A Direct Product Theorem for One-Way Quantum Communication

Rahul Jain ✉
Centre for Quantum Technologies & Department of Computer Science,
National University of Singapore, Singapore
Majulab, UMI 3654, Singapore

Srijita Kundu ✉
Centre for Quantum Technologies, National University of Singapore, Singapore

Abstract

We prove a direct product theorem for the one-way entanglement-assisted quantum communication complexity of a general relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. For any $0 < \varepsilon < \delta < \frac{1}{2}$ and any $k \geq 1$, we show that

$$\mathrm{Q}^1_{1-(1-\varepsilon)^{\Omega(k/\log|\mathcal{Z}|)}}(f^k) = \Omega\left(k \cdot \mathrm{Q}^1_\delta(f)\right),$$

where $\mathrm{Q}^1_\varepsilon(f)$ represents the one-way entanglement-assisted quantum communication complexity of f with worst-case error ε and f^k denotes k parallel instances of f.

As far as we are aware, this is the first direct product theorem for the quantum communication complexity of a general relation – direct sum theorems were previously known for one-way quantum protocols for general relations, while direct product theorems were only known for special cases. Our techniques are inspired by the parallel repetition theorems for the entangled value of two-player non-local games, under product distributions due to Jain, Pereszlényi and Yao [24], and under anchored distributions due to Bavarian, Vidick and Yuen [4], as well as message compression for quantum protocols due to Jain, Radhakrishnan and Sen [29]. In particular, we show that a direct product theorem holds for the distributional one-way quantum communication complexity of f under any distribution q on $\mathcal{X} \times \mathcal{Y}$ that is anchored on one side, i.e., there exists a y^* such that $q(y^*)$ is constant and $q(x|y^*) = q(x)$ for all x. This allows us to show a direct product theorem for general distributions, since for any relation f and any distribution p on its inputs, we can define a modified relation $\tilde{f}$ which has an anchored distribution q close to p, such that a protocol that fails with probability at most ε for $\tilde{f}$ under q can be used to give a protocol that fails with probability at most $\varepsilon + \zeta$ for f under p.

Our techniques also work for entangled non-local games which have input distributions anchored on any one side, i.e., either there exists a y^* as previously specified, or there exists an x^* such that $q(x^*)$ is constant and $q(y|x^*) = q(y)$ for all y. In particular, we show that for any game $G = (q, \mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}, \mathsf{V})$ where q is a distribution on $\mathcal{X} \times \mathcal{Y}$ anchored on any one side with constant anchoring probability, then

$$\omega^*(G^k) = \left(1 - (1 - \omega^*(G))^5\right)^{\Omega\left(\frac{k}{\log(|\mathcal{A}| \cdot |\mathcal{B}|)}\right)}$$

where $\omega^*(G)$ represents the entangled value of the game G. This is a generalization of the result of [4], who proved a parallel repetition theorem for games anchored on both sides, i.e., where both a special x^* and a special y^* exist, and potentially a simplification of their proof.

2012 ACM Subject Classification Theory of computation → Communication complexity; Theory of computation → Quantum complexity theory

Keywords and phrases Direct product theorem, parallel repetition theorem, quantum communication, one-way protocols, communication complexity

Digital Object Identifier 10.4230/LIPIcs.CCC.2021.27

Related Version *Full Version*: https://arxiv.org/abs/2008.08963

36th Computational Complexity Conference (CCC 2021).
Editor: Valentine Kabanets; Article No. 27; pp. 27:1–27:28

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Funding This work is supported by the National Research Foundation, including under NRF RF Award No. NRF-NRFF2013-13, the Prime Minister's Office, Singapore and the Ministry of Education, Singapore, under the Research Centres of Excellence program and by Grant No. MOE2012-T3-1-009 and in part by the NRF2017-NRF-ANR004 *VanQuTe* Grant.

1 Introduction

A fundamental question in complexity theory is: given k independent instances of a function or relation, does computing them require k times the amount of resources required to compute a single instance of the function or relation? Suppose solving one instance of some problem with success probability at least $1-\varepsilon$ requires c units of some resource. A natural way to solve k independent instances of this problem would be to solve them independently, which requires ck units of the resource. A *direct sum theorem* for this problem would state that any algorithm for solving k instances which uses $o(ck)$ units of resource has success probability at most $1-\varepsilon$. A *direct product theorem* for the problem would state that any algorithm for solving k instances that uses $o(ck)$ units of resource has success probability at most $(1-\varepsilon)^{\Omega(k)}$. Hence a direct product theorem is the stronger result of the two.

In this paper, we deal with direct product theorems in the model of communication complexity. In this model, there are two parties Alice and Bob, who receive inputs x and y respectively, and wish to jointly compute a relation f. They can use local computation, public coins, and communicate with each other using classical messages, in the classical model; use local unitaries, shared entanglement, and communicate with each other using quantum messages, in the quantum model. The resource of interest is the number of bits/qubits communicated; so the parties are allowed to share an arbitrary amount of randomness or entanglement, and perform local operations of arbitrary complexity.

Direct product theorems in communication are related to *parallel repetition theorems* for *non-local games*. In a non-local game, two parties Alice and Bob are given inputs x and y respectively from some specified distribution, and without communicating with each other, they are required to give answers a and b respectively to a referee. They are considered to win the game if $\mathsf{V}(a,b,x,y)$ holds for a specified predicate V. In the classical model, the players are allowed to share randomness, and in the quantum model they are allowed to share entanglement. A parallel repetition theorem shows that the maximum probability of winning k independent instances of a non-local game is $p^{\Omega(k)}$, if the maximum probability of winning a single instance of it is p, regardless of the amount of shared randomness or entanglement used. Direct product theorems in communication are often proved by combining techniques used to prove direct sum theorems in communication, which require message compression, and parallel repetition theorems for games.

In classical communication complexity, there is a long line of works on direct sum and direct-product theorems including [40, 14, 1, 41, 27, 28, 30, 5, 38, 44, 22, 21, 18, 35, 32, 2, 12, 11, 10, 7, 13, 20, 25, 37, 9, 43]. A parallel repetition theorem for the classical value of general two-player non-local games was first shown by Raz [39], and the proof was subsequently simplified by Holenstein [19].

In quantum communication complexity, a direct sum theorem is known for the entanglement-assisted one-way [30], *simultaneous-message-passing* (SMP), entanglement-assisted [30] and unassisted models [21]. A strong parallel repetition theorem for the quantum value of a general two-player non-local game is not known. Parallel repetition theorems were shown for special classes of games such as XOR games [15], unique games [34] and projection games [17]. When the type of game is not restricted but the input distribution is,

parallel repetition theorems have been shown under product distributions [24] and *anchored* distributions [4, 3]. For general games under general distributions, the best current result is due to Yuen [46], which shows that the quantum value of k parallel instances of a general game goes down polynomially in k, if the quantum value of the original game is strictly less than 1. No direct product theorems for quantum communication for a general function had previously been known. However, a direct product theorem has been shown for the generalized discrepancy method [42], which is a lower bound technique that often characterizes (multi-round) quantum communication complexity. [5] showed a direct product theorem for functions whose one-way quantum communication is characterized by VC dimension, and [36] showed a direct product theorem for symmetric functions.

Combining ideas from Jain, Pereszlényi and Yao [24] and the message compression scheme from Jain, Radhakrishnan and Sen [30], it is possible to show a strong direct product theorem for one-way quantum communication under product distributions. To deal with non-product distributions, we borrow the idea of anchored distributions due to Bavarian, Vidick and Yuen [4, 3], which allows us to prove a direct product theorem for the worst case one-way quantum communication complexity of a general function. We make some crucial changes in the definition of correlation-breaking random variable as used by [4] which help us use one-sided anchored distribution and simplify their proof. This simplification is in fact crucial for us to combine the anchored distribution technique with the message compression argument of [30] in the communication complexity setting. We elaborate further on our proof techniques in Section 1.2.

Parallel repetition and direct product theorems have a number of applications. For example, Raz's parallel repetition theorem [39] can be used to prove the PCP theorem [16]; the [4] parallel repetition theorem was used to prove the recent MIP* = RE result [33]. Sherstov's direct product theorem for generalized discrepancy was used in [8] to prove a near-optimal lower bound on the bounded-round quantum communication complexity of set disjointness. [36] used their direct product theorem to prove time-space tradeoffs for solving certain problems. We expect our result to have similar applications.

1.1 Our results

Let $\mathrm{Q}^1_\varepsilon(f)$ denote that the one-way entanglement-assisted quantum communication complexity of a relation f, with worst-case error ε. Let f^k denote k parallel instances of f. Our strong direct product theorem is as follows.

▶ **Theorem 1.** *For any relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, and any $0 < \varepsilon, \zeta < \frac{1}{2}$,*

$$\mathrm{Q}^1_{1-(1-\varepsilon)^{\Omega(\zeta^6 k/\log|\mathcal{Z}|)}}(f^k) = \Omega\left(k\left(\zeta^5 \cdot \mathrm{Q}^1_{\varepsilon+\zeta}(f) - \log\log(1/\zeta)\right)\right).$$

Let $\omega^*(G)$ represent the entangled value of a two-player non-local game G, and let G^k denote k parallel instances of G. We call a distribution q on $\mathcal{X} \times \mathcal{Y}$ *anchored on one side* with *anchoring probability* ζ if one of the following conditions holds:

(i) There exists an $x^* \in \mathcal{X}$ such that $q(x^*) = \zeta$ and $q(y|x^*) = q(y)$ for all $y \in \mathcal{Y}$,

(ii) There exists a $y^* \in \mathcal{Y}$ such that $q(y^*) = \zeta$ and $q(x|y^*) = q(x)$ for all $x \in \mathcal{X}$.

The game will be called *anchored on both sides* with anchoring probability ζ if both conditions hold simultaneously.

Then our parallel repetition theorem is stated as follows.

▶ **Theorem 2.** *For a two-player non-local game $G = (q, \mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}, \mathsf{V})$ such that q is a distribution anchored on one side with anchoring probability ζ,*

$$\omega^*(G^k) = \left(1 - (1 - \omega^*(G))^5\right)^{\Omega\left(\frac{\zeta^2 k}{\log(|\mathcal{A}|\cdot|\mathcal{B}|)}\right)}.$$

One can get a game anchored on one side (say the $\mathcal{Y}$ side) from a general game in the following way: in the anchored game, the referee chooses (x, y) from the original probability distribution, and with probability ζ replaces y with a new input y^*. If Bob's input is y^*, then the referee accepts any answer from the players. In a game anchored on both sides, the referee must instead replace x with x^* and y with y^* independently with probability ζ, and accept if either Alice's input is x^* or Bob's input is y^*. It is clear that anchoring makes the game easier. In this light, a parallel repetition theorem for anchoring games can be thought of as follows: for a general game G, there exists a simple transformation taking it to another game $\tilde{G}$ such that

1. If $\omega^*(G) = 1$, then $\omega^*(\tilde{G}^k) = 1$.
2. If $\omega^*(G) < 1$, then $\omega^*(\tilde{G}^k) = \exp(-\Omega(k))$.

The merit of our result here is that the transformation involved for anchoring on one side changes the game less than the transformation involved in anchoring it on both sides.

We note that the definition of anchoring used in [4, 3] is more general: instead of single inputs x^*, y^*, they consider anchoring sets $\mathcal{X}^* \subseteq \mathcal{X}$ and $\mathcal{Y}^* \subseteq \mathcal{Y}$, such that $q(\mathcal{X}^*), q(\mathcal{Y}^*) \geq \zeta$, and whenever $x \in \mathcal{X}^*$ or $y \in \mathcal{Y}^*$, $q(x, y) = q(x)q(y)$. However, it appears this generalized definition is not more useful from the perspective of anchoring transformations. While our technique could go through for the one-sided version of this definition of anchoring, we do not state or prove it as such for the sake of simplicity.

Unlike in the case of communication, worst-case success probability is usually not considered for non-local games. But one could define a game $G_{\text{wc}} = (\mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}, \mathsf{V})$ without an associated distribution, and the worst-case winning probability ω^*_{wc} over all inputs of this can be considered. As long as Alice and Bob are allowed to share randomness (which they are, in the quantum case), Yao's lemma [45] holds just like in the case of communication, relating the worst-case winning probability to distributional winning probability. Hence, by choosing $\zeta = (1 - \omega^*_{\text{wc}}(G_{\text{wc}}))/2$ and using the same arguments as in the case of communication, Theorem 2 leads to the following corollary about the worst-case winning probability of any game.

