5}\log(1/\alpha).
\end{aligned} \tag{15}$$

The main lemma of this subsection is the following, stating that if $S_{\text{low-cost}}$ contributes a nontrival amount of total advantage in $U$, then we can construct a protocol for $f^{\oplus k-1}$ satisfying the requirements of Lemma 26.

**Lemma 39.** *If we have*

$$\pi(S_{\text{low-cost}} \cap U) \cdot \mathop{\mathbb{E}}_{\pi|S_{\text{low-cost}} \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U)\right] \geq \frac{1}{3} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right],$$

*then Lemma 26 holds.*

The proof will use the following lemma that converts a generalized protocol with low costs to a standard protocol with low communication.

**Lemma 40.** *Let $\delta_1, \delta_2 \in (0, 1/2)$ be any fixed parameter. Let $\rho$ be an $r$-round generalized protocol and let $W$ be an event such that $(\rho \mid W)$ has the rectangle property with respect to $\mu$. Then for any function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, there is an $r$-round* standard *protocol $\tau$ such that*

- *in odd rounds of $\tau$, Alice sends a message of at most $\log \chi^2_{\mu,A}(\rho \mid W) + O(\log(r/\delta_1\delta_2) + \log\log\theta_\mu(\rho \mid W))$ bits;*
- *in even rounds of $\tau$, Bob sends a message of at most $\log \chi^2_{\mu,B}(\rho \mid W) + O(\log(r/\delta_1\delta_2))$ bits;*
- *$\tau$ computes $f$ correctly under input distribution $\mu$ with probability at least*

$$\frac{1}{2} + \frac{\delta_1}{32\theta_\mu(\rho \mid W)} \left(\rho(W) \cdot \mathop{\mathbb{E}}_{\rho|W} [\mathrm{adv}_\rho(f(X,Y) \mid X, \mathbf{M}, W)] - 6\delta_1\right) - 2r\delta_2.$$

To focus on the main proof, we will defer the proof of Lemma 40 to Section 7.

*Proof of Lemma 39.* Similar to the proof of Lemma 36, we first apply Lemma 37 to $\rho = \pi_{<k}$, event $W = S_{\text{low-cost}} \cap U$, input distribution $\nu = \mu^{k-1}$ and function $h = f^{\oplus k-1}$ for $\gamma = 2^{-12}$. We obtain sets $G_{\text{low-cost}}$, $B_{\text{low-cost},0}$, $B_{\text{low-cost},1}$ and $E_{\text{low-cost},0}$, $E_{\text{low-cost},1}$. Let $G_{\text{low-cost},0} = G_{\text{low-cost}} \cap E_{\text{low-cost},0}$, $G_{\text{low-cost},1} = G_{\text{low-cost}} \cap E_{\text{low-cost},1}$, and $(\pi_{<k})_{G_{\text{low-cost}}}$ be $\pi_{<k}$ conditioned on $G_{\text{low-cost}}$. Again, we have that $G_{\text{low-cost},0}$, $G_{\text{low-cost},1}$, $B_{\text{low-cost},0}$, $B_{\text{low-cost},1}$ and $G_{\text{low-cost}} \in \mathcal{S}_{\text{pt}}(\pi)$, $(\pi_{<k})_{G_{\text{low-cost}}}$ has the rectangle property with respect to $\mu^{k-1}$ and for $b = 0, 1$, $E_{\text{low-cost},b} \in \mathcal{S}_{\text{rec}}((\pi_{<k})_{G_{\text{low-cost}}})$.

Thus, the protocol $(\pi_{<k})_{G_{\text{low-cost}}}$ and the event $E_{\text{low-cost},b}$ are one possible candidate for Lemma 26. We will prove the following sufficient condition for them to satisfy the requirements of Lemma 26.

**Claim 41.** *If we have*

$$\begin{aligned}\pi(G_{\text{low-cost},b}) \cdot \mathop{\mathbb{E}}_{\pi|G_{\text{low-cost},b}} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b}))\right] \\ \geq 2^{-6} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right],\end{aligned} \tag{16}$$

*then $(\pi_{<k})_{G_{\text{low-cost}}}$ and $E_{\text{low-cost},b}$ satisfy the requirements of Lemma 26 for $\pi_{\text{new}}$ and $V_{\text{new}}$.*

Similar to the proof of Lemma 36, before proving the claim, we first show that it implies the lemma. If Equation (16) holds for either $b = 0$ or $b = 1$, then the lemma holds. Otherwise, we have

$$\pi(G_{\text{low-cost},b}) \cdot \mathop{\mathbb{E}}_{\pi|G_{\text{low-cost},b}} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b}))\right]$$

$$< 2^{-6} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)) \right],$$

for $b = 0, 1$. Lemma 37 guarantees that $\pi(B_{\text{low-cost},b} \mid U) \leq \pi(B_{\text{low-cost},b} \mid S_{\text{low-cost}} \cap U) \leq 2^{-12}$. By Proposition 32(i), we also have

$$\begin{aligned}
&\pi(B_{\text{low-cost},b}) \cdot \mathop{\mathbb{E}}_{\pi|B_{\text{low-cost},b}} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, B_{\text{low-cost},b})) \right] \\
&\leq 2^{-6} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)) \right].
\end{aligned}$$

By summing up the inequalities and applying Lemma 10, we obtain

$$\begin{aligned}
&\pi(S_{\text{low-cost}} \cap U) \cdot \mathop{\mathbb{E}}_{\pi|S_{\text{low-cost}} \cap U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U)) \right] \\
&< 2^{-4} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)) \right],
\end{aligned}$$

contracting with the lemma premise.

Now it suffices to prove the claim. By Proposition 32(i), Equation (16) implies that $\pi(G_{\text{low-cost},b} \mid U) \geq 2^{-12}$. Therefore, the probability of $E_{\text{low-cost},b}$ in the distribution $(\pi_{<k})_{G_{\text{low-cost}}}$ is at least $2^{-12}$ as required by Lemma 26. In the following, we show that the bound on $\phi_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid E_{\text{low-cost},b}) = \phi_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b})$ holds.

**Bounding $\phi_{k-1}^{\text{cost}}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b})$.** We first bound its $\phi_{k-1}^{\text{cost}}$ value. Similar to the proof of Lemma 36, Lemma 37 guarantees that

$$\begin{aligned}
&\phi_{k-1}^{\text{cost}}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \\
&\quad \leq \phi_{k-1}^{\text{cost}}(\pi_{<k} \mid S_{\text{low-cost}} \cap U) + O(r \log r) + 3\log(1/\pi(G_{\text{low-cost},b}))
\end{aligned}$$

which by Equation (15), is

$$\begin{aligned}
&\leq \phi_k^{\text{cost}}(\pi \mid U) + O(r \log r) + 3\log(1/\pi(G_{\text{low-cost},b})) + 3\log(1/\pi(S_{\text{low-cost}} \cap U)) + 3 \cdot 2^{-5} \log(1/\alpha) \\
&\leq \phi_k^{\text{cost}}(\pi \mid U) + 6\log(1/\pi(G_{\text{low-cost},b})) + 2^{-3} \log(1/\alpha), \qquad (17)
\end{aligned}$$

where we use the fact that $\alpha > r^{cr}$ for a sufficiently $c$, and $\pi(G_{\text{low-cost},b}) \leq \pi(S_{\text{low-cost}} \cap U)$.

To bound its $\phi_{k-1}^{\text{adv}}$ and then $\phi_{k-1}$, we will consider two cases: $\pi(U) \leq \alpha^{1/8}$ and $\pi(U) > \alpha^{1/8}$.

**Bounding $\phi_{k-1}^{\text{adv}}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b})$ when $\pi(U) \leq \alpha^{1/8}$.** We first bound its $\phi_{k-1}^{\text{adv}}$ when $\pi(U) \leq \alpha^{1/8}$. Similar to the proof of Lemma 36, Lemma 37 implies that

$$\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b}) \geq 1/2.$$

Thus, the first part of Lemma 35 for $S = G_{\text{low-cost},b}$ implies that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\pi|G_{\text{low-cost},b}} \left[ \mathrm{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid X_k, \mathbf{M}, G_{\text{low-cost},b}) \right] \\
&\qquad \geq \mathop{\mathbb{E}}_{\pi|G_{\text{low-cost},b}} \left[ \mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b}) \right].
\end{aligned}$$

That is,

$$\phi_{k-1}^{\text{adv}}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \leq \phi_{k,\text{pt}}^{\text{adv}}(\pi \mid G_{\text{low-cost},b}). \qquad (18)$$

**Bounding $\phi_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b})$ when $\pi(U) \le \alpha^{1/8}$.** By Equation (17) and (18), we have

$$\begin{aligned}
&\phi_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \\
&= \phi^{\text{cost}}_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) + \phi^{\text{adv}}_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \\
&\le \phi^{\text{cost}}_{k}(\pi \mid U) + \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b})) + 2^{-3}\log(1/\alpha) \\
&= \phi_{k,\text{pt}}(\pi \mid U) - \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid U) + \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b})) + 2^{-3}\log(1/\alpha)
\end{aligned}$$

which by Lemma 31, is

$$\begin{aligned}
&\le \phi_k(\pi \mid V) - (\phi^{\text{adv}}_{k,\text{pt}}(\pi \mid U) + 6\log(1/\pi(U))) + (\phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b}))) \\
&\qquad + 2^{-3}\log(1/\alpha) - 7\log(1/\pi(U)) + 13\log(1/\pi(V))
\end{aligned}$$

which by the assumption that $\pi(U) \le \alpha^{1/8}$ and $\pi(V) \ge 2^{-12}$, is

$$\begin{aligned}
&\le \phi_k(\pi \mid V) - (\phi^{\text{adv}}_{k,\text{pt}}(\pi \mid U) + 6\log(1/\pi(U))) + (\phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b}))) \\
&\qquad - \frac{3}{4}\log(1/\alpha) + 156.
\end{aligned}$$

By Proposition 32(ii), Equation (16) implies that

$$\phi^{\text{adv}}_{k,\text{pt}}(\pi \mid U) + 6\log(1/\pi(U)) \ge \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b})) - 192.$$

Thus, $\phi_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \le \phi_k(\pi \mid V) - \frac{1}{4}\log(1/\alpha)$, as $\alpha < r^{cr}$ for a large $c$. This proves Claim 41 when $\pi(U) \le \alpha^{1/8}$.

**Bounding $\phi^{\text{adv}}_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b})$ when $\pi(U) > \alpha^{1/8}$.** Next we consider the case where $\pi(U) > \alpha^{1/8}$. To bound $\phi^{\text{adv}}_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b})$ in this case, we will apply the second part of Lemma 35. To this end, we will first upper bound the advantage of $\pi_k$ for computing $f(X_k, Y_k)$ by applying Lemma 40 to $\rho = \pi_k$ and $W = G_{\text{low-cost},b}$ and using the assumption on the communication complexity of $f$.

To verify the premises of Lemma 40 are satisfied, note that $G_{\text{low-cost},b}$ is in $\mathcal{S}_{\text{pt}}(\pi)$, hence, $(\pi \mid G_{\text{low-cost},b})$ has the partial rectangle property with respect to $\mu^k$ by Proposition 22(ii). Hence, $(\pi_k \mid G_{\text{low-cost},b})$ has the rectangle property with respect to $\mu$ by Proposition 27. To bound the costs of $\pi_k$ conditioned on $G_{\text{low-cost},b}$, note that since $S_{\text{low-cost}} \cap U \in \mathcal{S}_{\text{pt}}(\pi)$, we have $Y_k$ and $X_{<k}$ are independent conditioned on $(X_k, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U)$ by Proposition 20. Note that $G_{\text{low-cost},b} \subseteq S_{\text{low-cost}} \cap U$, and note that whether the event $G_{\text{low-cost},b}$ happens is determined by $(X_{<k}, Y_{<k}, (X_k, \mathbf{M}), S_{\text{low-cost}} \cap U)$, and whether $S_{\text{low-cost}}$ happens is determined by $(X_k, Y_{<k}, \mathbf{M}, U)$. Therefore, when $\pi(X, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b}) > 0$, the distribution of $Y_k$ conditioned on $(X, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b})$ is the same as the distribution of $Y_k$ conditioned on $(X_k, Y_{<k}, \mathbf{M}, U)$, because

$$\begin{aligned}
&\pi(Y_k = y_k \mid X, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b}) \\
&= \pi(Y_k = y_k \mid X, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U, G_{\text{low-cost},b}) && (G_{\text{low-cost},b} \subseteq S_{\text{low-cost}} \cap U) \\
&= \pi(Y_k = y_k \mid X, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U) && (G_{\text{low-cost},b} \text{ implied by } (X, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U)) \\
&= \pi(Y_k = y_k \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U) && (X_{<k} \perp Y_k \text{ given } (X_k, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U)) \\
&= \pi(Y_k = y_k \mid X_k, Y_{<k}, \mathbf{M}, U, S_{\text{low-cost}} \cap U) && (S_{\text{low-cost}} \cap U \subseteq U)
\end{aligned}$$

$$= \pi(Y_k = y_k \mid X_k, Y_{<k}, \mathbf{M}, U) \qquad (S_{\text{low-cost}} \text{ implied by } (X_k, Y_{<k}, \mathbf{M})).$$

Thus, the $\theta$-cost of $\pi_k$ conditioned on $G_{\text{low-cost},b}$ is at most

$$\begin{aligned}
&\theta_\mu(\pi_k \mid G_{\text{low-cost},b}) \\
&\quad= \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\pi|G_{\text{low-cost},b}} [\theta_\mu(\pi_k \mathbin{@} X, Y, \mathbf{M})] \\
&\quad= \mathop{\mathbb{E}}_{(X,Y_{<k},\mathbf{M})\sim\pi|G_{\text{low-cost},b}} \left[ \mathop{\mathbb{E}}_{Y_k\sim\pi|(X,Y_{<k},\mathbf{M},G_{\text{low-cost},b})} [\theta_\mu(\pi_k \mathbin{@} X, Y, \mathbf{M})] \right] \\
&\quad= \mathop{\mathbb{E}}_{(X,Y_{<k},\mathbf{M})\sim\pi|G_{\text{low-cost},b}} \left[ \mathop{\mathbb{E}}_{Y_k\sim\pi|(X_k,Y_{<k},\mathbf{M},U)} [\theta_\mu(\pi_k \mathbin{@} X, Y, \mathbf{M})] \right] \\
&\quad< \mathop{\mathbb{E}}_{(X,Y_{<k},\mathbf{M})\sim\pi|G_{\text{low-cost},b}} \left[ \alpha^{-1/2} \right] \\
&\quad= \alpha^{-1/2}.
\end{aligned}$$

For the same reason, its $\chi^2$-cost by Alice is at most $2^{C_A - c\log(r/\alpha)}$, and its $\chi^2$-cost by Bob is at most $2^{C_B - c\log(r/\alpha)}$. By applying Lemma 40 to $\rho = \pi_k$ and event $W = G_{\text{low-cost},b}$ for $\delta_1 = \alpha^{1/4}$ and $\delta_2 = \alpha \cdot r^{-1}$, we obtain a *standard* $r$-round protocol $\tau$. Since $c$ is a sufficiently large constant, we have that in $\tau$,

- Alice sends at most

$$C_A - c \cdot \log(r/\alpha) + O(\log(r/\delta_1\delta_2) + \log\log\theta_\mu(\pi_k \mid G_{\text{low-cost},b})) \le C_A$$

  bits in every odd round;

- Bob sends at most

$$C_B - c \cdot \log(r/\alpha) + O(\log(r/\delta_1\delta_2)) \le C_B$$

  bits in every even round;

- $\tau$ computes $f$ correctly under input distribution $\mu$ with probability at least

$$\frac{1}{2} + \frac{\delta_1}{32\theta_\mu(\pi_k \mid G_{\text{low-cost},b})} \cdot \left( \pi_k(G_{\text{low-cost},b}) \cdot \mathop{\mathbb{E}}_{\pi_k|G_{\text{low-cost},b}} \left[ \mathrm{adv}_{\pi_k}(f(X_k, Y_k) \mid X_k, \mathbf{M}^{(\pi_k)}, G_{\text{low-cost},b}) \right] - 6\delta_1 \right) - 2r\delta_2.$$

By the our assumption on the communication complexity of $f$, the expected advantage of $\tau$ must be at most $\alpha$:

$$\frac{\delta_1}{16\theta_\mu(\pi_k \mid G_{\text{low-cost},b})} \cdot \left( \pi_k(G_{\text{low-cost},b}) \cdot \mathop{\mathbb{E}}_{\pi_k|G_{\text{low-cost},b}} \left[ \mathrm{adv}_{\pi_k}(f(X_k, Y_k) \mid X_k, \mathbf{M}^{(\pi_k)}, G_{\text{low-cost},b}) \right] - 6\delta_1 \right) - 4r\delta_2 \le \alpha.$$

It implies that

$$\begin{aligned}
&\pi_k(G_{\text{low-cost},b}) \cdot \mathop{\mathbb{E}}_{\pi_k|G_{\text{low-cost},b}} [\mathrm{adv}_{\pi_k}(f(X_k, Y_k) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b})] \\
&\quad\le (\alpha + 4r\delta_2) \cdot \frac{16\theta_\mu(\pi_k \mid G_{\text{low-cost},b})}{\delta_1} + 6\delta_1
\end{aligned}$$

$$\leq (\alpha + 4\alpha) \cdot \frac{16\alpha^{-1/2}}{\alpha^{1/4}} + 6\alpha^{1/4}$$
$$= 86\alpha^{1/4}.$$

Since $G_{\text{low-cost},b} \subseteq U$, we apply the second part of Lemma 35 for $S = G_{\text{low-cost},b}$ and

$$\eta = 86\alpha^{1/4} \cdot \pi(G_{\text{low-cost},b})^{-1}.$$

The premises of Lemma 35 are satisfied, because

(a) Lemma 37 gives that

$$\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) = b \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b}) \geq 1/2;$$

(b) by Proposition 32(i), Equation (16) implies that $\pi(G_{\text{low-cost},b} \mid U)^{1/2} \geq 2^{-6}$, hence,

$$\pi(G_{\text{low-cost},b}) \geq 2^{-12} \cdot \pi(U) \geq 2^{-12} \cdot \alpha^{1/8};$$

(c) then we have

$$\eta^{1/4} = \left(86\alpha^{1/4} \cdot \pi(G_{\text{low-cost},b})^{-1}\right)^{1/4}$$
$$\leq \left(2^{20}\alpha^{1/8}\right)^{1/4}$$
$$= 2^5\alpha^{1/32};$$

(d) Equation (16) implies

$$\frac{1}{2} \cdot \frac{\pi(G_{\text{low-cost},b})^{1/2} \cdot \mathbb{E}_{\pi|G_{\text{low-cost},b}}\left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b})\right]}{\pi(U)^{1/2} \cdot \mathbb{E}_{\pi|U}\left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right]}$$
$$= \frac{1}{2 \cdot \pi(G_{\text{low-cost},b} \mid U)^{1/2}} \cdot \frac{\pi(G_{\text{low-cost},b}) \cdot \mathbb{E}_{\pi|G_{\text{low-cost},b}}\left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b})\right]}{\pi(U) \cdot \mathbb{E}_{\pi|U}\left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)\right]}$$
$$\geq \frac{1}{2 \cdot \pi(G_{\text{low-cost},b} \mid U)^{1/2}} \cdot 2^{-6}$$
$$\geq 2^{-7},$$

which is at least $\eta^{1/4}$ by the upper bound on $\eta^{1/4}$ in (c) and the fact that $\alpha$ is sufficiently small.

Hence, we obtain the following by the second part of Lemma 35,

$$\mathop{\mathbb{E}}_{\pi|G_{\text{low-cost},b}}\left[\text{adv}_\pi(f^{\oplus k-1}(X_{<k}, Y_{<k}) \mid \mathbf{M}^{(\pi_{<k})}, G_{\text{low-cost},b})\right]$$
$$\geq \frac{1}{2} \cdot \eta^{-1/2} \cdot \mathop{\mathbb{E}}_{\pi|G_{\text{low-cost},b}}\left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b})\right]$$

which by the above upper bound on $\eta^{1/4}$ in (c), is

$$\geq 2^{-11} \cdot \alpha^{-1/16} \cdot \mathop{\mathbb{E}}_{\pi|G_{\text{low-cost},b}}\left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, G_{\text{low-cost},b})\right]$$

i.e.,

$$\phi^{\text{adv}}_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \leq \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) - 2\log(1/\alpha) + 352. \tag{19}$$

**Bounding $\phi_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b})$ when $\pi(U) > \alpha^{1/8}$.** Combining Equation (17) and (19), we have

$$\begin{aligned}
&\phi_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \\
&= \phi^{\text{cost}}_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) + \phi^{\text{adv}}_{k-1}((\pi_{<k})_{G_{\text{low-cost}}} \mid G_{\text{low-cost},b}) \\
&\le \phi^{\text{cost}}_{k}(\pi \mid U) + \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b})) + 2^{-3}\log(1/\alpha) - 2\log(1/\alpha) + 352 \\
&= \phi_{k,\text{pt}}(\pi \mid U) - \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid U) + \phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b})) - \frac{15}{8}\log(1/\alpha) + 352
\end{aligned}$$

which by Lemma 31 and the fact that $\log(1/\pi(U \mid V)) \ge 0$, is

$$\begin{aligned}
&\le \phi_k(\pi \mid V) - (\phi^{\text{adv}}_{k,\text{pt}}(\pi \mid U) + 6\log(1/\pi(U))) + (\phi^{\text{adv}}_{k,\text{pt}}(\pi \mid G_{\text{low-cost},b}) + 6\log(1/\pi(G_{\text{low-cost},b}))) \\
&\quad - \frac{15}{8}\log(1/\alpha) + 352 + 6\log(1/\pi(V))
\end{aligned}$$

which by Proposition 32(ii) and the fact that $\pi(V) \ge 2^{-12}$ and $\alpha$ is sufficiently small, is

$$\le \phi_k(\pi \mid V) - \log(1/\alpha).$$

This proves Claim 41 when $\pi(U) > \alpha^{1/8}$, and completes the proof of Lemma 39. □

## 6.4 Putting together

Now we are ready to prove Lemma 26. The two main lemmas in the previous two subsections show that if either $S_{\text{high-cost}}$ or $S_{\text{low-cost}}$ contributes a nontrivial advantage in $U$, then Lemma 26 holds. We will show that the complement of their union has very low probability, hence contributes a small amount of advantage by Proposition 32(i). Then the superadditivity of weighted advantage implies the lemma.

**Lemma 26 (restated).** *For $k \ge 2$,*

***if** there is a generalized protocol $\pi$ for $f^{\oplus k}$ with the rectangle property with respect to $\mu^k$ and an event $V \in \mathcal{S}_{\text{rec}}(\pi)$ such that $\pi(V) \ge 2^{-12}$,*

***then** there is a generalized protocol $\pi_{\text{new}}$ for $f^{\oplus k-1}$ with the rectangle property with respect to $\mu^{k-1}$ and an event $V_{\text{new}} \in \mathcal{S}_{\text{rec}}(\pi_{\text{new}})$ such that $\pi_{\text{new}}(V_{\text{new}}) \ge 2^{-12}$, and*

$$\phi_{k-1}(\pi_{\text{new}} \mid V_{\text{new}}) \le \phi_k(\pi \mid V) - \frac{1}{16}\log(1/\alpha).$$

*Proof.* If the premise of Lemma 36 or Lemma 39 holds, then Lemma 26 holds.

Otherwise, we have that

$$\begin{aligned}
&\pi(S_{\text{high-cost}} \cap U) \cdot \mathop{\mathbb{E}}_{\pi|S_{\text{high-cost}} \cap U} \left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{high-cost}} \cap U)\right] \\
&\quad < \frac{1}{3} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right], \qquad (20)
\end{aligned}$$

and

$$\pi(S_{\text{low-cost}} \cap U) \cdot \mathop{\mathbb{E}}_{\pi|S_{\text{low-cost}} \cap U} \left[\text{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{low-cost}} \cap U)\right]$$

$$< \frac{1}{3} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[ \mathrm{adv}_{\pi}(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U)) \right]. \tag{21}$$

On the other hand, by construction, the complement of $S_{\text{high-cost}} \cup S_{\text{low-cost}}$ is the set of all triples $(X_k, Y_{<k}, \mathbf{M})$ such that either

$$\mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] < \alpha^{2^{-5}} \cdot \pi(U), \text{ or}$$
$$\mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \chi^2_{\mu,A}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] < 2^{-2^{-5}(C_A - c\log(r/\alpha))} \cdot \pi(U), \text{ or}$$
$$\mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \chi^2_{\mu,B}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] < 2^{-2^{-5}(C_B - c\log(r/\alpha))} \cdot \pi(U).$$

Denote this set by $S_{\text{low-prob}}$. Clearly, we also have $S_{\text{low-prob}} \cap U \in \mathcal{S}_{\mathrm{pt}}(\pi)$.