▶ **Corollary 3.** *For any two-player non-local game $G_{\text{wc}} = (\mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}, \mathsf{V})$,*

$$\omega^*_{\text{wc}}(G^k_{\text{wc}}) = \left(1 - (1 - \omega^*_{\text{wc}}(G_{\text{wc}}))^7\right)^{\Omega\left(\frac{k}{\log(|\mathcal{A}|\cdot|\mathcal{B}|)}\right)}.$$

This is in fact also implied by the result of [4], although it is not explicitly observed by them.

1.2 Proof overview

We describe how to prove the parallel repetition and direct product theorems in the distributional setting first, and we shall later describe how to go from there to the worst case setting. We use the information theoretic framework for parallel repetition established by [39] and [19].The broad idea is as follows: for a given relation $\tilde{f} \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, let the one-way quantum communication required to compute a single copy with constant success be c. Now consider a one-way quantum protocol $\mathcal{P}$ for $\tilde{f}^k$ which has communication $o(ck)$, in which we can condition on the success of some t coordinates. If the success probability in these

t coordinates is already as small as we want, then we are done. Otherwise, we exhibit a $(t+1)$-th coordinate i, such that conditioned on the success on the t coordinates, the success of i in $\mathcal{P}$ is bounded away from 1. This is done by showing that if the success probability in the t coordinates is not too small, then we can give a protocol $\mathcal{P}'$ for $\tilde{f}$ whose communication is $o(c)$ and whose success probability is constant – a contradiction.

$\mathcal{P}'$ works by embedding its input into the i-th coordinate of a shared quantum state representing the final input, output, message and discarded registers of $\mathcal{P}$, conditioned on the success event in the t coordinates, which we denote by $\mathcal{E}$. Suppose the quantum state conditioned on $\mathcal{E}$, when Alice and Bob's inputs are x_i and y_i respectively at the i-th coordinates, is $|\varphi\rangle_{x_i y_i}$. On input (x_i, y_i) in $\mathcal{P}'$, Alice and Bob will by means of local unitaries and communication try to get the shared state close to $|\varphi\rangle_{x_i y_i}$, on which Bob can perform a measurement to get an outcome z_i. The state $|\varphi\rangle_{x_i y_i}$ is such that the resulting probability distribution $\mathsf{P}_{X_i Y_i Z_i}$ is the distribution of $X_i Y_i Z_i$ in $\mathcal{P}$ conditioned on success. Hence our proof mainly consists of showing how Alice and Bob can get the shared state close to $|\varphi\rangle_{x_i y_i}$. The proof technique for a parallel repetition theorem is the same, except one cannot, and need not, use communication to get the shared state $|\varphi\rangle_{x_i y_i}$ there.

1.2.1 Product distribution parallel repetition

In [24] the following three states are considered: $|\varphi\rangle_{x_i}$ which is the superposition of $|\varphi\rangle_{x_i y_i}$ over the distribution of Y_i, $|\varphi\rangle_{y_i}$ which is the superposition over the distribution of X_i, and $|\varphi\rangle$ which is the superposition over both. In this setting, $X_1 \dots X_k$ are initially in product with all of Bob's registers and $Y_1 \dots Y_k$ are in product with all of Alice's registers. If the probability of $\mathcal{E}$ is large, then conditioning on it, the following can be shown:

1. By chain rule of mutual information, there is an X_i whose mutual information with Bob's registers in $|\varphi\rangle$ is small. Hence by Uhlmann's theorem, there exist unitaries U_{x_i} acting on Alice's registers that take $|\varphi\rangle$ close to $|\varphi\rangle_{x_i}$.
2. Similarly, the mutual information between Y_i and Alice's registers in $|\varphi\rangle$ is small, and hence there exist unitaries U_{y_i} acting on Bob's registers that take $|\varphi\rangle$ close to $|\varphi\rangle_{y_i}$.
3. Since U_{x_i} and U_{y_i} act on disjoint registers, using a commuting argument and the monotonicity of ℓ_1 distance under quantum operations, $U_{x_i} \otimes U_{y_i}$ takes $|\varphi\rangle$ close to $|\varphi\rangle_{x_i y_i}$.

Alice and Bob can thus share $|\varphi\rangle$ as entanglement, and get close to $|\varphi\rangle_{x_i y_i}$ by local operations.

1.2.2 Product distribution direct product

It is possible to combine techniques from the product parallel repetition theorem above and a message compression technique from [30] to give a direct product theorem for one-way quantum communication complexity under product distributions, and we give a proof outline here.

If the communication protocol involves a message from Alice to Bob, we cannot then get the state $|\varphi\rangle_{x_i y_i}$ by applying Uhlmann unitaries on both Alice and Bob's registers: because of Alice's message, the dependence of $|\varphi\rangle_{x_i y_i}$ on x_i can be quite large. Instead, we use the result of [26, 30] to do the transformation from $|\varphi\rangle$ to $|\varphi\rangle_{x_i}$ on Alice's side via a projector instead. By [30], as long as $|\varphi\rangle$ is the superposition of $|\varphi\rangle_{x_i}$ over the X_i distribution, such a projector Π_{x_i} always exists and its success probability depends on the mutual information between X_i and Bob's registers. This success probability is not close to 1, but as long as it is not too small, Alice and Bob can share multiple copies of $|\varphi\rangle$ and Alice can perform the $\{\Pi_{x_i}, \mathbb{1} - \Pi_{x_i}\}$ measurement on all of them. With high probability, she succeeds on at least one copy, and her message to Bob is then just the index of the copy she succeeds on.

Overall, the steps analogous to the parallel repetition proof are as follows:

1. If the message size in $\mathcal{P}$ is $o(ck)$ bits, by the chain rule of mutual information, the information between X_i and Bob's registers is $o(c)$. Hence by [30], there exist projectors Π_{x_i} acting on Alice's registers, which succeed with probability $2^{-o(c)}$ on $|\varphi\rangle$, and on success, take $|\varphi\rangle$ close to $|\varphi\rangle_{x_i}$.
2. Since there is no communication from Bob to Alice, by the same argument as in the case for games, there exist unitaries U_{y_i} acting on Bob's registers, that take $|\varphi\rangle$ close to $|\varphi\rangle_{y_i}$.
3. By the same commuting argument, conditioned on the success of Π_{x_i}, $\Pi_{x_i} \otimes U_{y_i}$ takes $|\varphi\rangle$ close to $|\varphi\rangle_{x_i y_i}$.

Hence there is a communication protocol with prior shared entanglement between Alice and Bob to obtain a state close to $|\varphi\rangle_{x_i y_i}$ on inputs (x_i, y_i): Alice and Bob share $2^{o(c)}$ copies of $|\varphi\rangle_{y^*}$ as entanglement; Alice performs the Π_{x_i} measurement on all these copies, and succeeds on at least one copy with high probability. She sends the index of the copy on which she succeeds to Bob, who performs U_{y_i} on the same copy. This protocol has communication $o(c)$, since that is how many classical bits Alice needs in order to encode the index of the successful copy out of $2^{o(c)}$ copies.

1.2.3 Anchored distribution parallel repetition

[3] in their parallel repetition theorem use anchored distributions, which are non-product distributions that "look like" product distributions. However, since overall $X_1 \dots X_k$ are not initially in product with $Y_1 \dots Y_k$, one needs to use what are known as *correlation-breaking variables*. For each i, correlation-breaking variables $D_i G_i$ are such that conditioned on $D_i G_i$, X_i and Y_i are independent. In particular, D_i is a uniformly distributed bit, and G_i takes values in either $\mathcal{X}$ or $\mathcal{Y}$ depending on whether D_i is 0 or 1, and is highly correlated with either X_i or Y_i in the respective cases. This means that conditioned on $D_i = 0$, $G_i = x^*$ with probability $\Omega(\zeta)$ and conditioned on $D_i = 1$, $G_i = y^*$ with probability $\Omega(\zeta)$.

1. The mutual information between X_i and Bob's registers in $|\varphi\rangle$ conditioned on $D_i = 1$ and G_i is small. Further conditioning on $G_i = y^*$ (which happens with constant probability), the mutual information between X_i and Bob's registers in $|\varphi\rangle_{y^*}$ is small. Hence by Uhlmann's theorem, there exist unitaries U_{x_i} on Alice's registers, taking $|\varphi\rangle_{x^* y^*}$ close to $|\varphi\rangle_{x_i y^*}$.
2. Similarly, the mutual information between Y_i and Alice's registers in $|\varphi\rangle$ conditioning on $D_i = 0$ and $G_i = x^*$ is small, which means there exist unitaries U_{y_i} on Bob's registers, taking $|\varphi\rangle_{x^* y^*}$ close to $|\varphi\rangle_{x^* y_i}$.
3. Using an involved argument, it is possible to show that $U_{x_i} \otimes U_{y_i}$ takes $|\varphi\rangle_{x^* y^*}$ close to $|\varphi\rangle_{x_i y_i}$.

Alice and Bob can thus share $|\varphi\rangle_{x^* y^*}$ in this case, and get close to $|\varphi\rangle_{x_i y_i}$ by local operations.

1.2.4 Anchored distribution direct product

In our direct product proof, since the distribution is anchored on one side, we use correlation-breaking variables that are identical to those in [3] in the $D_i = 1$ case, but in the $D_i = 0$ we consider a simpler distribution where G_i is perfectly correlated with X_i. Here we also clarify what we mean by G_i and Y_i being highly correlated when $D_i = 1$: if $G_i = y^*$, then Y_i is always y^*; but if $G_i = y_i$ for $y_i \neq y^*$, then Y_i still takes value y^* with probability $\Omega(\zeta)$, and is y_i otherwise. The distribution of X_i conditioned on $G_i = y^*$ is the marginal distribution of X_i, while conditioned on y_i, it is the same as the distribution of X_i conditioned on $Y_i = y_i$ (potentially different from the marginal distribution of X_i). Our use of these correlation-breaking variables is quite different from that in [3], however.

1. If the message size is $o(ck)$, the mutual information between X_i and Bob's registers in $|\varphi\rangle$ is $o(c)$, conditioned on $D_i = 1, G_i = y^*$. Since the distribution is anchored on Bob's side, this means that the mutual information between X_i and Bob's registers in $|\varphi\rangle_{y^*}$ is $o(c)$. By [30], there exist projectors Π_{x_i} acting on Alice's registers, which succeed with probability $2^{-o(c)}$ on $|\varphi\rangle_{y^*}$, and on success take it close to $|\varphi\rangle_{x_i y^*}$.
2. The mutual information between Y_i and Alice's registers conditioned on $D_i = 1, G_i \neq y^*$ is small. For each value of $G_i \neq y^*$, there exist only two possible values of Y_i: y_i and y^*, and hence Alice's registers in $|\varphi\rangle_{y_i}$ and $|\varphi\rangle_{y^*}$ must be close on average. By Uhlmann's theorem, there exist unitaries U_{y_i} acting on Bob's registers, taking $|\varphi\rangle_{y^*}$ close to $|\varphi\rangle_{y_i}$.
3. Since the marginal distribution of X_i conditioned on $G_i = y_i$ is approximately the same as the marginal distribution of X_i conditioned on $Y_i = y_i$, we can show by the same commuting argument that conditioned on success of Π_{x_i}, $\Pi_{x_i} \otimes U_{y_i}$ takes $|\varphi\rangle_{y^*}$ close to $|\varphi\rangle_{x_i y_i}$.

Hence there is a communication protocol with prior shared entanglement which allows Alice and Bob to obtain a state close to $|\varphi\rangle_{x_i y_i}$ as a shared state on input (x_i, y_i): this works just like the communication protocol for the product case, except the initial shared entanglement is $2^{o(c)}$ copies of $|\varphi\rangle_{y^*}$ instead. We note that our step 3 above is the simpler argument used in [24] and the product distribution direct product, instead of the more involved technique from [4].

1.2.5 Simplified anchored distribution parallel repetition

Our anchored distribution parallel repetition proof is the same as the anchored direct product proof, except no communication is necessary, since there was no communication in the original protocol. Instead of a projector on Alice's registers taking $|\varphi\rangle_{y^*}$ close to $|\varphi\rangle_{x_i y^*}$, in this case we will have a unitary U_{x_i} doing it. We can argue identically to the direct product proof that there exist U_{y_i} taking $|\varphi\rangle_{y^*}$ close to $|\varphi\rangle_{y_i}$, and $U_{x_i} \otimes U_{y_i}$ takes $|\varphi\rangle_{y^*}$ close to $|\varphi\rangle_{x_i y_i}$.