However, by Proposition 13, we have

$$\mathop{\mathbb{E}}_{\pi|U} \left[ \theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})^{-1} \right] \le \pi(U)^{-1} \cdot \mathop{\mathbb{E}}_{\pi} \left[ \theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})^{-1} \right] = \pi(U)^{-1}.$$

Since $\theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})$ is a function of $(X_k, Y, \mathbf{M})$, by the convexity of $x^{-1}$, we also have

$$\left( \mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] \right)^{-1} \le \mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)})^{-1} \right],$$

and hence,

$$\mathop{\mathbb{E}}_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi | U} \left[ \left( \mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] \right)^{-1} \right] \le \pi(U)^{-1}.$$

By Markov's inequality, we obtain

$$\Pr_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi | U} \left[ \mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \theta_{\mu}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] < \alpha^{2^{-5}} \cdot \pi(U) \right] < \alpha^{2^{-5}}.$$

Similarly, by invoking Proposition 16, we have

$$\Pr_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi | U} \left[ \mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \chi^2_{\mu,A}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] < 2^{-2^{-5}(C_A - c\log(r/\alpha))} \cdot \pi(U) \right] < 2^{-2^{-5}(C_A - c\log(r/\alpha))},$$

and

$$\Pr_{(X_k, Y_{<k}, \mathbf{M}) \sim \pi | U} \left[ \mathop{\mathbb{E}}_{Y_k \sim \pi | X_k, Y_{<k}, \mathbf{M}, U} \left[ \chi^2_{\mu,B}(\pi_k \mathbin{@} X_k, Y_k, \mathbf{M}^{(\pi_k)}) \right] < 2^{-2^{-5}(C_B - c\log(r/\alpha))} \cdot \pi(U) \right] < 2^{-2^{-5}(C_B - c\log(r/\alpha))}.$$

Thus, by union bound, we have

$$\pi(S_{\text{low-prob}} \mid U) < \alpha^{2^{-5}} + 2^{-2^{-5}(C_A - c\log(r/\alpha))} + 2^{-2^{-5}(C_B - c\log(r/\alpha))} < 1/9,$$

since $\alpha < r^{cr}, C_A, C_B > 2c\log(r/\alpha)$ for a sufficiently large $c$. By applying Proposition 32(i) on $S_{\text{low-prob}} \cap U$, we obtain

$$\pi(S_{\text{low-prob}} \cap U) \cdot \mathop{\mathbb{E}}_{\pi|S_{\text{low-prob}} \cap U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, S_{\text{low-prob}} \cap U)\right]$$
$$< \frac{1}{3} \cdot \pi(U) \cdot \mathop{\mathbb{E}}_{\pi|U} \left[\mathrm{adv}_\pi(f^{\oplus k}(X,Y) \mid X_k, Y_{<k}, \mathbf{M}, U))\right]. \tag{22}$$

By summing up Equation (20), (21) and (22), we get a contradiction with Lemma 10. This completes the proof of Lemma 26. □

## 6.5 Proof of Lemma 37

In this subsection, we prove Lemma 37, which lets us convert a protocol conditioned on an event to a generalized protocol with bounded costs.

**Lemma 37 (restated).** *Fix any $\gamma \in (0, 1/2)$. Let $\rho$ be an $r$-round generalized protocol over $\mathcal{X} \times \mathcal{Y} \times \mathcal{M}$, $W$ be an event, $\nu$ be an input distribution and $h : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function of the inputs. Then there exists a* partition *of $W$ into three events $G, B_0, B_1$ and a* partition *of $\mathcal{Y} \times \mathcal{M}$ into $E_0, E_1$ such that the following holds:*

1. *all three events $G$, $B_0$, $B_1$ have the form $W \cap S$ for some $S \in \mathcal{S}_{\mathrm{rec}}(\rho)$;*
2. *$\rho(B_0 \cup B_1 \mid W) \le \gamma$;*
3. *let $\rho_G$ be the protocol $(\rho \mid G)$, $G_0 = G \cap E_0$ and $G_1 = G \cap E_1$, then for $b = 0, 1$,*
$$\log \theta_\nu(\rho_G \mid G_b) \le \log \theta_\nu(\rho \mid W) + (r+1)\log\left((r+3)/\gamma\right) + \log\left(1/((1-\gamma)\rho(G_b))\right),$$
$$\log \chi^2_{\nu,A}(\rho_G \mid G_b) \le \log \chi^2_{\nu,A}(\rho \mid W) + \log((r+3)/\gamma) + \log(1/((1-\gamma)\rho(G_b))),$$
$$\log \chi^2_{\nu,B}(\rho_G \mid G_b) \le \log \chi^2_{\nu,B}(\rho \mid W) + \log((r+3)/\gamma) + \log(1/((1-\gamma)\rho(G_b)));$$
4. *for $b = 0, 1$, and all $(Y, \mathbf{M})$ such that $\rho(Y, \mathbf{M} \mid G_b) > 0$,*
$$\rho(h(X,Y) = b \mid Y, \mathbf{M}, G_b) \ge 1/2.$$

*Proof.* Ideally, we could simply let $G_0$ be the intersection of $W$ and all $(Y, \mathbf{M})$ such that $\rho(h(X,Y) = 0 \mid Y, \mathbf{M}, W) \ge 1/2$, and $G_1$ be the intersection of $W$ and all other $(Y, \mathbf{M})$. In this way, the last line of the lemma holds, since $G_b$ is a set that depends only on $(Y, \mathbf{M}, W)$ and is a subset of $W$,

$$\rho(h(X,Y) = b \mid Y, \mathbf{M}, G_b) = \rho(h(X,Y) = b \mid Y, \mathbf{M}, W, G_b) = \rho(h(X,Y) = b \mid Y, \mathbf{M}, W).$$

However, the new protocol $\rho_G$ (for $G = W$ in this case) may not have low costs, since the denominators in the definitions may become arbitrarily small (recall Definition 12 and Definition 15). To ensure that the costs of the new protocol $\rho_G$ are bounded, we will identify all $(X, \mathbf{M})$ and $(Y, \mathbf{M})$ at which the denominators in the definition of $\theta$-cost and $\chi^2$-costs becomes much smaller, and repeatedly remove such pairs from the support.

More specifically, we repeatedly remove from the support of $\rho$, all $M_0$ whose probability becomes much smaller after conditioning on $G$, we also remove pairs $(X, \mathbf{M})$ [resp. $(Y, \mathbf{M})$] such that either $\rho(X, \mathbf{M})$ or for some odd $i$, $\rho(M_i \mid X, M_{<i})$ [resp. $\rho(Y, \mathbf{M})$ or for some even $i$, $\rho(M_i \mid Y, M_{<i})$] becomes much smaller after conditioning.

Formally, consider the following process:[6]

[6]Note that $W$ may not necessarily be a subset of $\mathcal{X} \times \mathcal{Y} \times \mathcal{M}$, it could be any event.

1. $W_{\mathrm{tmp}} \leftarrow W$ // the current $W$
2. $B_{X,M} \leftarrow \varnothing$ // the bad $(X, \mathbf{M})$ pairs
3. $B_{Y,M} \leftarrow \varnothing$ // the bad $(Y, \mathbf{M})$ pairs
4. repeat
5. if $\exists m_0$ such that $0 < \rho(m_0 \mid W_{\mathrm{tmp}}) < \frac{\gamma}{r+3} \cdot \rho(m_0)$
6. $B_{X,M} \leftarrow B_{X,M} \cup (\mathcal{X} \times \{\mathbf{M} : M_0 = m_0\})$
7. $W_{\mathrm{tmp}} \leftarrow W_{\mathrm{tmp}} \setminus (\mathcal{X} \times \mathcal{Y} \times \{\mathbf{M} : M_0 = m_0\})$
8. if $\exists x, \mathbf{m}$ such that $0 < \rho(x, \mathbf{m} \mid W_{\mathrm{tmp}}) < \frac{\gamma}{r+3} \cdot \rho(x, \mathbf{m})$
9. $B_{X,M} \leftarrow B_{X,M} \cup \{(x, \mathbf{m})\}$
10. $W_{\mathrm{tmp}} \leftarrow W_{\mathrm{tmp}} \setminus (\{x\} \times \mathcal{Y} \times \{\mathbf{m}\})$
11. if $\exists y, \mathbf{m}$ such that $0 < \rho(y, \mathbf{m} \mid W_{\mathrm{tmp}}) < \frac{\gamma}{r+3} \cdot \rho(y, \mathbf{m})$
12. $B_{Y,M} \leftarrow B_{Y,M} \cup \{(y, \mathbf{m})\}$
13. $W_{\mathrm{tmp}} \leftarrow W_{\mathrm{tmp}} \setminus (\mathcal{X} \times \{y\} \times \{\mathbf{m}\})$
14. if $\exists x, \mathrm{odd}\, i \in [r], m_{\le i}$ such that $0 < \rho(m_i \mid x, m_{<i}, W_{\mathrm{tmp}}) < \frac{\gamma}{r+3} \cdot \rho(m_i \mid x, m_{<i})$
15. $B_{X,M} \leftarrow B_{X,M} \cup (\{x\} \times \{\mathbf{M} : M_{\le i} = m_{\le i}\})$
16. $W_{\mathrm{tmp}} \leftarrow W_{\mathrm{tmp}} \setminus (\{x\} \times \mathcal{Y} \times \{\mathbf{M} : M_{\le i} = m_{\le i}\})$
17. if $\exists y, \mathrm{even}\, i \in [r], m_{\le i}$ such that $0 < \rho(m_i \mid y, m_{<i}, W_{\mathrm{tmp}}) < \frac{\gamma}{r+3} \cdot \rho(m_i \mid y, m_{<i})$
18. $B_{Y,M} \leftarrow B_{Y,M} \cup (\{y\} \times \{\mathbf{M} : M_{\le i} = m_{\le i}\})$
19. $W_{\mathrm{tmp}} \leftarrow W_{\mathrm{tmp}} \setminus (\mathcal{X} \times \{y\} \times \{\mathbf{M} : M_{\le i} = m_{\le i}\})$
20. until none of line 5,8,11,14,17 holds
21. $G \leftarrow W_{\mathrm{tmp}}$
22. $E_0 \leftarrow \{(Y, \mathbf{M}) : \rho(Y, \mathbf{M} \mid G) = 0 \vee \rho(h = 0 \mid Y, \mathbf{M}, G) \ge 1/2\}$
23. $E_1 \leftarrow \{(Y, \mathbf{M}) : \rho(Y, \mathbf{M} \mid G) > 0 \wedge \rho(h = 1 \mid Y, \mathbf{M}, G) > 1/2\}$
24. $G_0 \leftarrow G \cap E_0$
25. $G_1 \leftarrow G \cap E_1$
26. $B_0 \leftarrow W \cap \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \in B_{X,M}\}$
27. $B_1 \leftarrow W \cap \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \notin B_{X,M}, (Y, \mathbf{M}) \in B_{Y,M}\}$
28. return $(G, B_0, B_1, E_0, E_1)$

By construction, $(G, B_0, B_1)$ is a partition of $W$ and $(E_0, E_1)$ is a partition of $\mathcal{Y} \times \mathcal{M}$. To see that Item 1 holds, note that the set $B_{X,M}$ and its complement are in $\mathcal{U}_{X,M}$, the set $B_{Y,M}$ is in $\mathcal{U}_{Y,M}$ (recall Definition 21), and note that $G$ is also the set $W \cap \{(X, Y, \mathbf{M}) : (X, \mathbf{M}) \notin B_{X,M}, (Y, \mathbf{M}) \notin B_{Y,M}\}$. Thus, $G, B_0, B_1$ have the form $W \cap S$ for some $S \in \mathcal{S}_{\mathrm{rec}}$.

Since for $b = 0, 1$, for all $(Y, \mathbf{M})$ such that $\rho(Y, \mathbf{M} \mid G_b) > 0$, we have

$$\rho(h(X, Y) = b \mid Y, \mathbf{M}, G_b) = \rho(h(X, Y) = b \mid Y, \mathbf{M}, G, G_b) = \rho(h(X, Y) = b \mid Y, \mathbf{M}, G) \ge 1/2.$$

Item 4 also holds.

It remains to bound $\rho(B_0 \cup B_1)$, and bound the costs of $\rho_G$ conditioned on $G_b$.

**Claim 42.** *We have* $\rho(B_0 \cup B_1) \le \gamma \cdot \rho(W)$.

To see this, first observe that by construction, $B_0 \cup B_1$ contains all triples that are "removed from $W$" in the whole process. Let us first focus on step 5-7, and upper bound the total probability of all $m_0$ that are removed in step 7. Observe that each $m_0$ can only be removed at most once. Each time we remove a $m_0$, $\rho(W_{\mathrm{tmp}})$ decreases by $\rho(m_0, W_{\mathrm{tmp}})$ (for the $W_{\mathrm{tmp}}$ at the time of the removal). Since $\rho(m_0 \mid W_{\mathrm{tmp}}) < \frac{\gamma}{r+3} \cdot \rho(m_0)$ at the time of the removal, we have

$$\rho(m_0, W_{\mathrm{tmp}}) = \rho(W_{\mathrm{tmp}}) \cdot \rho(m_0 \mid W_{\mathrm{tmp}})$$

$$\leq \rho(W) \cdot \left( \frac{\gamma}{r+3} \cdot \rho(m_0) \right).$$

Therefore, during the entire process, $\rho(W_{\text{tmp}})$ can decrease in step 7 by at most

$$\sum_{m_0} \rho(W) \cdot \left( \frac{\gamma}{r+3} \cdot \rho(m_0) \right) = \frac{\gamma}{r+3} \cdot \rho(W).$$

Similarly, in step 10 and step 13, $\rho(W_{\text{tmp}})$ can also decrease by at most $\frac{\gamma}{r+3} \cdot \rho(W)$ respectively.

Next, consider step 14-16, and fix an odd $i$. Each time we remove a pair $(x, m_{\leq i})$, $\rho(W_{\text{tmp}})$ decreases by $\rho(x, m_{\leq i}, W_{\text{tmp}})$. Since $\rho(m_i \mid x, m_{<i}, W_{\text{tmp}}) < \frac{\gamma}{r+3} \cdot \rho(m_i \mid x, m_{<i})$, we have

$$\begin{aligned}\rho(x, m_{\leq i}, W_{\text{tmp}}) &= \rho(x, m_{<i}, W_{\text{tmp}}) \cdot \rho(m_i \mid x, m_{<i}, W_{\text{tmp}}) \\ &\leq \rho(x, m_{<i}, W) \cdot \left( \frac{\gamma}{r+3} \cdot \rho(m_i \mid x, m_{<i}) \right).\end{aligned}$$

Therefore, during the whole process, the probability $\rho(W_{\text{tmp}})$ can decrease in step 16 for a fixed $i$ by at most

$$\begin{aligned}&\sum_{(x, m_{\leq i})} \rho(x, m_{<i}, W) \cdot \left( \frac{\gamma}{r+3} \cdot \rho(m_i \mid x, m_{<i}) \right) \\ &= \frac{\gamma}{r+3} \cdot \sum_{(x, m_{<i})} \rho(x, m_{<i}, W) \\ &= \frac{\gamma}{r+3} \cdot \rho(W).\end{aligned}$$

Similarly, $\rho(W_{\text{tmp}})$ can decrease in step 19 for a fixed $i$ by at most $\frac{\gamma}{r+3} \cdot \rho(W)$.

Hence, summing over all $i$ for step 16 and 19 and over all steps, $\rho(W_{\text{tmp}})$ decreases by at most $\gamma \cdot \rho(W)$, i.e., $\rho(B_0 \cup B_1) \leq \gamma \cdot \rho(W)$, proving Claim 42. Equivalently, $\rho(B_0 \cup B_1 \mid W) \leq \gamma$, and $\rho(G \mid W) \geq 1 - \gamma$. Hence, Item 2 holds.

Finally, it remains to bound the costs of $\rho_G$ conditioned on $G_b$ for Item 3. Since $G$ is the final $W_{\text{tmp}}$, which passes line 20, $\rho_G$ must satisfy

$$\rho_G(M_0) \geq \frac{\gamma}{r+3} \cdot \rho(M_0) \tag{23}$$

for $M_0$ with $\rho_G(M_0) > 0$,

$$\rho_G(X, \mathbf{M}) \geq \frac{\gamma}{r+3} \cdot \rho(X, \mathbf{M}) \tag{24}$$

for $(X, \mathbf{M})$ with $\rho_G(X, \mathbf{M}) > 0$,

$$\rho_G(Y, \mathbf{M}) \geq \frac{\gamma}{r+3} \cdot \rho(Y, \mathbf{M}) \tag{25}$$

for $(Y, \mathbf{M})$ with $\rho_G(Y, \mathbf{M}) > 0$,

$$\rho_G(M_i \mid X, M_{<i}) \geq \frac{\gamma}{r+3} \cdot \rho(M_i \mid X, M_{<i}) \tag{26}$$

for all odd $i \in [r]$ and $(X, M_{\le i})$ with $\rho_G(X, M_{\le i}) > 0$, and

$$\rho_G(M_i \mid Y, M_{<i}) \ge \frac{\gamma}{r+3} \cdot \rho(M_i \mid Y, M_{<i}) \tag{27}$$

for all even $i \in [r]$ and $(Y, M_{\le i})$ with $\rho_G(Y, M_{\le i}) > 0$.

The $\theta$-cost of $\rho_G$ conditioned on $G_b$ is

$$\begin{aligned}
&\log \theta_\nu(\rho_G \mid G_b) \\
&= \log \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\rho_G|G_b} \left[ \frac{\rho_G(X,Y,\mathbf{M})}{\rho_G(M_0)\cdot\nu(X,Y)\cdot\prod_{\text{odd } i\in[r]}\rho_G(M_i \mid X, M_{<i})\cdot\prod_{\text{even } i\in[r]}\rho_G(M_i \mid Y, M_{<i})} \right]
\end{aligned}$$

which by (23), (26) and (27), is

$$\begin{aligned}
&\le \log \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\rho|G_b} \left[ \left(\frac{r+3}{\gamma}\right)^{r+1} \cdot \frac{\rho(X,Y,\mathbf{M} \mid G)}{\rho(M_0)\cdot\nu(X,Y)\cdot\prod_{\text{odd } i\in[r]}\rho(M_i \mid X, M_{<i})\cdot\prod_{\text{even } i\in[r]}\rho(M_i \mid Y, M_{<i})} \right] \\
&\le \log \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\rho|G_b} \left[ \left(\frac{r+3}{\gamma}\right)^{r+1} \cdot \frac{\rho(X,Y,\mathbf{M})/\rho(G)}{\rho(M_0)\cdot\nu(X,Y)\cdot\prod_{\text{odd } i\in[r]}\rho(M_i \mid X, M_{<i})\cdot\prod_{\text{even } i\in[r]}\rho(M_i \mid Y, M_{<i})} \right] \\
&= \log \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\rho|G_b} [\theta_\nu(\rho @ X, Y, \mathbf{M})] + (r+1)\log\left((r+3)/\gamma\right) + \log\left(1/\rho(G)\right) \\
&= \log \theta_\nu(\rho \mid G_b) + (r+1)\log\left((r+3)/\gamma\right) + \log\left(1/\rho(G)\right)
\end{aligned}$$

which by Proposition 14 and the fact that $G_b \subseteq W$, is

$$\begin{aligned}
&\le \log \theta_\nu(\rho \mid W) + \log(1/\rho(G_b \mid W)) + (r+1)\log\left((r+3)/\gamma\right) + \log\left(1/\rho(G)\right) \\
&= \log \theta_\nu(\rho \mid W) + (r+1)\log\left((r+3)/\gamma\right) + \log(1/\rho(G_b)) + \log(1/\rho(G \mid W)) \\
&\le \log \theta_\nu(\rho \mid W) + (r+1)\log\left((r+3)/\gamma\right) + \log(1/((1-\gamma)\rho(G_b))).
\end{aligned}$$

For the $\chi^2$-cost by Alice, we have

$$\begin{aligned}
\log \chi^2_{\nu,A}(\rho_G \mid G_b) &= \log \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\rho_G|G_b} \left[ \frac{\rho_G(X \mid \mathbf{M}, Y)}{\nu(X \mid Y)} \right] \\
&= \log \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\rho|G_b} \left[ \frac{\rho(X,Y,\mathbf{M} \mid G)}{\rho_G(\mathbf{M}, Y)\nu(X \mid Y)} \right]
\end{aligned}$$

which by (25), is

$$\begin{aligned}
&\le \log \mathop{\mathbb{E}}_{(X,Y,\mathbf{M})\sim\rho|G_b} \left[ \frac{\rho(X,Y,\mathbf{M})/\rho(G)}{(\gamma/(r+3))\rho(\mathbf{M}, Y)\nu(X \mid Y)} \right] \\
&= \log \chi^2_{\nu,A}(\rho \mid G_b) + \log((r+3)/\gamma) + \log(1/\rho(G))
\end{aligned}$$

which by Proposition 17, is

$$\begin{aligned}
&\le \log \chi^2_{\nu,A}(\rho \mid W) + \log(1/\rho(G_b \mid W)) + \log((r+3)/\gamma) + \log(1/\rho(G)) \\
&\le \log \chi^2_{\nu,A}(\rho \mid W) + \log((r+3)/\gamma) + \log(1/((1-\gamma)\rho(G_b))).
\end{aligned}$$

Similarly, $\log \chi^2_{\nu,B}(\rho_G) \le \log \chi^2_{\nu,B}(\rho \mid W) + \log((r+3)/\gamma) + \log(1/((1-\gamma)\rho(G_b)))$. This proves the lemma. □

# 7 Compression of Generalized Protocols: Proof of Lemma 40

In this subsection, we design a standard protocol with lower communication from a generalized protocol with low costs. We will use the following lemma as a subroutine, whose proof is similar to Theorem 4.1 in [BR11]. The lemma lets the players sample from a distribution $P$ with low communication, where only Alice knows $P$, and Bob knows a different distribution $Q$. The success probability depends on how "close" the two distributions are.

**Lemma 43.** *Let $P, Q$ be two distributions over $\mathcal{U}$, such that Alice knows $P$ and Bob knows $Q$. For any $C > 0$ and $\delta \in (0, 1/2)$, there is a (standard) one-way communication protocol with shared public random bits, where Alice sends one message of $C + O(\log(1/\delta))$ bits to Bob. Then Alice and Bob simultaneously output an element in $\mathcal{U}$ such that Alice outputs $x$ with probability $P(x)$ for every $x \in \mathcal{U}$; conditioned on Alice outputting $x$, Bob outputs*

- *the same $x$ with probability* at least $\min\{1, 2^C \cdot Q(x)/P(x)\} - \delta$,
- *some different $x \in \mathcal{U}$ with probability* at most $\delta$,
- $\perp$ *otherwise.*

In particular, for each $x$ such that $P(x) \le 2^C Q(x)$, the players will agree on $x$ with probability at least $(1-\delta)P(x)$.

*Proof.* Let $t = \lceil 2\log(2/\delta) \rceil$. Consider the following protocol.

**Protocol** `sample`$(P; Q)$:

**Part I: Alice samples** $x$

1. Alice and Bob view the public random bits as a sequence of $2t \cdot |\mathcal{U}|$ *uniform* samples $(x_i, p_i)$ in $\mathcal{U} \times [0, 1]$
2. Alice finds the first pair $(x_i, p_i)$ such that $p_i \le P(x_i)$
3. if such pair does not exist
4. Alice outputs an $x \sim P$, and sends "0" to Bob // "0" indicates "fail"
5. upon receiving "0", Bob outputs $\perp$, and the protocol aborts
6. otherwise, Alice outputs $x_i$, and sends 1

(to be cont'd)

So far, the protocol describes how Alice samples $x$. Now, we show that Alice indeed samples $x$ according to $P$. Fix $x \in \mathcal{U}$, for each sample $(x_i, p_i)$, the probability that $x_i = x$ and $p_i \le P(x)$ is

$$\frac{1}{|\mathcal{U}|} \cdot P(x).$$

Summing over all possible $x$, the probability that $p_i \le P(x_i)$ is equal to $\frac{1}{|\mathcal{U}|}$. The probability that Alice outputs $x$ is

$$\begin{aligned}
&\sum_{i=1}^{2t\cdot|\mathcal{U}|} \left(1 - \frac{1}{|\mathcal{U}|}\right)^{i-1} \cdot \frac{1}{|\mathcal{U}|} \cdot P(x) \\
&= P(x) \cdot \left(1 - \left(1 - \frac{1}{|\mathcal{U}|}\right)^{2t\cdot|\mathcal{U}|}\right).
\end{aligned}$$

Note that $(1 - 1/|\mathcal{U}|)^{2t\cdot|\mathcal{U}|} \le e^{-2\log(2/\delta)} < \delta/2$. The probability that Alice finds a pair in step 2 is at least $1 - \delta/2$, and conditioned on finding such a pair, $x_i$ is distributed according to $P$. Next, Alice tries to inform Bob by hashing the index $i$.