Our simplification of the techniques [4] is crucial to our direct product proof: we need to use the commuting argument from [30, 24] in order to make use of the message compression scheme. It is not clear whether the involved argument in [4] for the existence of $U_{x_i} \otimes V_{y_i}$ that takes $|\varphi\rangle_{x^* y^*}$ to $|\varphi\rangle_{x_i y_i}$ can work when there needs to be a projector rather than a unitary on Alice's side.

1.2.6 From anchored distribution to worst case direct product

The above argument proves a direct product theorem for the distributional one-way quantum communication complexity of under anchored distributions. However, what we are actually interested in is a direct product theorem for the worst case one-way quantum communication complexity. To get this for a relation f, we consider the distribution under which the distributional communication complexity is equal to the worst case communication complexity of f – this is guaranteed to exist by Yao's lemma. We do an anchoring transformation on f with this distributon to get $\tilde{f}$ with an anchored distribution. Note that it is fine if we can lower bound the distributional communication complexity of $\tilde{f}^k$ with success probability $(1-\varepsilon)^{\Omega(k)}$ under an anchored distribution by k times the worst case communication complexity of f with success probability δ. This is because f^k is harder than $\tilde{f}^k$, and the worst case communication complexity of $\tilde{f}^k$ is lower bounded by its distributional communication complexity under any distribution. By the argument described above, we can lower bound the distributional communication complexity of $\tilde{f}^k$ under the k-tensored anchored distribution with success probability $(1-\varepsilon)^{\Omega(k)}$ by k times the distributional communication complexity of $\tilde{f}$ under

the anchored distribution. Now it is easy to go from a distributional protocol for $\tilde{f}$ under the anchored distribution to a protocol for f under the original hard distribution decreasing the success probability by only $O(\zeta)$, since the anchoring transformation only disturbs the original distribution by this amount.

2 Preliminaries

2.1 Probability theory

We shall denote the probability distribution of a random variable X on some set $\mathcal{X}$ by P_X. For any event $\mathcal{E}$ on $\mathcal{X}$, the distribution of X conditioned on $\mathcal{E}$ will be denoted by $\mathsf{P}_{X|\mathcal{E}}$. For joint random variables XY, $\mathsf{P}_{X|Y=y}(x)$ is the conditional distribution of X given $Y = y$; when it is clear from context which variable's value is being conditioned on, we shall often shorten this to $\mathsf{P}_{X|y}$. We shall use $\mathsf{P}_{XY}\mathsf{P}_{Z|X}$ to refer to the distribution

$$(\mathsf{P}_{XY}\mathsf{P}_{Z|X})(x,y,z) = \mathsf{P}_{XY}(x,y) \cdot \mathsf{P}_{Z|X=x}(z).$$

For two distributions P_X and $\mathsf{P}_{X'}$ on the same set $\mathcal{X}$, the ℓ_1 distance between them is defined as

$$\|\mathsf{P}_X - \mathsf{P}_{X'}\|_1 = \sum_{x\in\mathcal{X}} |\mathsf{P}_X(x) - \mathsf{P}_{X'}(x)|.$$

▶ **Fact 4.** *For joint distributions P_{XY} and $\mathsf{P}_{X'Y'}$ on the same sets,*

$$\|\mathsf{P}_X - \mathsf{P}_{X'}\|_1 \leq \|\mathsf{P}_{XY} - \mathsf{P}_{X'Y'}\|_1.$$

▶ **Fact 5.** *For two distributions P_X and $\mathsf{P}_{X'}$ on the same set and an event $\mathcal{E}$ on the set,*

$$|\mathsf{P}_X(\mathcal{E}) - \mathsf{P}_{X'}(\mathcal{E})| \leq \frac{1}{2}\|\mathsf{P}_X - \mathsf{P}_{X'}\|_1.$$

▶ **Fact 6.** *For two distributions P_X and $\mathsf{P}_{X'}$ on the same set, and any joint distribution $\mathsf{P}_{XX'}$ whose marginals are P_X and $\mathsf{P}_{X'}$ respectively, we have*

$$\|\mathsf{P}_X - \mathsf{P}_{X'}\|_1 \leq 2\mathsf{P}_{XX'}(X \neq X').$$

▶ **Fact 7.** *Suppose probability distributions $\mathsf{P}_X, \mathsf{P}_{X'}$ satisfy $\|\mathsf{P}_X - \mathsf{P}_{X'}\|_1 \leq \varepsilon$, and an event $\mathcal{E}$ satisfies $\mathsf{P}_X(\mathcal{E}) \geq \alpha$, where $\alpha > \varepsilon$. Then,*

$$\|\mathsf{P}_{X|\mathcal{E}} - \mathsf{P}_{X'|\mathcal{E}}\|_1 \leq \frac{2\varepsilon}{\alpha}.$$

Proof. From Fact 5, $\alpha - \varepsilon/2 \leq \mathsf{P}_{X'}(\mathcal{E}) \leq \alpha + \varepsilon/2$. By definition, there exists an event $\mathcal{E}'$ such that $2(\mathsf{P}_{X|\mathcal{E}}(\mathcal{E}') - \mathsf{P}_{X'|\mathcal{E}}(\mathcal{E}')) = \|\mathsf{P}_{X|\mathcal{E}} - \mathsf{P}_{X'|\mathcal{E}}\|_1$. Now, $\mathsf{P}_X(\mathcal{E} \wedge \mathcal{E}') = \mathsf{P}_X(\mathcal{E})\mathsf{P}_{X|\mathcal{E}}(\mathcal{E}') \geq \alpha\mathsf{P}_{X|\mathcal{E}}(\mathcal{E}')$. Similarly, $\mathsf{P}_{X'}(\mathcal{E} \wedge \mathcal{E}') \leq (\alpha + \varepsilon/2)\mathsf{P}_{X'|\mathcal{E}}(\mathcal{E}') \leq \alpha\mathsf{P}_{X'|\mathcal{E}}(\mathcal{E}') + \frac{1}{2}\|\mathsf{P}_X - \mathsf{P}_{X'}\|_1$.

Now,

$$\begin{aligned}\|\mathsf{P}_X - \mathsf{P}_{X'}\|_1 &\geq 2(\mathsf{P}_X(\mathcal{E} \wedge \mathcal{E}') - \mathsf{P}_{X'}(\mathcal{E} \wedge \mathcal{E}')) \\ &\geq 2\alpha(\mathsf{P}_{X|\mathcal{E}}(\mathcal{E}') - \mathsf{P}_{X'|\mathcal{E}}(\mathcal{E}')) - \|\mathsf{P}_X - \mathsf{P}_{X'}\|_1 \\ &\geq \alpha\|\mathsf{P}_{X|\mathcal{E}} - \mathsf{P}_{X'|\mathcal{E}}\|_1 - \|\mathsf{P}_X - \mathsf{P}_{X'}\|_1\end{aligned}$$

which gives the required result. ◀

▶ **Fact 8** ([3], Lemma 16). *Suppose XYZ are random variables satisfying $\mathsf{P}_{XY}(x,y^*) = \alpha \cdot \mathsf{P}_X(x)$ for all x. Then,*

$$\left\|\mathsf{P}_{XYZ} - \mathsf{P}_{XY}\mathsf{P}_{Z|X,y^*}\right\|_1 \leq \frac{2}{\alpha}\left\|\mathsf{P}_{XYZ} - \mathsf{P}_{XY}\mathsf{P}_{Z|X}\right\|_1.$$

▶ **Corollary 9.** *Supose P_{XY} and $\mathsf{P}_{X'Y'Z'}$ are distributions such that $\mathsf{P}_X(x,y^*) = \alpha \cdot \mathsf{P}_X(x)$ for all x. Then,*

$$\|\mathsf{P}_{X'Z'|y^*} - \mathsf{P}_{X'Z'}\|_1 \leq \frac{11}{\alpha}\|\mathsf{P}_{X'Y'Z'} - \mathsf{P}_{XY}\mathsf{P}_{Z'|X'}\|_1.$$

Proof. Let $\|\mathsf{P}_{X'Y'Z'} - \mathsf{P}_{XY}\mathsf{P}_{Z'|X'}\|_1 = \varepsilon$. Note that

$$\|\mathsf{P}_{X|y^*} - \mathsf{P}_{X'|y^*}\|_1 \leq \frac{2\varepsilon}{\alpha}$$

by Fact 7. Let $\mathsf{P}_{XYZ''}$ denote the distribution $\mathsf{P}_{XY}\mathsf{P}_{Z'|X'Y'}$.

$$\begin{aligned}
\|\mathsf{P}_{X'Z'} - \mathsf{P}_{XZ''}\|_1 &= \sum_{x,z}\left|\mathsf{P}_{X'}(x)\sum_y \mathsf{P}_{Y'|x}(y)\mathsf{P}_{Z'|xy}(z) - \mathsf{P}_X(x)\sum_y \mathsf{P}_{Y|x}(y)\mathsf{P}_{Z'|xy}(z)\right| \\
&\leq \sum_{x,y,z}\left|\mathsf{P}_{X'}(x)\mathsf{P}_{Y'|x}(y) - \mathsf{P}_X(x)\mathsf{P}_{Y|x}(y)\right|\mathsf{P}_{Z'|xy}(z) \\
&= \|\mathsf{P}_{X'Y'} - \mathsf{P}_{XY}\|_1 \leq \varepsilon.
\end{aligned}$$

$$\begin{aligned}
\|\mathsf{P}_{XYZ''} - \mathsf{P}_{XY}\mathsf{P}_{Z''|X}\|_1 &\leq \|\mathsf{P}_{XYZ''} - \mathsf{P}_{X'Y'Z'}\|_1 + \|\mathsf{P}_{X'Y'Z'} - \mathsf{P}_{XY}\mathsf{P}_{Z'|X'}\|_1 \\
&\quad + \|\mathsf{P}_{XY}\mathsf{P}_{Z'|X'} - \mathsf{P}_{XY}\mathsf{P}_{Z''|X}\|_1 \\
&= \|\mathsf{P}_{XY} - \mathsf{P}_{X'Y'}\|_1 + \|\mathsf{P}_{X'Y'Z'} - \mathsf{P}_{XY}\mathsf{P}_{Z'|X'}\|_1 \\
&\quad + \sum_{x,y}\mathsf{P}_{XY}(x,y)\|\mathsf{P}_{Z'|x} - \mathsf{P}_{Z''|x}\|_1 \\
&\leq 2\varepsilon + \sum_x \mathsf{P}_X(x)\sum_{y,z}|\mathsf{P}_{Y|x}(y) - \mathsf{P}_{Y'|x}(y)|\mathsf{P}_{Z'|xy}(z) \\
&\leq 2\varepsilon + \sum_{x,y}|\mathsf{P}_X(x)\mathsf{P}_{Y|x}(y) - \mathsf{P}_{X'}(x)\mathsf{P}_{Y'|x}(y)| \\
&\quad + \sum_{x,y}|\mathsf{P}_{X'}(x) - \mathsf{P}_X(x)|\mathsf{P}_{Y'|x}(y) \\
&\leq 2\varepsilon + 2\|\mathsf{P}_{XY} - \mathsf{P}_{X'Y'}\|_1 \leq 4\varepsilon.
\end{aligned}$$

Combining all this,

$$\begin{aligned}
\|\mathsf{P}_{X'Z'|y^*} - \mathsf{P}_{X'Z'}\|_1 &\leq \|\mathsf{P}_{X'Z'|y^*} - \mathsf{P}_{XZ''|y^*}\|_1 + \|\mathsf{P}_{XZ''|y^*} - \mathsf{P}_{XZ''}\|_1 + \|\mathsf{P}_{XZ''} - \mathsf{P}_{X'Z'}\|_1 \\
&\leq \|\mathsf{P}_{X|y^*} - \mathsf{P}_{X'|y^*}\|_1 + \|\mathsf{P}_{XZ''|y^*} - \mathsf{P}_{XZ''}\|_1 + \|\mathsf{P}_{XZ''} - \mathsf{P}_{X'Z'}\|_1 \\
&\leq \frac{2\varepsilon}{\alpha} + \frac{2}{\alpha}\|\mathsf{P}_{XYZ''} - \mathsf{P}_{XY}\mathsf{P}_{Z''|X}\|_1 + \varepsilon \\
&\leq \frac{2\varepsilon}{\alpha} + \frac{8\varepsilon}{\alpha} + \varepsilon \leq \frac{11\varepsilon}{\alpha}.
\end{aligned}$$

where we have used Lemma 8 in the third inequality. ◀

▶ **Fact 10** ([19], Corollary 6). *Let $\mathsf{P}_{TU_1\ldots U_kV} = \mathsf{P}_T\mathsf{P}_{U_1|T}\mathsf{P}_{U_2|T}\ldots\mathsf{P}_{U_k|T}\mathsf{P}_{V|TU_1\ldots U_k}$ be a probability distribution over $\mathcal{T}\times\mathcal{U}^k\times\mathcal{V}$, and let $\mathcal{E}$ be any event. Then,*

$$\sum_{i=1}^k \|\mathsf{P}_{TU_iV|\mathcal{E}} - \mathsf{P}_{TV|\mathcal{E}}\mathsf{P}_{U_i|T}\|_1 \leq \sqrt{k\left(\log(|\mathcal{V}|) + \log\left(\frac{1}{\Pr[\mathcal{E}]}\right)\right)}.$$