**Part II: Bob outputs $x$**

7. the players view the remaining public random bits as a uniformly random hash function $h : [2t \cdot |\mathcal{U}|] \to \{0,1\}^{C+\lceil \log(2/\delta)+\log(2t)\rceil}$
8. Alice sends $v = h(i)$ to Bob
9. upon receiving $v$, Bob finds all $i \in [2t \cdot |\mathcal{U}|]$ such that $h(i) = v$ and $p_i \le 2^C \cdot Q(x_i)$
10. if there is only one such $i$
11. Bob outputs $x_i$
12. else
13. Bob outputs $\perp$

It is clear that the communication cost is at most

$$C + \log(1/\delta) + \log\log(1/\delta) + 7 \le C + O(\log(1/\delta))$$

bits.

Conditioned on Alice outputting $x_i$, $p_i$ is uniform in $[0, P(x_i)]$. Hence, the correct $i$ will be found in step 9 with probability $\min\{1, 2^C \cdot Q(x_i)/P(x_i)\}$. Next, we bound the probability that Bob finds any other $j \neq i$, conditioned on Alice outputting $x_i$.

For each $(x_j, p_j)$, if $j > i$, the probability that $p_j \le 2^C \cdot Q(x_j)$ is $\min\left\{1, 2^C \cdot Q(x_j)\right\} \le 2^C \cdot Q(x_j)$. If $j < i$, conditioned on Alice outputting $x_i$, $p_j$ is uniform in $(P(x_j), 1]$. The probability that $p_j \le 2^C \cdot Q(x_j)$ is

$$\frac{\max\{0, \min\left\{1, 2^C \cdot Q(x_j)\right\} - P(x_j)\}}{1 - P(x_j)} \le 2^C \cdot Q(x_j).$$

Independently, the probability that $h(j) = v$ is equal to $2^{-(C+\lceil \log(2/\delta)+\log(2t)\rceil)}$.

Therefore, the probability $(x_j, p_j)$ satisfies both conditions is at most

$$\sum_x \Pr[x_j = x] \cdot 2^{-(C+\lceil \log(2/\delta)+\log(2t)\rceil)} \cdot 2^C \cdot Q(x) < \frac{\delta}{4t \cdot |\mathcal{U}|}.$$

By union bound, the probability that any other $(x_j, p_j)$ satisfies both conditions is at most $\delta/2$.

To conclude, Bob outputs the same $x_i$ when

- Bob does not output $\perp$ in step 5 (with probability $\ge 1 - \delta/2$), and
- Bob finds the correct $i$ in step 9 (with probability $\min\{1, 2^C \cdot Q(x_i)/P(x_i)\}$), and
- Bob does not find any other $j \neq i$ in step 9 (with probability $\ge 1 - \delta/2$).

By union bound, Bob outputs the same $x_i$ with probability at least $\min\{1, 2^C \cdot Q(x_i)/P(x_i)\} - \delta$. Bob outputs some different $x_i$ *only* when

- Bob does not output $\perp$ in step 5, and
- Bob does not find the correct $i$ in step 9, and
- Bob find some other $j \neq i$ (and $x_j \neq x_i$) in step 9 (with probability $\le \delta/2$).

Bob outputs some different $x_i$ with probability at most $\delta/2$. Otherwise, Bob outputs $\perp$. This proves the lemma. □

We will use the above lemma to sample messages $M_i$ given $M_{<i}$. The next lemma proves that most of time, in Alice and Bob's view, the probabilities of $M_i$ are not too different.

**Lemma 44.** *Let $\rho$ be an $r$-round generalized protocol and $W$ be an event such that $(\rho \mid W)$ has the rectangle property with respect to $\mu$, and let $(X, Y, \mathbf{M}) \sim \rho \mid W$. Then for any $T > 1$, the probability that*

- *there exists an odd $i \in [r]$ such that*

$$\frac{\rho(M_i \mid X, M_{<i})}{\rho(M_i \mid Y, M_{<i})} > T \cdot \chi^2_{\mu,A}(\rho \mid W),$$

  *or*

- *there exists an even $i \in [r]$ such that*

$$\frac{\rho(M_i \mid Y, M_{<i})}{\rho(M_i \mid X, M_{<i})} > T \cdot \chi^2_{\mu,B}(\rho \mid W),$$

*is at most $6r \cdot T^{-1/5} \cdot \rho(W)^{-1}$.*

*Proof.* We first fix an odd $i \in [r]$, and upper bound the probability that $\frac{\rho(M_i|X,M_{<i})}{\rho(M_i|Y,M_{<i})} > T \cdot \chi^2_{\mu,A}(\rho \mid W)$. Recall that

$$\chi^2_{\mu,A}(\rho \mid W) = \mathop{\mathbb{E}}_{\rho|W}\left[\frac{\rho(X \mid Y, \mathbf{M})}{\mu(X \mid Y)}\right],$$

and we have

$$\begin{aligned}
\frac{\rho(X \mid Y, \mathbf{M})}{\mu(X \mid Y)} &= \frac{\rho(X \mid Y, \mathbf{M})}{\rho(X \mid Y)} \cdot \frac{\rho(X \mid Y)}{\mu(X \mid Y)} \\
&= \frac{\rho(\mathbf{M} \mid X, Y)}{\rho(\mathbf{M} \mid Y)} \cdot \frac{\rho(X \mid Y)}{\mu(X \mid Y)} \\
&= \frac{\rho(M_{<i} \mid X, Y)}{\rho(M_{<i} \mid Y)} \cdot \frac{\rho(M_i \mid M_{<i}, X, Y)}{\rho(M_i \mid M_{<i}, Y)} \cdot \frac{\rho(M_{>i} \mid M_{\le i}, X, Y)}{\rho(M_{>i} \mid M_{\le i}, Y)} \cdot \frac{\rho(X \mid Y)}{\mu(X \mid Y)} \\
&= \frac{\rho(M_i \mid M_{<i}, X)}{\rho(M_i \mid M_{<i}, Y)} \cdot \left(\frac{\rho(M_{<i} \mid X, Y)}{\rho(M_{<i} \mid Y)} \cdot \frac{\rho(M_i \mid M_{<i}, X, Y)}{\rho(M_i \mid M_{<i}, X)} \cdot \frac{\rho(M_{>i} \mid M_{\le i}, X, Y)}{\rho(M_{>i} \mid M_{\le i}, Y)} \cdot \frac{\rho(X \mid Y)}{\mu(X \mid Y)}\right) \\
&=: \frac{\rho(M_i \mid M_{<i}, X)}{\rho(M_i \mid M_{<i}, Y)} \cdot (F_1 \cdot F_2 \cdot F_3 \cdot F_4), \qquad (28)
\end{aligned}$$

where $F_1, F_2, F_3, F_4$ denote the four fractions in the parenthesis respectively. Note that the fraction outside the parenthesis is what we want to upper bound. We now show that $F_1, F_2, F_3, F_4$ are all not-too-small with high probability.

For $F_1$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\rho}[1/F_1] &= \mathop{\mathbb{E}}_{\rho}\left[\frac{\rho(M_{<i} \mid Y)}{\rho(M_{<i} \mid X, Y)}\right] \\
&= \sum_{X,Y,M_{<i}} \rho(X, Y, M_{<i}) \cdot \frac{\rho(M_{<i} \mid Y)}{\rho(M_{<i} \mid X, Y)} \\
&= \sum_{X,Y,M_{<i}} \rho(X, Y) \cdot \rho(M_{<i} \mid Y) \\
&= \sum_{X,Y} \rho(X, Y)
\end{aligned}$$

$$= 1.$$

Similarly, we can show that

$$\mathop{\mathbb{E}}_{\rho}[1/F_2] = \mathop{\mathbb{E}}_{\rho}[1/F_3] = 1,$$

and

$$\mathop{\mathbb{E}}_{\rho}[1/F_4] = \sum_{X,Y} \rho(X,Y) \cdot \frac{\mu(X \mid Y)}{\rho(X \mid Y)} = \sum_{X,Y} \rho(Y)\mu(X \mid Y) = 1.$$

Since $F_1, F_2, F_3, F_4$ are all nonnegative, by Markov's inequality, we have

$$\Pr_{\rho}\left[1/F_j \geq \delta^{-1}\right] \leq \delta,$$

for $j = 1, 2, 3, 4$ and any $\delta \in (0, 1)$. Thus, $\Pr_{\rho|W}\left[1/F_j \geq \delta^{-1}\right] \leq \delta/\rho(W)$.

By union bound and plugging into (28), we have

$$\Pr_{\rho|W}\left[\frac{\rho(M_i \mid M_{<i}, X)}{\rho(M_i \mid M_{<i}, Y)} \geq \delta^{-4} \cdot \frac{\rho(X \mid Y, \mathbf{M})}{\mu(X \mid Y)}\right] \leq 4\delta/\rho(W).$$

Thus by union bound over all odd $i \in [r]$, we have

$$\Pr_{\rho|W}\left[\exists \text{odd}\, i \in [r], \frac{\rho(M_i \mid M_{<i}, X)}{\rho(M_i \mid M_{<i}, Y)} \geq \delta^{-4} \cdot \frac{\rho(X \mid Y, \mathbf{M})}{\mu(X \mid Y)}\right] \leq 4\lceil r/2 \rceil \cdot \delta/\rho(W).$$

By Markov's inequality again, we have

$$\Pr_{\rho|W}\left[\frac{\rho(X \mid Y, \mathbf{M})}{\mu(X \mid Y)} \geq \delta^{-1} \cdot \chi^2_{\mu,A}(\rho \mid W)\right] \leq \delta.$$

Combining the two inequalities, we obtain

$$\Pr_{\rho|W}\left[\exists \text{odd}\, i \in [r], \frac{\rho(M_i \mid M_{<i}, X)}{\rho(M_i \mid M_{<i}, Y)} \geq \delta^{-5} \cdot \chi^2_{\mu,A}(\rho \mid W)\right] \leq 4\lceil r/2 \rceil \cdot \delta/\rho(W) + \delta.$$

Similarly, for even $i$, we can prove that

$$\Pr_{\rho|W}\left[\exists \text{even}\, i \in [r], \frac{\rho(M_i \mid M_{<i}, Y)}{\rho(M_i \mid M_{<i}, X)} \geq \delta^{-5} \cdot \chi^2_{\mu,B}(\rho \mid W)\right] \leq 4\lfloor r/2 \rfloor \cdot \delta/\rho(W) + \delta.$$

Finally, by setting $\delta = T^{-1/5}$ and applying a union bound on the odd and the even case, the probability is at most

$$4r \cdot T^{-1/5}/\rho(W) + 2T^{-1/5} \leq 6rT^{-1/5}/\rho(W).$$

This proves the lemma. □

We will also use the following lemma in the proof.

**Lemma 45.** *Let $\rho$ be an $r$-round generalized protocol and $W$ be an event such that $(\rho \mid W)$ has the rectangle property with respect to $\mu$, and let $(X, Y, \mathbf{M}) \sim \rho \mid W$. Then for any $T > 1$, the probability that*

$$\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})} > T \cdot \theta_\mu(\rho \mid W),$$

*or*

$$\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})} < 1/T,$$

*is at most $2 \cdot T^{-1} \cdot \rho(W)^{-1}$.*

*Proof.* The first half is bounded using an application of Markov's inequality and the fact that $\frac{\rho(X,Y,\mathbf{M},W)}{\rho(X,Y,\mathbf{M})} \leq 1$:

$$\Pr_{\rho|W}\left[\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) > T \cdot \theta_\mu(\rho \mid W)\right] < 1/T,$$

implying that

$$\Pr_{\rho|W}\left[\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})} > T \cdot \theta_\mu(\rho \mid W)\right] < 1/T.$$

For the second half, similar to the proof of Lemma 44, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\rho|W}\left[\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M})^{-1} \cdot \frac{\rho(X, Y, \mathbf{M})}{\rho(X, Y, \mathbf{M}, W)}\right] \\
&= \sum_{X,Y,\mathbf{M}} \rho(X, Y, \mathbf{M} \mid W) \cdot \frac{\rho(M_0) \cdot \mu(X, Y) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i})}{\rho(X, Y, \mathbf{M}, W)} \\
&= \frac{1}{\rho(W)} \cdot \sum_{X,Y,\mathbf{M}} \rho(M_0) \cdot \mu(X, Y) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}) \\
&= \frac{1}{\rho(W)} \cdot \sum_{X,Y,M_{<r}} \rho(M_0) \cdot \mu(X, Y) \cdot \prod_{\text{odd } i \in [r-1]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r-1]} \rho(M_i \mid Y, M_{<i}) \\
&= \cdots \\
&= \frac{1}{\rho(W)} \cdot \sum_{X,Y,M_0} \rho(M_0) \cdot \mu(X, Y) \\
&= \frac{1}{\rho(W)}.
\end{aligned}$$

Hence, by Markov's inequality,

$$\Pr_{\rho}\left[\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M})^{-1} \cdot \frac{\rho(X, Y, \mathbf{M})}{\rho(X, Y, \mathbf{M}, W)} > T\right] < T^{-1} \cdot \rho(W)^{-1}.$$

We prove the lemma by an application of the union bound. □

Finally, we are ready to prove Lemma 40.

**Lemma 40 (restated).** *Let $\delta_1, \delta_2 \in (0, 1/2)$ be any fixed parameter. Let $\rho$ be an $r$-round generalized protocol and let $W$ be an event such that $(\rho \mid W)$ has the rectangle property with respect to $\mu$. Then for any function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, there is an $r$-round* standard *protocol $\tau$ such that*

- *in odd rounds of $\tau$, Alice sends a message of at most* $\log \chi^2_{\mu,A}(\rho \mid W) + O(\log(r/\delta_1\delta_2) + \log\log\theta_\mu(\rho \mid W))$ *bits;*
- *in even rounds of $\tau$, Bob sends a message of at most* $\log \chi^2_{\mu,B}(\rho \mid W) + O(\log(r/\delta_1\delta_2))$ *bits;*
- *$\tau$ computes $f$ correctly under input distribution $\mu$ with probability at least*

$$\frac{1}{2} + \frac{\delta_1}{32\theta_\mu(\rho \mid W)} \left( \rho(W) \cdot \mathop{\mathbb{E}}_{\rho|W} [\mathrm{adv}_\rho(f(X,Y) \mid X, \mathbf{M}, W)] - 6\delta_1 \right) - 2r\delta_2.$$

*Proof.* Let us first consider the following "ideal protocol" $\tau^*$ that cannot necessarily be implemented in the standard communication setting. But we can still analyze the probability that $\tau^*$ computes $f(X,Y)$. Then we construct a standard protocol $\tau$ with low communication and statistically close to $\tau^*$ when $(X,Y)$ is sampled from $\mu$.

The ideal protocol consists of two parts: In the first part, the players generate a transcript $\mathbf{M}$ given the inputs $(X,Y)$; in the second part, they use rejection sampling, and accept $\mathbf{M}$ with some carefully chosen probability (and output a random bit if they reject).

**"Ideal protocol"** $\tau^*(X;Y)$:

**Part I**

1. Alice and Bob use public random bits to sample $M_0$ from $\rho(M_0)$
2. for $i = 1, \ldots, r-1$
3. if $i$ is odd, Alice samples $M_i$ from $\rho(M_i \mid X, M_{<i})$ and sends it to Bob
4. if $i$ is even, Bob samples $M_i$ from $\rho(M_i \mid Y, M_{<i})$ and sends it to Alice
5. Bob locally samples $M_r$ from $\rho(M_r \mid Y, M_{<r})$ // recall that $M_r \in \{0,1\}$
6. for $j = 0, 1$
7. Alice examines the distribution of $\rho(f(X,Y) \mid X, \mathbf{M}, W)$ pretending $M_r = j$
8. Alice sends Bob the more likely value $p_j \in \{0,1\}$ of $f(X,Y)$ in this conditional distribution

**Part II**

9. Alice and Bob accept $\mathbf{M}$ with probability equal to $\gamma \cdot \theta_\rho(\mu \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X,Y,\mathbf{M},W)}{\rho(X,Y,\mathbf{M})}$ (assuming it is at most 1), for some fixed parameter $\gamma$
10. if the players decide to accept
11. Bob sends $p_{M_r}$
12. else
13. Bob sends a random bit

**Success probability of $\tau^*$.** Given $X, Y$, Alice and Bob generate transcript $\mathbf{M}$ with probability

$$\rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}),$$

where $M_r$ is only known to Bob. Then it is accepted with probability $\gamma \cdot \theta_\rho(\mu \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X,Y,W|\mathbf{M})}{\rho(X,Y|\mathbf{M})}$. Recall that

$$\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) = \frac{\rho(X,Y,\mathbf{M})}{\mu(X,Y) \cdot \rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i})}.$$

Hence, for $(X,Y) \sim \mu$, the probability that the players generate and accept $(X,Y,\mathbf{M})$ is

$$\mu(X,Y) \cdot \rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}) \cdot \gamma \cdot \theta_\rho(\mu \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X,Y,\mathbf{M},W)}{\rho(X,Y,\mathbf{M})}$$

$= \gamma \cdot \rho(X, Y, \mathbf{M}, W)$.

Thus, the probability that $\tau^*$ accepts is $\gamma \cdot \rho(W)$. Alice does not know $M_r$, so she sends the more likely value $p_j$ conditioned on $(X, \mathbf{M})$ for both possibilities of $M_r$, and Bob outputs this bit when they accept.

Since conditioned on accepting, $(X, Y, \mathbf{M})$ follows the distribution of $\rho(X, Y, \mathbf{M} \mid W)$, and Alice has told Bob in advance what is the more like value of $f(X, Y)$ conditioned on $(X, \mathbf{M}, W)$. Intuitively, this should imply that the overall advantage should be $\gamma \cdot \rho(W) \cdot \mathbb{E}_{\rho|W}[\mathrm{adv}(f(X, Y) \mid X, \mathbf{M}, W)]$. We now formally prove that this holds. We use $\tau^*(\mathbf{R})$ to denote the probability of $\mathbf{R}$ in the distribution induced by running $\tau^*$ on $(X, Y) \sim \mu$. Note that the transcript of $\tau^*$ is $(M_{<r}, p_0, p_1, p)$. The expected overall advantage of $\tau^*$ is at least

$$
\begin{aligned}
&\sum_{M_{<r},p} \tau^*(M_{<r}, p) \cdot |2\tau^*(f(X, Y) = 1 \mid M_{<r}, p) - 1| \\
&\quad \geq \sum_{M_{<r},p} \tau^*(M_{<r}, p) \cdot (2\tau^*(f(X, Y) = p \mid M_{<r}, p) - 1) \\
&\quad = 2 \sum_{M_{<r},p} \tau^*(f(X, Y) = p, M_{<r}, p) - 1 \\
&\quad = \left( 2 \sum_{M_{<r},p} \tau^*(f(X, Y) = p, M_{<r}, p, \text{accept}) \right) + \left( 2 \sum_{M_{<r},p} \tau^*(f(X, Y) = p, M_{<r}, p, \text{reject}) - 1 \right).
\end{aligned}
$$

The first term is

$$
\begin{aligned}
&2 \sum_{X,Y,\mathbf{M},p=f(X,Y)} \tau^*(X, Y, \mathbf{M}, p_{M_r} = p, \text{accept}) \\
&= 2 \sum_{X,Y,\mathbf{M}:p_{M_r}=f(X,Y)} \gamma \cdot \rho(X, Y, \mathbf{M}, W) \\
&= 2 \sum_{X,\mathbf{M}} \gamma \cdot \rho(X, \mathbf{M}, W) \cdot \rho(p_{M_r} = f(X, Y) \mid X, \mathbf{M}, W)
\end{aligned}
$$

which by the fact that $p_{M_r}$ is the more likely value of $f(X, Y)$ conditioned on $(X, \mathbf{M}, W)$, is

$$
\begin{aligned}
&= 2 \sum_{X,\mathbf{M}} \gamma \cdot \rho(X, \mathbf{M}, W) \cdot \left( \frac{1}{2} + \frac{1}{2} \cdot \mathrm{adv}_\rho(f(X, Y) \mid X, \mathbf{M}, W) \right) \\
&= \gamma \cdot \rho(W) + \gamma \cdot \rho(W) \cdot \mathop{\mathbb{E}}_{\rho|W} \left[ \mathrm{adv}_\rho(f(X, Y) \mid X, \mathbf{M}, W) \right].
\end{aligned}
$$

The second term is equal to

$$
\begin{aligned}
&2 \sum_{M_{<r},p} \frac{1}{2} \cdot \tau^*(f(X, Y) = p, M_{<r}, \text{reject}) - 1 \\
&\quad = -\tau^*(\text{accept}) \\
&\quad = -\gamma \cdot \rho(W).
\end{aligned}
$$

The two terms sum up to $\gamma \cdot \rho(W) \cdot \mathbb{E}_{\rho|W} \left[ \mathrm{adv}_\rho(f(X, Y) \mid X, \mathbf{M}, W) \right]$.

Hence, we have proved the following claim.

**Claim 46.** *If the probability that a protocol generates and accepts a triple $(X, Y, \mathbf{M})$ is equal to $\gamma \cdot \rho(X, Y, \mathbf{M}, W)$, and it outputs a random bit otherwise, then this protocol computes $f(X, Y)$ correctly with probability at least*

$$\frac{1}{2} + \frac{\gamma}{2} \cdot \rho(W) \cdot \mathop{\mathbb{E}}_{\rho|W} \left[\mathrm{adv}_\rho(f(X, Y) \mid X, \mathbf{M}, W)\right].$$

**Standard protocol $\tau$.** Next, we will construct a standard protocol $\tau$ that simulates $\tau^*$. Similar to $\tau^*$, protocol $\tau$ also has two parts: In the first part, the players generate a transcript $\mathbf{M}$; in the second part, they decide if they will accept $\mathbf{M}$.

For the first part, the players first use public randomness to sample $M_0$. Then for the subsequent messages $M_i$, Alice knows the distribution $\rho(M_i \mid X, M_{<i})$, and Bob knows the distribution $\rho(M_i \mid Y, M_{<i})$. For odd $i$, the players use Lemma 43 to sample from $\rho(M_i \mid X, M_{<i})$ where Alice sends a message; for even $i$, they sample from $\rho(M_i \mid Y, M_{<i})$ with Bob sending a message. Finally, Bob locally samples the last message $M_r$. We will show that Lemma 44 guarantees that the probability of sampling $\mathbf{M}$ is a good approximation of

$$\rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}).$$

**Protocol** $\tau(X; Y)$:

**Part I**

1. fix parameters $\delta_1, \delta_2 \in (0, 1/2)$
2. Alice and Bob use public random bits to sample $M_0$ from $\rho(M_0)$
3. for $i = 1, \ldots, r-1$
4. for odd $i$, use Lemma 43 to sample $M_i$ from $\rho(M_i \mid X, M_{<i})$ given that Bob only knows $\rho(M_i \mid Y, M_{<i})$, where we set $C := C_0 = \log \chi^2_{\mu,A}(\rho \mid W) + 5\log(6r/\delta_1)$ and $\delta := \delta_2$ // Alice sends one message
5. for even $i$, use Lemma 43 to sample $M_i$ from $\rho(M_i \mid Y, M_{<i})$ given that Alice only knows $\rho(M_i \mid X, M_{<i})$, where we set $C := C_1 = \log \chi^2_{\mu,B}(\rho \mid W) + 5\log(6r/\delta_1)$ and $\delta := \delta_2$ // Bob sends one message
6. in Bob's local memory: $\texttt{acc} \leftarrow 1$ // the final value of `acc` indicates if they will accept
7. if any player outputs $\perp$ in any round
8. $\texttt{acc} \leftarrow 0$
9. Bob *locally* samples $M_r$ from $\rho(M_r \mid Y, M_{<r})$ (to be cont'd)

Each player will send one extra bit indicating whether they output $\perp$ in the previous round. Hence, Bob knows if any player outputs $\perp$ in the first $r-1$ rounds (including round $r-1$, for which he does not need to send the extra bit).

Next, we use rejection sampling, and accept $\mathbf{M}$ with probability roughly $\gamma \cdot \theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X,Y,\mathbf{M},W)}{\rho(X,Y,\mathbf{M})}$ for some carefully chosen $\gamma > 0$. The rectangle property of $(\rho \mid W)$ ensures that this rejection sampling can be done approximately using very little communication.