▶ **Definition 11** ([19]). *For two distributions P_{XY} and $\mathsf{P}_{X'Y'ST}$, we say (X,Y) is $(1-\varepsilon)$-embeddable in $(X'S,Y'T)$ if there exists a random variable R on a set $\mathcal{R}$ independent of XY and functions $f_A : \mathcal{X}\times\mathcal{R}\to\mathcal{S}$ and $f_B : \mathcal{Y}\times\mathcal{R}\to\mathcal{T}$, such that*

$$\|\mathsf{P}_{XYf_A(X,R)f_B(X,R)} - \mathsf{P}_{X'Y'ST}\|_1 \leq \varepsilon.$$

▶ **Fact 12** ([19, 25]). *If two distributions P_{XY} and $\mathsf{P}_{X'Y'R'}$ satisfy*

$$\|\mathsf{P}_{X'Y'R'} - \mathsf{P}_{XY}\mathsf{P}_{R'|X'}\|_1 \leq \varepsilon \qquad \|\mathsf{P}_{X'Y'R'} - \mathsf{P}_{XY}\mathsf{P}_{R'|Y'}\|_1 \leq \varepsilon,$$

then (X,Y) is $(1-5\varepsilon)$-embeddable in $(X'R',Y'R')$.[1]

2.2 Quantum information

The ℓ_1 distance between two quantum states ρ and σ is given by

$$\|\rho-\sigma\|_1 = \mathrm{Tr}\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)} = \mathrm{Tr}|\rho-\sigma|.$$

The fidelity between two quantum states is given by

$$\mathsf{F}(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1.$$

ℓ_1 distance and fidelity are related in the following way.

▶ **Fact 13** (Fuchs-van de Graaf inequality). *For any pair of quantum states ρ and σ,*

$$2(1-\mathsf{F}(\rho,\sigma)) \leq \|\rho-\sigma\|_1 \leq 2\sqrt{1-\mathsf{F}(\rho,\sigma)^2}.$$

For two pure states $|\psi\rangle$ and $|\phi\rangle$, we have

$$\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 = \sqrt{1-\mathsf{F}\left(|\psi\rangle\langle\psi|,|\phi\rangle\langle\phi|\right)^2} = \sqrt{1-|\langle\psi|\phi\rangle|^2}.$$

▶ **Fact 14** (Uhlmann's theorem). *Suppose ρ and σ are mixed states on register X which are purified to $|\rho\rangle$ and $|\sigma\rangle$ on registers XY, then it holds that*

$$\mathsf{F}(\rho,\sigma) = \max_U |\langle\rho|\mathbb{1}_X\otimes U|\sigma\rangle|$$

where the maximization is over unitaries acting only on register Y.

▶ **Fact 15** (Data-processing inequality). *For a quantum channel $\mathcal{E}$ and states ρ and σ,*

$$\|\mathcal{E}(\rho)-\mathcal{E}(\sigma)\|_1 \leq \|\rho-\sigma\|_1 \qquad \text{and} \qquad \mathsf{F}(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \geq \mathsf{F}(\rho,\sigma).$$

The entropy of a quantum state ρ on a register Z is given by

$$\mathsf{S}(\rho) = -\mathrm{Tr}(\rho\log\rho).$$

The relative entropy between two states ρ and σ of the same dimensions is given by

$$\mathsf{S}(\rho\|\sigma) = \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\rho\log\sigma).$$

[1] This fact is equivalent to Lemma 2.11 in [25], although this lemma is stated in terms of relative entropies instead of trace distances between the various distributions. In the proof of the lemma, the relative entropies are converted to the same trace distances as we consider, using Pinsker's inequality. This justifies our statement of the fact, which is tailored towards our application.

The relative min-entropy between ρ and σ is defined as

$$\mathsf{S}_\infty(\rho\|\sigma) = \min\{\lambda : \rho \leq 2^\lambda \sigma\}.$$

It is easy to see that $\mathsf{S}(\rho\|\sigma)$ and $\mathsf{S}_\infty(\rho\|\sigma)$ only take finite values when the support of ρ is contained in the support of σ. Moreover, clearly $0 \leq \mathsf{S}(\rho\|\sigma) \leq \mathsf{S}_\infty(\rho\|\sigma)$ for all ρ and σ.

The ε-smooth relative min-entropy between ρ and σ is defined as

$$\mathsf{S}^\varepsilon_\infty(\rho\|\sigma) = \inf_{\rho' : \|\rho-\rho'\|_1 \leq \varepsilon} \mathsf{S}(\rho'\|\sigma).$$

$\mathsf{S}^\varepsilon_\infty(\rho\|\sigma)$ can take a finite value even if the support of ρ is not contained in the support of σ, for example if ρ is ε-close to a state contained within the support of σ. $\mathsf{S}_\infty(\rho\|\sigma)$ cannot be upper bounded by $\mathsf{S}(\rho\|\sigma)$, but $\mathsf{S}^\varepsilon_\infty(\rho\|\sigma)$ can be, due to the Quantum Substate Theorem.

▶ **Fact 16** (Quantum Substate Theorem, [31, 23]). *For any two states ρ and σ such that the support of ρ is contained in the support of σ, and any $\varepsilon > 0$,*

$$\mathsf{S}^\varepsilon_\infty(\rho\|\sigma) \leq \frac{4\mathsf{S}(\rho\|\sigma)}{\varepsilon^2} + \log\left(\frac{1}{1-\varepsilon^2/4}\right).$$

▶ **Fact 17** (Pinsker's Inequality). *For any two states ρ and σ, $\|\rho-\sigma\|_1 \leq \sqrt{\mathsf{S}(\rho\|\sigma)}$.*

▶ **Fact 18.** *If $\sigma = \varepsilon\rho + (1-\varepsilon)\rho'$, then $\mathsf{S}_\infty(\rho\|\sigma) \leq \log(1/\varepsilon)$.*

▶ **Fact 19.** *For any three quantum states ρ, σ, φ such that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\varphi) \subseteq \mathrm{supp}(\sigma)$,*

$$\mathsf{S}_\infty(\rho\|\sigma) \leq \mathsf{S}_\infty(\rho\|\varphi) + \mathsf{S}_\infty(\varphi\|\sigma).$$

▶ **Fact 20.** *For any unitary U, $\mathsf{S}_\infty(U\rho U^\dagger\|U\sigma U^\dagger) = \mathsf{S}_\infty(\rho\|\sigma)$.*

A state of the form

$$\rho_{XY} = \sum_x \mathsf{P}_X(x)|x\rangle\langle x|_X \otimes \rho_{Y|x}$$

is called a CQ (classical-quantum) state, with X being the classical register and Y being quantum. We shall use X to refer to both the classical register and the classical random variable with the associated distribution. As in the classical case, here we are using $\rho_{Y|x}$ to denote the state of the register Y conditioned on $X = x$, or in other words the state of the register Y when a measurement is done on the X register and the outcome is x. Hence $\rho_{XY|x} = |x\rangle\langle x|_X \otimes \rho_{Y|x}$. When the registers are clear from context we shall often write simply ρ_x.

The mutual information between Y and Z with respect to a state ρ on YZ is defined as

$$\mathsf{I}(Y:Z)_\rho = \mathsf{S}(\rho_{YZ}\|\rho_Y \otimes \rho_Z).$$

The conditional mutual information between Y and Z conditioned on a classical register X, is defined as

$$\mathsf{I}(Y:Z|X) = \mathop{\mathbb{E}}_{\mathsf{P}_X}[\mathsf{I}(Y:Z)_{\rho_x}].$$

Mutual information can be seen to satisfy the chain rule

$$\mathsf{I}(XY:Z)_\rho = \mathsf{I}(X:Z)_\rho + \mathsf{I}(Y:Z|X)_\rho.$$

▶ **Fact 21** ([6], Lemma B.7). *For any quantum state ρ_{YZ},*

$$\inf_{\sigma_Z} \mathsf{S}_\infty(\rho_{YZ}\|\rho_Y \otimes \sigma_Z) \leq 2\min\{\log|\mathcal{Y}|, \log|\mathcal{Z}|\}.$$

▶ **Fact 22.** *For CQ states*

$$\rho_{XY} = \sum_x \mathsf{P}_X(x)|x\rangle\langle x|_X \otimes \rho_{Y|x} \qquad \sigma_{XY} = \sum_x \mathsf{P}_{X'}(x)|x\rangle\langle x|_X \otimes \sigma_{Y|x},$$

their relative entropy is given by

$$\mathsf{S}(\rho_{XY}\|\sigma_{XY}) = \mathsf{S}(\mathsf{P}_X\|\mathsf{P}_{X'}) + \mathop{\mathbb{E}}_{\mathsf{P}_X}[\mathsf{S}(\rho_{Y|x}\|\sigma_{Y|x})].$$

▶ **Fact 23.** *Suppose σ_{XYZ} and ρ_{XYZ} are CQ states defined as follows*

$$\sigma_{XYZ} = \sum_{x,y} \mathsf{P}_{XY}(x,y)|x,y\rangle\langle x,y| \otimes \sigma_{Z|xy} \qquad \rho_{XYZ} = \sum_{x,y} \mathsf{P}_{X'Y'}(x,y)|x,y\rangle\langle x,y| \otimes \sigma_{Z|xy},$$

where $\|\mathsf{P}_{XY} - \mathsf{P}_{X'Y'}\|_1 \leq \delta$. Let $\mathsf{I}(Y:Z|X)_\sigma \leq c$. Then, for any $0 < \varepsilon < \frac{1}{4}$,

$$\mathsf{P}_{X'Y'}\left(\mathsf{S}^\varepsilon_\infty(\sigma_{Z|xy}\|\sigma_{Z|x}) > \frac{4c+1}{\varepsilon^3}\right) \leq \varepsilon + \frac{\delta}{2}.$$

Proof. We have $\mathbb{E}_{\mathsf{P}_{XY}}[\mathsf{S}(\sigma_{Z|xy}\|\sigma_{Z|x})] = \mathsf{I}(Y:Z|X)_\sigma \leq c$. By Markov's inequality, this means that

$$\mathsf{P}_{XY}\left(\mathsf{S}(\sigma_{Z|xy}\|\sigma_{Z|x}) > \frac{c}{\varepsilon}\right) \leq \varepsilon.$$

Using the Quantum Substate Theorem, this implies

$$\mathsf{P}_{XY}\left(\mathsf{S}^\varepsilon_\infty(\sigma_{Z|xy}\|\sigma_{Z|x}) > \frac{4c+1}{\varepsilon^3}\right) \leq \mathsf{P}_{XY}\left(\mathsf{S}^\varepsilon_\infty(\sigma_{Z|xy}\|\sigma_{Z|x}) > \frac{4c}{\varepsilon^3} + \log\left(\frac{1}{1-\varepsilon^2/4}\right)\right) \leq \varepsilon.$$

Since $\|\mathsf{P}_{XY} - \mathsf{P}_{X'Y'}\|_1 \leq \delta$, this gives us the required bound of the probability under $\mathsf{P}_{X'Y'}$. ◀

▶ **Fact 24** (Quantum Raz's Lemma, [3]). *Let ρ_{XY} and σ_{XY} be two CQ states with $X = X_1 \ldots X_k$ being classical, and σ being product across all registers. Then,*

$$\sum_{i=1}^k \mathsf{I}(X_i:Y)_\rho \leq \mathsf{S}(\rho_{XY}\|\sigma_{XY}).$$

▶ **Fact 25** ([29], Lemma 2). *Suppose the state*

$$|\sigma\rangle_{X\tilde{X}AB} = \sum_x \sqrt{\mathsf{P}_X(x)}|xx\rangle_{X\tilde{X}}|\sigma\rangle_{AB|x}$$

satisfies $\mathsf{P}_X(\mathsf{S}^\varepsilon_\infty(\sigma_{B|x}\|\sigma_B) > c) \leq \delta$ for some $\delta > 0$. Then there is a family of measurement operators $\{\Pi_x\}_x$ acting only on $X\tilde{X}A$ such that:

- **(i)** *Each Π_x succeeds with probability $\alpha = 2^{-c/\delta}$ on $|\sigma\rangle_{X\tilde{X}AB}$, i.e., $\|\Pi_x \otimes \mathbb{1}_B|\sigma\rangle\|_2^2 = 2^{-c/\delta}$,*
- **(ii)** *$(\Pi_x \otimes \mathbb{1}_B)|\sigma\rangle\langle\sigma|(\Pi_x \otimes \mathbb{1}_B)$ is of the form $|xx\rangle\langle xx| \otimes \rho_x$, for some state ρ_x on AB, and*

$$\mathop{\mathbb{E}}_{\mathsf{P}_X}\left\|\frac{1}{\alpha}(\Pi_x \otimes \mathbb{1}_B)|\sigma\rangle\langle\sigma|_{X\tilde{X}AB}(\Pi_x \otimes \mathbb{1}_B) - |xx\rangle\langle xx|_{X\tilde{X}} \otimes |\sigma\rangle\langle\sigma|_{AB|x}\right\|_1 \leq \varepsilon + 2\delta.$$

The version of the above fact stated here is slightly different from the original statement in [29], in order to suit our application. In the original statement, $\mathsf{I}(X:B)_\sigma$ is used instead, and the superposition state lacks the $\tilde{X}$ register. However, in the proof of the fact in [29], $\mathsf{I}(X:B)_\sigma$ is converted to $\mathsf{P}_X(\mathsf{S}^\varepsilon_\infty(\sigma_{B|x}\|\sigma_B) > c)$ anyway, so the first change makes no difference. The second change also makes no difference as the same projector that takes the superposition state without the $\tilde{X}$ register to $|x\rangle\langle x| \otimes |\sigma\rangle\langle\sigma|_{AB|x}$ takes the superposition state with the $\tilde{X}$ register to $|xx\rangle\langle xx| \otimes |\sigma\rangle\langle\sigma|_{AB|x}$.

2.3 Quantum communication & entangled games

We briefly describe a quantum communication protocol $\mathcal{P}$ for computing a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, between two parties Alice and Bob sharing prior entanglement, with inputs x and y respectively.

In each round, either Alice or Bob will apply a unitary on their classical input register, along with the quantum register they received as a message from the other party in the last round, and memory registers they may have kept from previous rounds; after the unitary they will keep some registers as memory and send the rest to the other party as the message for that round. We can always assume that players make "safe" copies of their inputs using CNOT gates in such protocols, so that the input registers come out as is after each round. We also note that though in general we need not consider shared classical randomness in quantum communication protocols, protocols with shared randomness fall under the shared entanglement framework we have described. This is because shared randomness can be obtained by sharing entanglement and then both parties measuring in the same basis.

In a one-way, i.e., a single round protocol, the memory from previous rounds is replaced by Alice's (who we consider to be sending the single message) part of the shared entangled state, and any register she does not send as a message is simply discarded. After Alice's message, Bob performs a projective measurement on his input register, his part of the shared entanglement, and Alice's message, and gives the outcome of this measurement as the output of the protocol, which we shall denote by $\mathcal{P}(x, y)$. We can of course think of this measurement as Bob performing a unitary on the three registers, and then doing a measurement in the computational basis on some $\log |\mathcal{Z}|$ qubits which are designated for the output.

▶ **Definition 26.** *The one-way entanglement-assisted quantum communication complexity, with error $0 < \varepsilon < 1$, of a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, denoted by $\mathrm{Q}^1_\varepsilon(f)$, is the minimum message size, i.e., number of qubits sent, in a one-way entanglement-assisted quantum protocol $\mathcal{P}$ such that for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$,*

$$\Pr[\mathcal{P}(x, y) \in f(x, y)] \geq 1 - \varepsilon,$$

where the probability is taken over the inherent randomness in the protocol.

▶ **Definition 27.** *For a probability distribution p on $\mathcal{X} \times \mathcal{Y}$, the distributional one-way entanglement-assisted quantum communication complexity of a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, with error $0 < \varepsilon < 1$ with respect to p, is defined as the minimum message size of a one-way entanglement-assisted quantum protocol $\mathcal{P}$ such that*

$$\Pr[\mathcal{P}(x, y) \in f(x, y)] \geq 1 - \varepsilon,$$

where the probability is taken over the distribution p on (x, y) as well as the inherent randomness in the protocol.

▶ **Fact 28** (Yao's lemma, [45])**.** *For any $0 < \varepsilon < 1$, and any relation f, $\mathrm{Q}^1_\varepsilon(f) = \max_p \mathrm{Q}^1_{p,\varepsilon}(f)$.*

A two-player non-local game G is described as $(q, \mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}, \mathsf{V})$ where q is a distribution over the input set $\mathcal{X} \times \mathcal{Y}$, $\mathcal{A} \times \mathcal{B}$ is the output set, and $\mathsf{V} : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B} \to \{0, 1\}$ is a predicate. It is played as follows: a referee selects inputs (x, y) according to q, sends x to Alice and y to Bob. If Alice and Bob are allowed to share entanglement, they perform measurements on their respective halves of the entangled state along with their respective input registers (which we model as performing unitaries and then measuring in the computational basis on some $\log |\mathcal{A}|$ and $\log |\mathcal{B}|$ qubits designated for outputs respectively), and send their outputs (a, b) back to the referee. The referee accepts and Alice and Bob win the game iff $\mathsf{V}(x, y, a, b) = 1$.

▶ **Definition 29.** *The entangled value of a game $G = (q, \mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}, \mathsf{V})$, denoted by $\omega^*(G)$, is the maximum winning probability of Alice and Bob, averaged over the distribution q as well as inherent randomness in the strategy, over all shared entanglement strategies for G.*

3 Proof of direct product theorem

In this section, we prove Theorem 1, whose statement we recall below.

▶ **Theorem 1.** *For any relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, and any $0 < \varepsilon, \zeta < \frac{1}{2}$,*

$$\mathrm{Q}^1_{1-(1-\varepsilon)^{\Omega(\zeta^6 k/\log|\mathcal{Z}|)}}(f^k) = \Omega\left(k\left(\zeta^5 \cdot \mathrm{Q}^1_{\varepsilon+\zeta}(f) - \log\log(1/\zeta)\right)\right).$$

3.1 Setup

Let p be the hard distribution on $\mathcal{X} \times \mathcal{Y}$ for $\mathrm{Q}^1_{\varepsilon+12\zeta}(f)$ from Yao's lemma, i.e., $\mathrm{Q}^1_{\varepsilon+12\zeta}(f) = \mathrm{Q}^1_{p,\varepsilon+12\zeta}(f)$. Consider the relation $\tilde{f} \subseteq \mathcal{X} \times (\mathcal{Y} \cup \{y^*\}) \times \mathcal{Z}$ which is the same as f on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ and additionally,

$$(x, y^*, z) \in \tilde{f} \quad \forall x \in \mathcal{X}, \forall z \in \mathcal{Z}.$$

We can think of p as a distribution on $\mathcal{X} \times (\mathcal{Y} \cup \{y^*\})$ as well, which has $p(y^*) = 0$. Clearly,

$$\mathrm{Q}^1_{p,\gamma}(\tilde{f}) = \mathrm{Q}^1_{p,\gamma}(f) \tag{1}$$

for any error γ, since p has no support on the extra inputs on which $\tilde{f}$ is defined. We also note that

$$\mathrm{Q}^1_{\gamma}(f^k) \geq \mathrm{Q}^1_{\gamma}(\tilde{f}^k) \tag{2}$$

for any γ. This is because any protocol for f^k is also a protocol for $\tilde{f}^k$: on the indices where Bob's input is y^* instead of an element of $\mathcal{Y}$, he pretends he has gotten an input from $\mathcal{Y}$, runs the protocol with this input and gives the answer accordingly. This gives a correct output if the original protocol gives a correct output, since any output is correct when Bob's input in y^*.

For a distribution q related to p, we shall show that

$$\mathrm{Q}^1_{q^k, 1-(1-\varepsilon)^{\Omega(\zeta^6 k/\log|\mathcal{Z}|)}}(\tilde{f}^k) \geq \frac{\zeta^5 k}{60} \cdot \mathrm{Q}^1_{p,\varepsilon+12\zeta}(\tilde{f}) - k\log\log\left(\frac{24}{5\zeta}\right). \tag{3}$$

Since $\mathrm{Q}^1_{\gamma}(\tilde{f}^k) \geq \mathrm{Q}^1_{q^k,\gamma}(\tilde{f}^k)$, (1), (2) and (3) imply the theorem. The distribution q is defined as follows

$$q(x, y) = (1-\zeta) \cdot p(x, y) \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$
$$q(x, y^*) = \zeta \cdot p(x) \quad \forall x \in \mathcal{X}.$$

Clearly, $q(x, y^*) = q(x)q(y^*)$ for all x, and

$$\|p(x, y) - q(x, y)\|_1 \leq 2\zeta. \tag{4}$$

Following [3], for each $i \in [k]$, we shall define a joint distribution $\mathsf{P}_{X_i Y_i D_i G_i}$, where the marginal on $X_i Y_i$ is $q(x, y)$, and $D_i G_i$ are correlation-breaking variables such that conditioned on $D_i G_i = d_i g_i$, X_i and Y_i are independent. Each $X_i Y_i D_i G_i$ is distributed independently of the rest. Each D_i is distributed uniformly in $\{0, 1\}$. Depending on the value of D_i, G_i is distributed in the following way:

$$G_i = \begin{cases} x & \text{w.p. } p(x) & \text{if } D_i = 0 \\ y^* & \text{w.p. } 1-(1-\zeta)^{2/3} & \text{if } D_i = 1 \\ y & \text{w.p. } (1-\zeta)^{2/3} \cdot p(y) & \text{if } D_i = 1 \end{cases}$$

Now depending on the value of D_iG_i, X_iY_i is distributed in the following way:

$$X_iY_i = \begin{cases} (x,y^*) & \text{w.p. } \zeta & \text{if } D_i=0, G_i=x \\ (x,y) & \text{w.p. } (1-\zeta)\cdot p(y|x) & \text{if } D_i=0, G_i=x \\ (x,y^*) & \text{w.p. } p(x) & \text{if } D_i=1, G_i=y^* \\ (x,y^*) & \text{w.p. } \left(1-(1-\zeta)^{1/3}\right)\cdot p(x|y) & \text{if } D_i=1, G_i=y \\ (x,y) & \text{w.p. } (1-\zeta)^{1/3}\cdot p(x|y) & \text{if } D_i=1, G_i=y. \end{cases}$$

The following lemma is similar to Claim 18 from [3]; we provide a proof for completeness.

▶ **Lemma 30.** *For all* $(x,y) \in \mathcal{X} \times (\mathcal{Y} \cup \{y^*\})$, $\mathsf{P}_{X_iY_i}(x,y) = q(x,y)$.