More specifically, by the rectangle property of $(\rho \mid W)$ with respect to $\mu$ (see Definition 18), there exists $g_1, g_2$ such that $\rho(X, Y, \mathbf{M} \mid W) = \mu(X, Y) \cdot g_1(X, \mathbf{M}) \cdot g_2(Y, \mathbf{M})$. Hence, $\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X,Y,\mathbf{M},W)}{\rho(X,Y,\mathbf{M})}$ can be written as

$$\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})} = g_A(X, \mathbf{M}) \cdot g_B(Y, \mathbf{M}),$$

by letting $g_A(X, \mathbf{M}) := \frac{\rho(W) \cdot g_1(X,\mathbf{M})}{\prod_{\text{odd } i \in [r]} \rho(M_i|X,M_{<i})}$ and $g_B(Y, \mathbf{M}) := \frac{g_2(Y,\mathbf{M})}{\rho(M_0) \cdot \prod_{\text{even } i \in [r]} \rho(M_i|Y,M_{<i})}$.

Suppose we let Alice accept with probability $\gamma_A \cdot g_A(X, \mathbf{M})$ and Bob accept with probability $\gamma_B \cdot g_B(Y, \mathbf{M})$ for $\gamma_A\gamma_B = \gamma$, then they will be able to accept with the correct probability by sending only one bit, i.e., whether they accept locally. We will also need to choose $\gamma_A$ and $\gamma_B$ carefully so that both probabilities are at most one. This is done by applying Lemma 45, which ensures that most of the time $\theta_\mu(\rho \mathbin{@} X, Y, \mathbf{M}) \cdot \frac{\rho(X,Y,\mathbf{M},W)}{\rho(X,Y,\mathbf{M})}$ is between $\delta$ and $\theta_\mu(\rho \mid W)/\delta$ for small $\delta > 0$. Thus, they can coordinate the values of $\gamma_A$ and $\gamma_B$ by Alice sending a small hash value of some approximation of $g_A(X, \mathbf{M})$.

**Part II**

10. for $j = 0, 1$, Alice computes $g_A(X, \mathbf{M})$ pretending $M_r = j$, and computes $e_{A,j} := \lfloor \log g_A(X, \mathbf{M}) \rfloor$
    Bob computes $g_B(Y, \mathbf{M})$ and $e_B := \lfloor \log g_B(Y, \mathbf{M}) \rfloor$
11. let $R := \lceil \log(32\theta_\mu(\rho \mid W)/\delta_1^2)/\delta_2 \rceil$, Alice and Bob use public random bits to sample a hash function $h : \mathbb{Z} \to [R]$
12. for $j = 0, 1$
13. Alice samples a bit $b_{A,j}$ such that $\Pr[b_{A,j} = 1] = g_A(X, \mathbf{M}) \cdot 2^{-(e_{A,j}+1)}$ pretending $M_r = j$
14. Alice examines the distribution of $\rho(f(X, Y) \mid X, \mathbf{M}, W)$ pretending $M_r = j$
15. Alice sets $p_j \in \{0, 1\}$ to the more likely value of $f(X, Y)$ in this conditional distribution
16. Alice appends $(h(e_{A,0}), h(e_{A,1}), b_{A,0}, b_{A,1}, p_0, p_1)$ to her last message $M_{r-1}$

17. let $L_1 := \lceil \log(4/\delta_1) \rceil$ and $L_2 := \lceil \log(\theta_\mu(\rho \mid W)/\delta_1) \rceil$
18. upon receiving $(v_0, v_1, b_{A,0}, b_{A,1}, p_0, p_1)$, Bob checks:
    if there is one *unique* integer $e'_A \in [-e_B - L_1, -e_B + L_2]$ such that $h(e'_A) = v_{M_r}$
19. Bob samples a bit $b_B$ such that $\Pr[b_B = 1] = g_B(Y, \mathbf{M}) \cdot 2^{e'_A - L_2 - 1}$
20. set $\mathtt{acc} \leftarrow 0$ if there is no such $e'_A$, or $e'_A$ is not unique, or $b_{A,M_r} = 0$, or $b_B = 0$
21. if $\mathtt{acc} = 1$
22. Bob sends $p_{M_r}$
23. else
24. Bob sends a random bit

Note that Alice's new messages are sent before Bob starts sending the last message, hence, it is still part of round $r - 1$. In step 13, since $g_A(X, \mathbf{M}) < 2^{e_{A,j}+1}$ for $j = 0, 1$, the probability is at most 1. In step 19, since $e'_A - L_2 - 1 \le -(e_B + 1)$, the probability is also at most 1. Hence, the protocol is well-defined.

**Communication cost.** By Lemma 43, in odd rounds, Alice sends a message of length at most

$$\log \chi^2_{\mu,A}(\rho \mid W) + 5 \log(6r/\delta_1) + O(\log(1/\delta_2)),$$

in even rounds, Bob sends a message of length at most

$$\log \chi^2_{\mu,B}(\rho \mid W) + 5 \log(6r/\delta_1) + O(\log(1/\delta_2)).$$

They also send one extra bit indicating if the lemma outputs $\perp$ in the previous round. In Alice's last message (round $r-1$), Alice further sends two hash values $h(e_{A,0}), h(e_{A,1})$ and the bits $b_{A,0}, b_{A,1}, p_0, p_1$, which takes at most

$$2\lceil \log R \rceil + 4 \le O(\log \log \theta_\mu(\rho \mid W) + \log \log(1/\delta_1) + \log(1/\delta_2))$$

bits in total. This proves the communication bound of $\tau$ we claimed.

**The first part of $\tau$.** We first analyze the first part of $\tau$ and estimate the probability that we generate a triple $(X, Y, \mathbf{M})$. By Lemma 44, for $(X, Y, \mathbf{M}) \sim \rho \mid W$, the probability that (recall the value of $C_0$ in line 4 and the value of $C_1$ in line 5 of $\tau$)

- there exists an odd $i \in [r]$ such that

$$\frac{\rho(M_i \mid X, M_{<i})}{\rho(M_i \mid Y, M_{<i})} > 2^{C_0},$$

or

- there exists an even $i \in [r]$ such that

$$\frac{\rho(M_i \mid Y, M_{<i})}{\rho(M_i \mid X, M_{<i})} > 2^{C_1},$$

is at most $6r \cdot \left(2^{5\log(6r/\delta_1)}\right)^{-1/5} \cdot \rho(W)^{-1} = \delta_1 \cdot \rho(W)^{-1}$. Denote this set of $(X, Y, \mathbf{M})$ by $\mathcal{B}_1$, hence, $\rho(\mathcal{B}_1 \mid W) \le \delta_1 \cdot \rho(W)^{-1}$.

Now consider any $(X, Y, \mathbf{M}) \notin \mathcal{B}_1$, and we estimate the probability that $\mathbf{M}$ is generated by the players given $X, Y$. By Lemma 43, conditioned on $(X, Y, M_{<i})$, for odd $i \in [r]$, the probability that both players agree on $M_i$ in step 4 is at least

$$\left(\min\left\{1, 2^{C_0} \cdot \frac{\rho(M_i \mid Y, M_{<i})}{\rho(M_i \mid X, M_{<i})}\right\} - \delta_2\right) \rho(M_i \mid X, M_{<i}) \ge (1-\delta_2)\rho(M_i \mid X, M_{<i}),$$

as $\frac{\rho(M_i|X,M_{<i})}{\rho(M_i|Y,M_{<i})} \le 2^{C_0}$ for $(X, Y, \mathbf{M}) \notin \mathcal{B}_1$. Similarly, for even $i \in [r]$, both players agree on $M_i$ in step 5 is at least

$$(1-\delta_2)\rho(M_i \mid Y, M_{<i}).$$

Bob generates the last message $M_r$ with probability $\rho(M_r \mid Y, M_{<r})$. Thus, conditioned on $(X, Y)$, the probability that the players generate and agree on $\mathbf{M}$ is at least

$$(1-(r-1)\delta_2)\rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}),$$

where we used the fact that $(1-\delta_2)^{r-1} \ge (1-(r-1)\delta_2)$.

On the other hand, for *all* $(X, Y, \mathbf{M})$ (not necessarily in $\overline{\mathcal{B}_1}$), the probability that the players agree on $M_i$ is at most $\rho(M_i \mid X, M_{<i})$ for odd $i \in [r]$ (since this is the probability that Alice outputs $M_i$ by Lemma 43), and $\rho(M_i \mid Y, M_{<i})$ for even $i \in [r]$. Thus, the probability that they agree on $\mathbf{M}$ is at most

$$\rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}).$$

Also, by Lemma 43 and union bound, the probability that the players do not agree on the same $M_i$ in some round is at most $(r-1)\delta_2$. Otherwise, some player outputs $\perp$, and `acc` is set to 0. Thus, we obtain the following claim.

**Claim 47.** *There is a set $\mathcal{B}_1$ such that $\rho(\mathcal{B}_1 \mid W) \le \delta_1 \cdot \rho(W)^{-1}$, and given $(X, Y)$, the protocol $\tau$ generates $\mathbf{M}$ in Part I of $\tau$ with probability at most*

$$\rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i});$$

*furthermore, if $(X, Y, \mathbf{M}) \notin \mathcal{B}_1$, $\tau$ generates $\mathbf{M}$ with probability at least*

$$(1-(r-1)\delta_2)\rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}).$$

*Moreover, the probability that the players do not agree on the same $\mathbf{M}$ is at most $(r-1)\delta_2$.*

**The second part of $\tau$.** Consider a triple $(X, Y, \mathbf{M})$, we analyze the probability that it is accepted in the second part, *conditioned on* it being generated in the first part. Alice does not know $M_r$, but it has only two possible values. Alice pretends that $M_r = j$ for $j = 0, 1$, and computes the corresponding $g_A(X, \mathbf{M}), e_{A,j}, b_{A,j}$ and $p_j$. She sends both copies (for $j = 0, 1$) to Bob, and Bob only looks at the copy corresponding to the actual $M_r$. In terms of the correctness, this is equivalent to Alice knowing $M_r$. For simplicity of notations, we will omit the subscript $j$, and use $g_A(X, \mathbf{M}), e_A, b_A, p$ to denote the copy for the actual $M_r$.

If for a triple $(X, Y, \mathbf{M})$, we have $-L_1 \le e_A + e_B \le L_2$ (note that $e_A, e_B$ are determined by the triple), then the probability that there is a *unique* integer $e'_A \in [-e_B - L_1, -e_B + L_2]$ such that $h(e'_A) = h(e_A)$ is *equal to*

$$(1 - 1/R)^{L_1+L_2},$$

and in this case, we must have $e'_A = e_A$. Then the probability that $b_A = 1$ is

$$g_A(X, \mathbf{M}) \cdot 2^{-(e_A+1)},$$

and the probability that $b_B = 1$ is

$$g_B(Y, \mathbf{M}) \cdot 2^{e_A - L_2 - 1}.$$

The players do not set `acc` to 0 in Part II with probability

$$\begin{aligned}
&(1 - 1/R)^{L_1+L_2} \cdot g_A(X, \mathbf{M}) \cdot 2^{-(e_A+1)} \cdot g_B(Y, \mathbf{M}) \cdot 2^{e_A - L_2 - 1} \\
&= (1 - 1/R)^{L_1+L_2} \cdot 2^{-L_2-2} \cdot \theta_\mu(\rho \circledcirc X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})}.
\end{aligned}$$

On the other hand, if for a triple $(X, Y, \mathbf{M})$, either $e_A + e_B < -L_1$ or $e_A + e_B > L_2$, then the probability that the players accept it conditioned on it being generated is *at most* the probability that there exists some $e'_A$ that matches the hash value of $e_A$, which by union bound, is at most

$$\begin{aligned}
(L_1 + L_2 + 1)/R &= (\lceil \log(4/\delta_1) \rceil + \lceil \log(\theta_\mu(\rho \mid W)/\delta_1) \rceil + 1)/\lceil \log(32\theta_\mu(\rho \mid W)/\delta_1^2)/\delta_2 \rceil \\
&\le (\log(4\theta_\mu(\rho \mid W)/\delta_1^2) + 3)/(\log(32\theta_\mu(\rho \mid W)/\delta_1^2)/\delta_2) \\
&= \delta_2.
\end{aligned}$$

We denote the set of $(X, Y, \mathbf{M})$ such that either $e_A + e_B < -L_1$ or $e_A + e_B > L_2$ by $\mathcal{B}_2$. If $e_A + e_B > L_2$, then we have

$$\begin{aligned}
&\log\left(\theta_\mu(\rho \circledcirc X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})}\right) \\
&= \log(g_A(X, \mathbf{M}) g_B(Y, \mathbf{M})) \\
&\ge e_A + e_B \\
&> L_2 \\
&> \log(\theta_\mu(\rho \mid W)/\delta_1).
\end{aligned}$$

If $e_A + e_B < -L_1$, then we have

$$\log\left(\theta_\mu(\rho \circledcirc X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})}\right)$$

$$
\begin{aligned}
&= \log(g_A(X, \mathbf{M}) g_B(Y, \mathbf{M})) \\
&< e_A + e_B + 2 \\
&< -L_1 + 2 \\
&\leq \log \delta_1.
\end{aligned}
$$

However, by Lemma 45, for $(X, Y, \mathbf{M}) \sim \rho \mid W$, the probability that

$$\theta_\mu(\rho @ X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})} > \theta_\mu(\rho \mid W)/\delta_1,$$

or

$$\theta_\mu(\rho @ X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})} < \delta_1$$

is at most $2\delta_1 \cdot \rho(W)^{-1}$. This implies that $\rho(\mathcal{B}_2 \mid W) \leq 2\delta_1 \cdot \rho(W)^{-1}$. Hence, we obtain the following claim.

**Claim 48.** *There is a set $\mathcal{B}_2$ such that $\rho(\mathcal{B}_2 \mid W) \leq 2\delta_1 \cdot \rho(W)^{-1}$, and for $(X, Y, \mathbf{M}) \notin \mathcal{B}_2$, the probability that $\tau$ accepts $(X, Y, \mathbf{M})$ conditioned on $\tau$ generating $(X, Y, \mathbf{M})$ is* equal to

$$(1 - 1/R)^{L_1+L_2} \cdot 2^{-L_2-2} \cdot \theta_\mu(\rho @ X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})},$$

*for $(X, Y, \mathbf{M}) \in \mathcal{B}_2$, the probability that $\tau$ accepts $(X, Y, \mathbf{M})$ conditioned on $\tau$ generating $(X, Y, \mathbf{M})$ is* at most $\delta_2$.

**Overall success probability.** If all $(X, Y, \mathbf{M})$ were generated in the first part with probability equal to

$$\rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}),$$

and accepted in the second part with probability equal to

$$(1 - 1/R)^{L_1+L_2} \cdot 2^{-L_2-2} \cdot \theta_\mu(\rho @ X, Y, \mathbf{M}) \cdot \frac{\rho(X, Y, \mathbf{M}, W)}{\rho(X, Y, \mathbf{M})},$$

then $\tau$ would be the ideal protocol in Claim 46 for $\gamma = (1 - 1/R)^{L_1+L_2} \cdot 2^{-L_2-2}$. Hence, to lower bound the overall success probability, it suffices to compare $\tau$ with $\tau^*$, and bound the total probability difference in generating and accepting a triple $(X, Y, \mathbf{M})$.

By Claim 47 and Claim 48, for $(X, Y, \mathbf{M}) \notin \mathcal{B}_1 \cup \mathcal{B}_2$, the probability that it is generated and accepted is at most

$$(1 - 1/R)^{L_1+L_2} \cdot 2^{-L_2-2} \cdot \rho(X, Y, \mathbf{M}, W) = \gamma \cdot \rho(X, Y, \mathbf{M}, W),$$

and at least

$$(1 - (r-1)\delta_2) \cdot (1 - 1/R)^{L_1+L_2} \cdot 2^{-L_2-2} \cdot \rho(X, Y, \mathbf{M}, W) \geq (1 - (r-1)\delta_2)\gamma \cdot \rho(X, Y, \mathbf{M}, W).$$

Hence, the total probability difference between $\tau$ and $\tau^*$ for these $(X, Y, \mathbf{M})$ is at most

$$\sum_{(X,Y,\mathbf{M}) \notin \mathcal{B}_1 \cup \mathcal{B}_2} (r-1)\delta_2\gamma \cdot \rho(X, Y, \mathbf{M}, W) \leq (r-1)\delta_2\gamma \cdot \rho(W). \tag{29}$$

For $(X, Y, \mathbf{M}) \in \mathcal{B}_1 \setminus \mathcal{B}_2$, the probability that it is generated and accepted is at most

$$\gamma \cdot \rho(X, Y, \mathbf{M}, W).$$

Hence, the total probability difference for these $(X, Y, \mathbf{M})$ is at most

$$\sum_{(X,Y,\mathbf{M}) \in \mathcal{B}_1 \setminus \mathcal{B}_2} \gamma \cdot \rho(X, Y, \mathbf{M}, W) \leq \gamma \cdot \rho(\mathcal{B}_1 \mid W) \cdot \rho(W) \leq \gamma\delta_1. \tag{30}$$

For $(X, Y, \mathbf{M}) \in \mathcal{B}_2$, the probability that it is generated and accepted is at most

$$\delta_2 \cdot \rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}).$$

Hence, the total probability difference for these $(X, Y, \mathbf{M})$ is at most

$$\begin{aligned}
&\sum_{(X,Y,\mathbf{M}) \in \mathcal{B}_2} \max \left\{ \delta_2 \cdot \rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}),\ \gamma \cdot \rho(X, Y, \mathbf{M}, W) \right\} \\
&\leq \sum_{(X,Y,\mathbf{M}) \in \mathcal{B}_2} \left( \delta_2 \cdot \rho(M_0) \cdot \prod_{\text{odd } i \in [r]} \rho(M_i \mid X, M_{<i}) \cdot \prod_{\text{even } i \in [r]} \rho(M_i \mid Y, M_{<i}) + \gamma \cdot \rho(X, Y, \mathbf{M}, W) \right) \\
&\leq \delta_2 + \gamma \cdot \rho(\mathcal{B}_2 \mid W) \cdot \rho(W) \\
&\leq \delta_2 + 2\gamma\delta_1.
\end{aligned} \tag{31}$$

Finally, the players do not agree on the same $\mathbf{M}$ with probability at most $(r-1)\delta_2$.

Summing up Equation (29), (30), (31) and the probability that they do no agree, the statistical distance between $\tau$ and $\tau^*$ is at most $3\gamma\delta_1 + 2r\delta_2$, where we used the fact that $\gamma \leq 1$ and $\rho(W) \leq 1$. Combining it with Claim 46, we obtain that $\tau$ computes $f$ correctly with probability at least

$$\frac{1}{2} + \frac{\gamma}{2} \cdot \left( \rho(W) \cdot \mathop{\mathbb{E}}_{\rho|W} \left[ \mathrm{adv}_\rho(f(X, Y) \mid X, \mathbf{M}, W) \right] - 6\delta_1 \right) - 2r\delta_2.$$

Finally, note that if $\rho(W) \cdot \mathbb{E}_\rho \left[ \mathrm{adv}_\rho(f(X, Y) \mid X, \mathbf{M}) \right] - 6\delta_1 < 0$, then the lemma holds trivially by setting $\tau$ to the protocol that outputs a random bit, otherwise, we have

$$\begin{aligned}
\gamma &= (1 - 1/R)^{L_1 + L_2} \cdot 2^{-L_2 - 2} \\
&\geq (1 - (L_1 + L_2)/R) \cdot 2^{-L_2 - 2} \\
&\geq (1 - \delta_2) \cdot \frac{1}{8\theta_\mu(\rho \mid W)/\delta_1} \\
&\geq \frac{\delta_1}{16\theta_\mu(\rho \mid W)}.
\end{aligned}$$

Hence, the success probability is as claimed in the statement. This finishes the proof of the lemma. □

# References


[AN21] Sepehr Assadi and Vishvajeet N. Graph streaming lower bounds for parameter estimation and property testing via a streaming XOR lemma. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 612–625. ACM, 2021.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013.

[BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[BKLS20] Joshua Brody, Jae Tak Kim, Peem Lerdputtipongporn, and Hariharan Srinivasulu. A strong XOR lemma for randomized query complexity. *CoRR*, abs/2007.05580, 2020.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In Rafail Ostrovsky, editor, *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 748–757. IEEE Computer Society, 2011.

[BRWY13a] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part I*, volume 7965 of *Lecture Notes in Computer Science*, pages 232–243. Springer, 2013.

[BRWY13b] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 746–755. IEEE Computer Society, 2013.

[BW15] Mark Braverman and Omri Weinstein. An interactive information odometer and applications. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 341–350. ACM, 2015.

[CKP+21] Lijie Chen, Gillat Kol, Dmitry Paramonov, Raghuvansh R. Saxena, Zhao Song, and Huacheng Yu. Almost optimal super-constant-pass streaming lower bounds for reachability. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 570–583. ACM, 2021.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 270–278. IEEE Computer Society, 2001.

[Dru12] Andrew Drucker. Improved direct product theorems for randomized query complexity. *Comput. Complex.*, 21(2):197–244, 2012.

[GNW11] Oded Goldreich, Noam Nisan, and Avi Wigderson. On yao's xor-lemma. In Oded Goldreich, editor, *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation - In Collaboration with Lidor Avigad, Mihir Bellare, Zvika Brakerski, Shafi Goldwasser, Shai Halevi, Tali Kaufman, Leonid Levin, Noam Nisan, Dana Ron, Madhu Sudan, Luca Trevisan, Salil Vadhan, Avi Wigderson, David Zuckerman*, volume 6650 of *Lecture Notes in Computer Science*, pages 273–301. Springer, 2011.

[Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *36th Annual Symposium on Foundations of Computer Science, Milwaukee, Wisconsin, USA, 23-25 October 1995*, pages 538–545. IEEE Computer Society, 1995.

[IW97] Russell Impagliazzo and Avi Wigderson. $P = BPP$ if $E$ requires exponential circuits: Derandomizing the XOR lemma. In Frank Thomson Leighton and Peter W. Shor, editors, *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA, May 4-6, 1997*, pages 220–229. ACM, 1997.

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 167–176. IEEE Computer Society, 2012.

[Lev87] Leonid A. Levin. One-way functions and pseudorandom generators. *Comb.*, 7(4):357–363, 1987.

[Sha03] Ronen Shaltiel. Towards proving strong direct product theorems. *Comput. Complex.*, 12(1-2):1–22, 2003.

[She11] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 41–50. ACM, 2011.

[VW08] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory Comput.*, 4(1):137–168, 2008.

[Yao82] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *23rd Annual Symposium on Foundations of Computer Science, Chicago, Illinois, USA, 3-5 November 1982*, pages 80–91. IEEE Computer Society, 1982.

# A Theorem 2 Implies Shaltiel's XOR Lemma

Recall that the discrepancy of a function $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$ is[7]

$$\mathrm{disc}(f) := \frac{1}{|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot \max_{\mathbf{R}=\mathbf{X}\times\mathbf{Y}:\mathbf{X}\subseteq\mathcal{X},\mathbf{Y}\subseteq\mathcal{Y}} \left\{ \left| \sum_{x\in\mathbf{X},y\in\mathbf{Y}} f(x,y) \right| \right\}.$$

[7]For $\{0, 1\}$-valued functions, we map the value to $\{-1, 1\}$ then apply this definition.

Shaltiel's XOR lemma states that

$$\mathrm{disc}(f^{\oplus n}) = O(\mathrm{disc}(f))^{\Omega(n)}.$$

We show that Theorem 2 implies this bound. First by viewing a protocol with $C$ bits of communication as a partitioning of $\mathcal{X} \times \mathcal{Y}$ into $2^C$ combinatorial rectangles, any $C$-bit communication protocol cannot compute $f$ with probability better than

$$\frac{1}{2} + 2^C \cdot \mathrm{disc}(f),$$

when the inputs are sampled from the uniform distribution $\mu$ over $\mathcal{X} \times \mathcal{Y}$. In particular, it applies to 2-round protocols, and we obtain that

$$\mathrm{suc}_\mu\left(f; \frac{1}{4}\log(1/\mathrm{disc}(f)), \frac{1}{4}\log(1/\mathrm{disc}(f)), 2\right) \leq \frac{1}{2} + \mathrm{disc}(f)^{1/2}.$$

Thus, for $\mathrm{disc}(f)$ smaller than a sufficiently small constant (otherwise, the bound holds trivially), Theorem 2 with $C_A = C_B = \frac{1}{4}\log(1/\mathrm{disc}(f))$, $\alpha = \mathrm{disc}(f)^{1/16c} (\geq 2\mathrm{disc}(f)^{1/2})$ and $r = 2$ implies that no 2-round protocol with communication $O(n\log(1/\mathrm{disc}(f)))$ solves $f^{\oplus n}$ with probability

$$\frac{1}{2} + \mathrm{disc}(f)^{\Omega(n)}.$$

In particular, no protocol with *two* bits of communication can solve $f^{\oplus n}$ with this probability.

Finally, as pointed out in Remark 3.12 in [VW08], the discrepancy of a function is equal to the maximum advantage over $1/2$ that a 2-bit protocol can achieve on the uniform distribution (up to a constant factor). This proves that $\mathrm{disc}(f^{\oplus n}) = O(\mathrm{disc}(f))^{\Omega(n)}$.



ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il