Proof. It is trivial to see that $\mathsf{P}_{G_iY_i|D_i=0}(x,y) = \mathsf{P}_{X_iY_i|D_i=0}(x,y) = q(x,y)$, since $G_i = X_i$ conditioned on $D_i = 0$. We now prove the $D_i = 1$ case. First consider a $y \in \mathcal{Y}$. Y_i can only take value y if G_i takes value y. Hence,

$$\begin{aligned} \mathsf{P}_{X_iY_i|D_i=1}(x,y) &= \mathsf{P}_{G_i|D_i=1}(y)\cdot \mathsf{P}_{X_iY_i|D_i=1,G_i=y}(x,y) \\ &= (1-\zeta)^{2/3}p(y)\cdot(1-\zeta)^{1/3}p(x|y) \\ &= (1-\zeta)\cdot p(x,y) = q(x,y). \end{aligned}$$

On the other hand, Y_i can take value y^* when $G_i = y^*$ or when $G_i = y$ for any $y \in \mathcal{Y}$. Hence,

$$\begin{aligned} \mathsf{P}_{X_iY_i|D_i=1}(x,y^*) &= \mathsf{P}_{G_i|D_i=1}(y^*)\cdot \mathsf{P}_{X_iY_i|D_i=1,G_i=y^*}(x,y^*) \\ &\quad + \sum_{y\in\mathcal{Y}} \mathsf{P}_{G_i|D_i=1}(y)\cdot \mathsf{P}_{X_iY_i|D_i=1,G_i=y}(x,y^*) \\ &= \left(1-(1-\zeta)^{2/3}\right)\cdot p(x) + (1-\zeta)^{2/3}\left(1-(1-\zeta)^{1/3}\right)\sum_{y\in\mathcal{Y}} p(y)\cdot p(x|y) \\ &= \left(1-(1-\zeta)^{2/3}\right)\cdot p(x) + \left((1-\zeta)^{2/3}-(1-\zeta)\right)\cdot p(x) \\ &= \zeta\cdot p(x) = q(x,y^*). \end{aligned}$$

◀

In particular the lemma means $\mathsf{P}_{X_iY_i}(x,y^*) = \mathsf{P}_{X_i}(x)\mathsf{P}_{Y_i}(y^*)$. We also note

$$\mathsf{P}_{Y_iG_i|D_i=1}(Y_i \neq G_i) = (1-\zeta)^{2/3}(1-(1-\zeta)^{1/3}) \leq 1-2\zeta/3-1+\zeta = \zeta/3. \tag{5}$$

Let $\mathcal{P}$ be any quantum one-way protocol between Alice and Bob, for $\tilde{f}^k \subseteq \mathcal{X}^k \times (\mathcal{Y} \cup \{y^*\})^k \times \mathcal{Z}^k$, which has communication cost ck. $\mathcal{P}$ is depicted in Figure 1. Alice and Bob's inputs are in registers $X = X_1 \dots X_k$ and $Y = Y_1 \dots Y_k$, and they share an entangled pure state uncorrelated with the inputs on registers $E^{\mathrm{A}}E^{\mathrm{B}}$, with Alice holding E^{A} and Bob holding E^{B}. Alice applies a unitary V^{A} on XE^{A}, to get the message register M, and the register A to be discarded. We shall use $|\theta\rangle_{AME^{\mathrm{B}}|x}$ to refer to the pure state in AME^{B} in the protocol after Alice's unitary, for inputs xy ($|\theta\rangle_x$ only depends on y via x). When Alice and Bob's inputs are distributed according to P_{XY}, the state of the protocol after Alice's message, will be given by the following CQ state:

$$\theta_{XYAME^{\mathrm{B}}} = \sum_{xy} \mathsf{P}_{XY}(xy)|xy\rangle\langle xy|_{XY} \otimes |\theta\rangle\langle\theta|_{AME^{\mathrm{B}}|x}.$$

We shall also consider the following purification of it, with the purifying registers $\tilde{X}$ and $\tilde{Y}$:

$$|\theta\rangle_{X\tilde{X}Y\tilde{Y}AME^{\mathrm{B}}} = \sum_{xy} \sqrt{\mathsf{P}_{XY}(xy)}|xxyy\rangle_{X\tilde{X}Y\tilde{Y}}|\theta\rangle_{AME^{\mathrm{B}}|x}.$$

After receiving Alice's message, Bob applies a unitary V^{B} to YME^{B}, after which ME^{B} gets converted to BZ, where $Z = Z_1 \dots Z_k$ are the answer registers. We shall use $|\rho\rangle_{X\tilde{X}Y\tilde{Y}ABZ}$ to refer to $|\theta\rangle_{X\tilde{X}Y\tilde{Y}AME^{\mathrm{B}}}$ after V^{B}. We shall use P_{XYDGZ} to refer to the joint distribution where $XYDG$ are as previously defined; Z is independent of DG given XY, and the conditional distribution of Z given XY is what is obtained by measuring the Z register in the computational basis in $|\rho\rangle$.

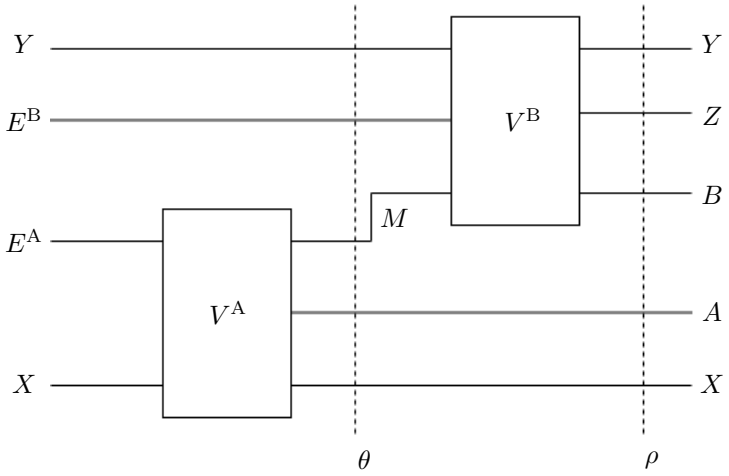

Figure 1 One-way quantum protocol $\mathcal{P}$.

3.2 Proof of Theorem 1

We shall show that if the communication cost ck of $\mathcal{P}$ is $< \frac{\zeta^5 k}{300} \cdot \mathrm{Q}^1_{p,\varepsilon+12\zeta}(\tilde{f}) - k \log\log(24/5\zeta)$, then the success probability of $\mathcal{P}$ is $(1-\varepsilon)^{\Omega(\zeta^6 k/\log|\mathcal{Z}|)}$. This is implied by the following claim, which the rest of the proof will show.

▶ **Lemma 31.** *Let $\delta = \frac{\zeta^6}{1440000}$ and $\delta' = \frac{\zeta^6}{1440000 \log|\mathcal{Z}|}$. For $i \in [k]$, let T_i be the random variable which takes value 1 if $\mathcal{P}$ computes $f(X_i, Y_i)$ correctly, and value 0 otherwise. If the communication cost of $\mathcal{P}$ is $< \frac{\zeta^5 k}{60} \cdot \mathrm{Q}^1_{p,\varepsilon+12\zeta}(\tilde{f}) - k \log\log(24/5\zeta)$, then there exist $\lfloor \delta' k \rfloor$ coordinates $\{i_1, \dots, i_{\lfloor \delta' k \rfloor}\} \subseteq [k]$, such that for all $1 \le r \le \lfloor \delta' k \rfloor - 1$, at least one of the following two conditions holds*

(i) $\Pr\left[\prod_{j=1}^{r} T_{i_j} = 1\right] \le (1-\varepsilon)^{\delta k}$

(ii) $\Pr\left[T_{i_{r+1}} = 1 \middle| \prod_{j=1}^{r} T_{i_j} = 1\right] \le 1 - \varepsilon.$

Lemma 31 can be proved inductively. Suppose we have already identified $1 \le t \le \lfloor \delta' k \rfloor$ coordinates in $C = \{i_1, \dots i_t\}$, such that for all $1 \le r \le t-1$, $\Pr\left[T_{i_{r+1}} = 1 | \prod_{j=1}^{r} T_{i_j} = 1\right] \le 1 - \varepsilon$. Let $\mathcal{E}$ refer to the event $\prod_{i \in C} T_i = 1$. If $\Pr[\mathcal{E}] \le (1-\varepsilon)^{\delta k}$, then we are already done. If not, then we shall show how to identify the $(t+1)$-th coordinate i such that $\Pr[T_i = 1 | \mathcal{E}] \le 1 - \varepsilon$. The process of identifying the first coordinate is also similar, except in that case the conditioning event is empty. Since we only use the lower bound $(1-\varepsilon)^{\delta k}$ on the probability of the conditioning event in our proof, the proof goes through for that case as well.

We shall use the state $|\varphi\rangle$, which is $|\rho\rangle_{X\tilde{X}Y\tilde{Y}ABZ}$ conditioned on $\mathcal{E}$, for the proof of Lemma 31. For any value $DG = dg$, $|\varphi\rangle_{X\tilde{X}Y\tilde{Y}ABZ|dg}$ is defined as:

$$|\varphi\rangle_{X\tilde{X}Y\tilde{Y}ABZ|dg} = \frac{1}{\sqrt{\gamma_{dg}}} \sum_{xy} \sqrt{\mathsf{P}_{XY|dg}(xy)} |xxyy\rangle_{X\tilde{X}Y\tilde{Y}} \otimes \sum_{z_C:(x_C,y_C,z_C)\in\tilde{f}^t} |z_C\rangle_{Z_C} |\tilde{\varphi}\rangle_{ABZ_{\bar{C}}|xyz_C}.$$

Here $|\tilde{\varphi}\rangle_{xyz_C}$ is a subnormalized state with $\||\tilde{\varphi}\rangle_{ABZ_{\bar{C}}|xyz_C}\|_2^2 = \mathsf{P}_{Z_C|xy}(z_C)$. The overall normalization factor γ_{dg} is the probability of $\mathcal{E}$ conditioned on dg, and satisfies

$$\sum_{dg} \mathsf{P}_{DG}(dg) \cdot \gamma_{dg} = \Pr[\mathcal{E}].$$

It is clear that the distribution of XYZ in $|\varphi\rangle_{X\tilde{X}Y\tilde{Y}ABZ|dg}$ is $\mathsf{P}_{XYZ|\mathcal{E},dg}$. Note that we are using the notation $|\varphi\rangle_{dg}$ without explicitly considering registers DG on which a measurement is done to obtain $|\varphi\rangle_{dg}$. We shall also sometimes use $|\varphi\rangle_{d_{-i}g_{-i}}$ in which the xy distributions are conditioned on $d_{-i}g_{-i}$ instead, which changes the normalization factor to some $\gamma_{d_{-i}g_{-i}}$, everything else remaining the same. $\varphi_{x_iy_id_{-i}g_{-i}}$ refers as usual to the state obtained when a measurement done on the X_iY_i registers (which are actually present in $|\varphi\rangle$) in $|\varphi\rangle_{d_{-i}g_{-i}}$. For $i \notin \bar{C}$, we shall use the states $|\varphi\rangle_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|x_iy_ix_Cy_Cz_Cd_{-i}g_{-i}}$ in our proof, which we note are pure states.

Lemma 31 will be proved with the help of the following lemma, whose proof we give later.

▶ **Lemma 32.** *If $\Pr[\mathcal{E}] \geq (1-\varepsilon)^{\delta k}$, then there exist a coordinate $i \in \bar{C}$, a random variable $R_i = X_CY_CZ_CD_{-i}G_{-i}$ and for each $R_i = r_i$ a state $|\varphi'\rangle_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|y^*r_i}$ such that the following conditions hold:*

- **(i)** $\|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},X_i}\|_1 \leq \frac{7\zeta}{120}$
- **(ii)** $\|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},Y_i}\|_1 \leq \frac{7\zeta}{120}$.
- **(iii)** *There exist projectors $\{\Pi_{x_ir_i}\}_{x_ir_i}$ acting only on registers $X_{\bar{C}}\tilde{X}_{\bar{C}}A$ and unitaries $\{U_{y_ir_i}\}_{y_ir_i}$ acting only on $Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}$, such that each $\Pi_{x_ir_i}$ succeeds on $|\varphi'\rangle_{r_i}$ with probability $\alpha = 2^{-c'}$ where $c' \leq \frac{60c}{\zeta^5}$, and*

$$\mathop{\mathbb{E}}_{\mathsf{P}_{X_iY_iR_i|\mathcal{E}}} \left\| \frac{1}{\alpha}(\Pi_{x_ir_i} \otimes U_{y_ir_i})|\varphi'\rangle\langle\varphi'|_{y^*r_i}(\Pi_{x_ir_i} \otimes U_{y_ir_i}^\dagger) - |\varphi\rangle\langle\varphi|_{x_iy_ir_i} \right\|_1 \leq 21\zeta.$$

Proof of Lemma 31. We give a one-way quantum protocol $\mathcal{P}'$ for $\tilde{f}$, whose inputs are distributed according to $\mathsf{P}_{X_iY_i}$, i.e., q, by embedding Alice and Bob's inputs into the i-th coordinate of $|\varphi\rangle_{x_iy_ir_i}$, as follows:

- Alice and Bob have r according to the distribution required by Fact 12 as shared randomness, and $2^{60c/\zeta^5}\log(24/5\zeta)$ copies of $|\varphi'\rangle_{y^*r_i}$ as shared entanglement, with Alice holding registers $X_{\bar{C}}\tilde{X}_{\bar{C}}A$ and Bob holding registers $Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}$ of each copy.
- On input (x_i, y_i) from $\mathsf{P}_{X_iY_i}$, using items (i), (ii) of Lemma 32, their shared randomness, and the protocol from Fact 12, Alice and Bob generate random variables $R_i^{\mathrm{A}}R_i^{\mathrm{B}}$ such that

$$\|\mathsf{P}_{X_iY_iR_i^{\mathrm{A}}R_i^{\mathrm{B}}} - \mathsf{P}_{X_iY_iR_iR_i|\mathcal{E}}\|_1 \leq \frac{7\zeta}{24}.$$

  where R_iR_i denotes two perfectly correlated copies of R_i in $\mathsf{P}_{X_iY_iR_iR_i|\mathcal{E}}$.
- Alice applies the $\{\Pi_{x_ir_i^{\mathrm{A}}}, \mathbb{1} - \Pi_{x_ir_i^{\mathrm{A}}}\}$ measurement according to her input and R_i^{A} on her registers for each copy of the shared entangled state. If the $\Pi_{x_ir_i^{\mathrm{A}}}$ measurement does not succeed on any copy, then she aborts. Otherwise, she sends to Bob a $(\frac{60c}{\zeta^5} + \log\log(24/5\zeta))$-bit message indicating an index where $\Pi_{x_ir_i^{\mathrm{A}}}$ measurement succeeded.

- Bob applies the unitary $U_{y_i r_i^{\mathrm{B}}}$ according to his input and R_i^{B} on the copy of the shared entangled state whose index Alice has sent, and measures the Z_i register of the resulting state to give his output.

To analyze the success of this protocol, first note that

$$\mathop{\mathbb{E}}_{\mathsf{P}_{X_i Y_i R_i|\mathcal{E}}} \Pr[\text{Result of } Z_i \text{ measurement on } |\varphi\rangle_{x_i y_i r_i} \in \tilde{f}(x_i, y_i)] = \Pr[T_i = 1|\mathcal{E}].$$

Let us first assume Alice and Bob have $(x_i, y_i, r_i^{\mathrm{A}}, r_i^{\mathrm{B}})$ distributed exactly according to $\mathsf{P}_{X_i Y_i R_i R_i|\mathcal{E}}$ – we shall denote both r_i^{A} and r_i^{B} by r_i in this case. Alice aborts the protocol if none of her measurements succeed. This happens with probability

$$(1 - 2^{-c'})^{2^{60c/\zeta^5} \cdot \log(24/5\zeta)} \leq \frac{5\zeta}{24}.$$

If Alice does not abort, then Alice and Bob's state after Bob's unitary is $\frac{1}{\sqrt{\alpha}} \Pi_{x_i r_i} \otimes U_{y_i r_i} |\varphi'\rangle_{y^* r_i}$. From (iii), the expected probability of the Z_i measurement on this state giving an answer $\in \tilde{f}(x_i, y_i)$ is at least $\Pr[T_i = 1|\mathcal{E}] - \frac{21\zeta}{2}$. Hence, if Alice and Bob had $(x_i, y_i, r_i^{\mathrm{A}}, r_i^{\mathrm{B}})$ distributed according to $\mathsf{P}_{X_i Y_i R_i R_i|\mathcal{E}}$, then their expected success probability would have been at least $\Pr[T_i = 1|\mathcal{E}] - \frac{21\zeta}{2} - \frac{5\zeta}{24}$. Since Alice and Bob have $(x_i, y_i, r_i^{\mathrm{A}}, r_i^{\mathrm{B}})$ according to $\mathsf{P}_{X_i Y_i R_i^{\mathrm{A}} R_i^{\mathrm{B}}}$ instead, their expected success probability is at least

$$\Pr[T_i = 1|\mathcal{E}] - \frac{21\zeta}{2} - \frac{5\zeta}{24} - \frac{7\zeta}{24} \geq \Pr[T_i = 1|\mathcal{E}] - 11\zeta.$$

Since $\|q(x, y) - p(x, y)\|_1 \leq 2\zeta$, when the same protocol is run on $X_i Y_i$ distributed according to p instead, it must succeed with probability at least $\Pr[T_i = 1|\mathcal{E}] - 12\zeta$. Since the communication in $\mathcal{P}'$ is at most $(\frac{60c}{\zeta^5} + \log\log(24/5\zeta)) < \mathrm{Q}^1_{p,\varepsilon+12\zeta}(\tilde{f})$, $\Pr[T_i = 1|\mathcal{E}] \geq 1 - \varepsilon$ gives the error probability of $\mathcal{P}'$ to be $\leq \varepsilon + 12\zeta$, which is a contradiction. Hence we must have $\Pr[T_i = 1|\mathcal{E}] \leq 1 - \varepsilon$. The desired result thus follows by setting $i_{t+1} = i$. ◀

3.3 Proof of Lemma 32

First we shall show that on expectation over $i \in \bar{C}$, a number of probability distributions conditioned on $\mathcal{E}$ are close to those unconditioned on $\mathcal{E}$. Applying Fact 10 with T and V being trivial and $U_i = X_i Y_i D_i G_i$ for $i \in \bar{C}$, we get,

$$\mathop{\mathbb{E}}_{i \in \bar{C}} \|\mathsf{P}_{X_i Y_i D_i G_i|\mathcal{E}} - \mathsf{P}_{X_i Y_i D_i G_i}\|_1 \leq \frac{1}{k-t} \sqrt{k \cdot \log((1-\varepsilon)^{-\delta k})} \leq \sqrt{2\delta}. \tag{6}$$

In particular, due to (5), this means

$$\mathop{\mathbb{E}}_{i \in \bar{C}} \mathsf{P}_{Y_i G_i|\mathcal{E}, D_i = 1}(Y_i = G_i) \geq 1 - \zeta/3 - \sqrt{2\delta}. \tag{7}$$

And since $\mathsf{P}_{G_i|D_i=1}(y^*) = 1 - (1-\zeta)^{2/3}$, $\mathsf{P}_{Y_i|D_i=1, G_i=y_i}(y_i) = (1-\zeta)^{1/3}$ for $y_i \in \mathcal{Y}$, we have

$$\zeta + \sqrt{2\delta} \geq 1 - (1-\zeta)^{2/3} + \sqrt{2\delta} \geq \mathop{\mathbb{E}}_{i \in \bar{C}} \mathsf{P}_{G_i|\mathcal{E}, D_i=1}(y^*) \geq 1 - (1-\zeta)^{2/3} - \sqrt{2\delta} \geq 2\zeta/3 - \sqrt{2\delta} \tag{8}$$

$$\left(1 - \frac{\zeta}{3} + \sqrt{2\delta}\right) \mathop{\mathbb{E}}_{i \in \bar{C}} \mathsf{P}_{G_i|\mathcal{E}, D_i=1}(y_i) \geq \mathop{\mathbb{E}}_{i \in \bar{C}} \mathsf{P}_{Y_i G_i|\mathcal{E}, D_i=1}(y_i, y_i) \geq (1 - \zeta - \sqrt{2\delta}) \mathop{\mathbb{E}}_{i \in \bar{C}} \mathsf{P}_{G_i|\mathcal{E}, D_i=1}(y_i). \tag{9}$$

Fact 10 can again be applied with $U_i = X_i Y_i$, $T = X_C Y_C D G$ and $V = Z_C$. Let $\delta_1 = \delta + \delta' \log|\mathcal{Z}| = \frac{\zeta^6}{720000}$. Then we have,

$$\sqrt{2\delta_1} \geq \mathop{\mathbb{E}}_{i \in \bar{C}} \|\mathsf{P}_{X_i Y_i X_C Y_C Z_C D G|\mathcal{E}} - \mathsf{P}_{X_C Y_C Z_C D G|\mathcal{E}} \mathsf{P}_{X_i Y_i|X_C Y_C D G}\|_1$$

$$= \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iX_CY_CZ_CDG|\mathcal{E}} - \mathsf{P}_{X_CY_CZ_CDG|\mathcal{E}}\mathsf{P}_{X_iY_i|D_iG_i}\|_1$$
$$= \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iD_iG_iR_i|\mathcal{E}} - \mathsf{P}_{D_iG_iR_i|\mathcal{E}}\mathsf{P}_{X_iY_i|D_iG_i}\|_1. \tag{10}$$

We note that D_i takes value uniformly in $\{0,1\}$ even conditioned on $\mathcal{E}$. Hence from (10),

$$\sqrt{2\delta_1} \geq \frac{1}{2} \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iG_iR_i|\mathcal{E},D_i=0} - \mathsf{P}_{G_iR_i|\mathcal{E},D_i=0}\mathsf{P}_{X_iY_i|G_i,D_i=0}\|_1$$
$$= \frac{1}{2} \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iR_i|\mathcal{E}}\mathsf{P}_{Y_i|X_i}\|_1$$

where we have used the fact that $X_i = G_i$ conditioned on $D_i = 0$. Combining this with the fact that $\mathbb{E}_{i\in\bar{C}} \|\mathsf{P}_{X_i|\mathcal{E}} - \mathsf{P}_{X_i}\|_1 \leq \sqrt{2\delta}$, we have,

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},X_i}\|_1 \leq 3\sqrt{2\delta_1} < \frac{7\zeta^3}{600}. \tag{11}$$

Due to Corollary 9 we also have from (11),

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iR_i|\mathcal{E},y^*} - \mathsf{P}_{X_iR_i|\mathcal{E}}\|_1 \leq \frac{33\sqrt{2\delta_1}}{\zeta}. \tag{12}$$

Let $\mathcal{F}_i$ denote the event $Y_i = G_i$. We know $\mathbb{E}_{i\in\bar{C}} \mathsf{P}_{X_iY_iG_i|D_i=1}(\mathcal{F}_i) \geq 1 - \zeta/3 - \sqrt{2\delta}$, from (7). Hence, using Fact 7,

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{Y_iR_i|\mathcal{E}}\mathsf{P}_{X_i|Y_i}\|_1$$
$$= \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iG_iR_i|\mathcal{E},D_i=1,\mathcal{F}_i} - \mathsf{P}_{G_iR_i|\mathcal{E},D_i=1,\mathcal{F}_i}\mathsf{P}_{X_iY_i|G_iD_i=1,\mathcal{F}_i}\|_1$$
$$\leq 6 \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iD_iR_i|\mathcal{E},D_i=1} - \mathsf{P}_{G_iR_i|\mathcal{E},D_i=1}\mathsf{P}_{X_iY_i|G_iD_i=1}\|_1 \leq 6\sqrt{2\delta_1}.$$

Using $\mathbb{E}_{i\in\bar{C}} \|\mathsf{P}_{Y_i|\mathcal{E}} - \mathsf{P}_{Y_i}\|_1 \leq \sqrt{2\delta}$, we have as before,

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},Y_i}\|_1 \leq 7\sqrt{2\delta_1} = \frac{7\zeta^3}{600}. \tag{13}$$

Next we shall show the existence of projectors $\Pi_{x_ir_i}$ which take $|\varphi'\rangle_{y^*r_i}$ (which will be defined soon) close to $|\varphi\rangle_{x_iy^*r_i}$. Since M is ck qubits, by Fact 21, for any value $DG = dg$, there exists some state $\sigma_{M|dg}$ such that

$$\mathsf{S}_\infty(\theta_{XY\tilde{Y}E^{\mathrm{B}}M|dg} \| \theta_{XY\tilde{Y}E^{\mathrm{B}}|dg} \otimes \sigma_{M|dg}) \leq 2ck.$$

By Fact 20 we have,

$$\mathsf{S}_\infty\left(\rho_{XY\tilde{Y}BZ|dg} \| V^{\mathrm{B}}(\theta_{XY\tilde{Y}E^{\mathrm{B}}|dg} \otimes \sigma_{M|dg})(V^{\mathrm{B}})^\dagger\right) \leq 2ck.$$

Let $\psi_{X_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}|dg} = \mathrm{Tr}_{Z_C}(V^{\mathrm{B}}(\theta_{XYE^{\mathrm{B}}|dg} \otimes \sigma_{M|x_Cy_Cdg})(V^{\mathrm{B}})^\dagger)$. Note that $\theta_{XY\tilde{Y}E^{\mathrm{B}}|dg} \otimes \sigma_{M|dg}$ is product across X and the other registers, and V^{B} does not act on X. Hence $\psi_{X_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}|dg}$ is also product across X and the other registers, and moreover, all the X_i-s are in product with each other as well. We have,

$$\mathsf{S}_\infty\left(\rho_{XY\tilde{Y}BZ_{\bar{C}}|dg} \| \psi_{XY\tilde{Y}BZ_{\bar{C}}|dg}\right) \leq 2ck.$$

Using Facts 22 and 19, this gives us

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathsf{P}_{X_C Y_C Z_C DG|\mathcal{E}}} \left[\mathsf{S}\left(\varphi_{X_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}|x_C y_C z_C dg} \| \psi_{X_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}|x_C y_C dg}\right)\right] \\
&\leq \mathop{\mathbb{E}}_{\mathsf{P}_{Z_C DG|\mathcal{E}}} \left[\mathsf{S}\left(\varphi_{XY\tilde{Y}BZ_{\bar{C}}|z_C dg} \| \psi_{XY\tilde{Y}BZ_{\bar{C}}|dg}\right)\right] \\
&\leq \mathop{\mathbb{E}}_{\mathsf{P}_{Z_C DG|\mathcal{E}}} \left[\mathsf{S}_\infty\left(\varphi_{XY\tilde{Y}BZ_{\bar{C}}|z_C dg} \| \psi_{XY\tilde{Y}BZ_{\bar{C}}|dg}\right)\right] \\
&\leq \mathop{\mathbb{E}}_{\mathsf{P}_{Z_C DG|\mathcal{E}}} \Big[\mathsf{S}_\infty\left(\varphi_{XY\tilde{Y}BZ_{\bar{C}}|z_C dg} \| \varphi_{XY\tilde{Y}BZ_{\bar{C}}|dg}\right) \\
&\quad + \mathsf{S}_\infty\left(\varphi_{XY\tilde{Y}BZ_{\bar{C}}|dg} \| \rho_{XY\tilde{Y}BZ_{\bar{C}}|dg}\right) + \mathsf{S}_\infty\left(\rho_{XY\tilde{Y}BZ_{\bar{C}}|dg} \| \psi_{XY\tilde{Y}BZ_{\bar{C}}|dg}\right)\Big] \\
&\leq \mathop{\mathbb{E}}_{\mathsf{P}_{Z_C DG|\mathcal{E}}} \left[\log(1/\mathsf{P}_{Z_C|\mathcal{E}}(z_C)) + \log(1/\Pr[\mathcal{E}]) + 2ck\right] \\
&\leq |C| \log|\mathcal{Z}| + \delta k + 2ck \leq (\delta_1 + 2c)k.
\end{aligned}$$

By Quantum Raz's Lemma,

$$\begin{aligned}
4c + 2\delta_1 &\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_C Y_C Z_C DG|\mathcal{E}}} \mathsf{I}(X_i : Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}})_{\varphi_{x_C y_C z_C dg}} \\
&= \mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{D_i G_i R_i|\mathcal{E}}} \mathsf{I}(X_i : Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}})_{\varphi_{d_i g_i r_i}} \\
&\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \frac{1}{2}\mathsf{P}_{G_i|\mathcal{E},D_i=1}(y^*) \mathop{\mathbb{E}}_{\mathsf{P}_{R_i|\mathcal{E},D_i=1,G_i=y^*}} \mathsf{I}(X_i : Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}})_{\varphi_{r_i|D_i=1,G_i=y^*}} \\
&\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \frac{1}{2}(2\zeta/3 - \sqrt{2\delta}) \mathop{\mathbb{E}}_{\mathsf{P}_{R_i|\mathcal{E},D_i=1,G_i=y^*}} \mathsf{I}(X_i : Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}})_{\varphi_{r_i,D_i=1,G_i=y^*}} \qquad (14)
\end{aligned}$$

where we have used (8) in the last inequality.

Note that $\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|x_i r_i, D_i=1, G_i=y^*}$ is the same state as $\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|x_i y^* r_i}$, where the value of Y_i is being conditioned on, instead of G_i. $|\varphi\rangle_{r_i,D_i=1,G_i=y^*}$ is the superposition over X_i of $|\varphi\rangle_{x_i r_i, D_i=1, G_i=y^*}$, with the X_i distribution being $\mathsf{P}_{X_i|\mathcal{E},r_i,D_i=1,G_i=y^*}$. The only difference between $|\varphi\rangle_{y^* r_i}$ and $|\varphi\rangle_{r_i,D_i=1,G_i=y^*}$ is the X_i distribution, which in the former is $\mathsf{P}_{X_i|\mathcal{E},y^* r_i}$ instead. We shall refer to $|\varphi\rangle_{r_i,D_i=1,G_i=y^*}$ as simply $|\varphi\rangle_{r_i,1,y^*}$ as now on – note that there is no ambiguity between this and $|\varphi\rangle_{y^* r_i}$. The same goes for the distributions $\mathsf{P}_{X_i R_i|\mathcal{E},1,y^*}$ and $\mathsf{P}_{X_i R_i|\mathcal{E},y^*}$.

$\mathsf{P}_{X_i|1,y^*}$ is the same distribution as $\mathsf{P}_{X_i|y^*}$ and $\mathsf{P}_{R_i|\mathcal{E},x_i,1,y^*}$ is the same distribution as $\mathsf{P}_{R_i|\mathcal{E},x_i y^*}$ for any x_i. Hence,

$$\begin{aligned}
\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_i R_i|\mathcal{E},y^*} - \mathsf{P}_{X_i R_i|\mathcal{E},1,y^*}\|_1 &\leq \mathop{\mathbb{E}}_{i\in\bar{C}} \big[\|\mathsf{P}_{X_i R_i|\mathcal{E},y^*} - \mathsf{P}_{X_i|y^*}\mathsf{P}_{R_i|\mathcal{E},X_i,y^*}\|_1 \\
&\quad + \|(\mathsf{P}_{X_i|1,y^*} - \mathsf{P}_{X_i|\mathcal{E},1,y^*})\mathsf{P}_{R_i|\mathcal{E},X_i,y^*}\|_1\big] \\
&\leq \mathop{\mathbb{E}}_{i\in\bar{C}} \left[\frac{\|\mathsf{P}_{X_i R_i|\mathcal{E}} - \mathsf{P}_{X_i}\mathsf{P}_{R_i|\mathcal{E},X_i}\|_1}{2\zeta/3 - \sqrt{2\delta}} + \frac{\|\mathsf{P}_{X_i|\mathcal{E}} - \mathsf{P}_{X_i}\|_1}{2\zeta/3 - \sqrt{2\delta}}\right] \\
&\leq \frac{7\sqrt{2\delta_1}}{\zeta}
\end{aligned}$$

where we have used (8) in the second inequality. Using the above computation and (12), we get,

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_i R_i|\mathcal{E}} - \mathsf{P}_{X_i R_i|\mathcal{E},1,y^*}\|_1 \leq \frac{40\sqrt{2\delta_1}}{\zeta}.$$

Let

$$|\varphi'\rangle_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|y^*r_i} = \sum_{x_i} \sqrt{\mathsf{P}_{X_i|\mathcal{E},r_i}}|\varphi\rangle_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|x_iy^*r_i},$$

i.e., $|\varphi'\rangle_{y^*r_i}$ is the same state as $|\varphi\rangle_{y^*r_i}$ except that the distribution of distribution of X_i is unconditioned on $Y_i = y^*$. From (14) and Fact 23, we then have that,

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \mathsf{P}_{X_iR_i|\mathcal{E}} \left(\mathsf{S}^{\zeta}_{\infty}\left(\varphi'_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}|x_iy^*r_i}\|\varphi'_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}|y^*r_i}\right) > \frac{28(2c+\delta_1)+1}{\zeta^4}\right) \leq \zeta + \frac{20\sqrt{2\delta_1}}{\zeta}.$$

Hence by Fact 25, there exist projectors $\Pi_{x_ir_i}$ acting on registers $X_{\bar{C}}\tilde{X}_{\bar{C}}A$, such that $\Pi_{x_ir_i}$ succeeds with probability $\alpha = 2^{-c'}$ on $|\varphi'\rangle_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|y^*r_i}$, where $c' = \frac{60c}{\zeta^5}$, and

$$\begin{aligned}\mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_iR_i|\mathcal{E}}} \left\|\frac{1}{\alpha}(\Pi_{x_ir_i}\otimes\mathbb{1})|\varphi'\rangle\langle\varphi'|_{y^*r_i}(\Pi_{x_ir_i}\otimes\mathbb{1}) - |\varphi\rangle\langle\varphi|_{x_iy^*r_i}\right\|_1 &\leq 3\zeta + \frac{40\sqrt{2\delta_1}}{\zeta^2} \\ &\leq \frac{7\zeta}{2}. \end{aligned} \tag{15}$$

Next we shall show the existence of unitaries $U_{y_ir_i}$ taking $|\varphi\rangle_{y^*r_i,D_i=1,G_i=y_i}$ close to $|\varphi\rangle_{y_ir_i,D_i=1,G_i=y_i}$. By similar arguments as the ones leading to (14) on Bob's side (except the first step where we consider the information due to Alice's message, which does not apply here), we can alo upper bound $\mathbb{E}_{\mathsf{P}_{X_CY_CZ_CDG|\mathcal{E}}}\left[\mathsf{S}\left(\varphi_{Y_{\bar{C}}X_{\bar{C}}\tilde{X}_{\bar{C}}A|x_Cy_Cz_Cdg}\|\rho_{Y_{\bar{C}}X_{\bar{C}}\tilde{X}_{\bar{C}}A|x_Cy_Cdg}\right)\right]$. Hence by Raz's lemma again,

$$\begin{aligned}
2\delta_1 &\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{D_iG_iR_i|\mathcal{E}}} \mathsf{I}(Y_i : X_{\bar{C}}\tilde{X}_{\bar{C}}A)_{\varphi_{d_ig_ir_i}} \\
&\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \frac{1}{2}(1-\zeta-\sqrt{2\delta}) \mathop{\mathbb{E}}_{\mathsf{P}_{R_iG_i|\mathcal{E},D_i=1,G_i\neq y^*}} \mathsf{I}(Y_i : X_{\bar{C}}\tilde{X}_{\bar{C}}A)_{\varphi_{r_i,D_i=1,g_i}} \\
&= \mathop{\mathbb{E}}_{i\in\bar{C}} \frac{1}{2}(1-\zeta-\sqrt{2\delta}) \mathop{\mathbb{E}}_{\mathsf{P}_{R_iG_iY_i|\mathcal{E},D_i=1,G_i\neq y^*}} \left[\mathsf{S}\left(\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|y_i,D_i=1,g_i}\|\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|D_i=1,g_i}\right)\right] \\
&\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \frac{1}{2}(1-\zeta-\sqrt{2\delta}) \sum_{y_i\in\mathcal{Y}} \mathop{\mathbb{E}}_{\mathsf{P}_{R_i|\mathcal{E},D_i=1,G_i=y_i}} \mathsf{P}_{G_i|\mathcal{E},D_i=1}(y_i)\cdot \\
&\quad \Big[(1-\zeta-\sqrt{2\delta})\|\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|y_i,r_i,D_i=1,G_i=y_i} - \varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|r_i,D_i=1,G_i=y_i}\|_1^2 \\
&\quad +(\zeta/3-\sqrt{2\delta})\|\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|y^*,r_i,D_i=1,G_i=y_i} - \varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|r_i,D_i=1,G_i=y_i}\|_1^2\Big] .
\end{aligned}$$

where we have used (9) and Pinsker's inequality in the last line. Since the ℓ_1 norm obeys triangle inequality, we have,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in\bar{C}} \sum_{y_i\in\mathcal{Y}} \mathop{\mathbb{E}}_{\mathsf{P}_{R_i|\mathcal{E},1,y_i}} \mathsf{P}_{G_i|\mathcal{E},1}(y_i)\|\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|y_ir_i,1,y_i} - \varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|y^*r_i,1,y_i}\|_1^2 \\
&\leq \mathop{\mathbb{E}}_{i\in\bar{C}} \sum_{y_i\in\mathcal{Y}} \mathop{\mathbb{E}}_{\mathsf{P}_{R_i|\mathcal{E},1,y_i}} \mathsf{P}_{G_i|\mathcal{E},1}(y_i)\cdot 2\Big[\|\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|y_i,r_i,1,y_i} - \varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|r_i,1,y_i}\|_1^2 \\
&\quad + \|\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|y^*,r_i,1,y_i} - \varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}A|r_i,1,y_i}\|_1^2\Big] \\
&\leq \frac{4\delta_1}{1-\zeta-\sqrt{2\delta}}\left(\frac{1}{1-\zeta-\sqrt{2\delta}} + \frac{1}{\zeta/3-\sqrt{2\delta}}\right) \\
&\leq \frac{32\delta_1}{\zeta}.
\end{aligned}$$

We note that $\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|y_ir_i,1,y_i}$ and $\varphi_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}ABZ_{\bar{C}}|y^*r_i,1,y_i}$ are pure states. Hence, using the Fuchs-van de Graaf inequality and Uhlmann's theorem, there exist unitaries $U_{y_ir_i}$ acting only on $Y_{\bar{C}}\tilde{Y}_{\bar{C}}BZ_{\bar{C}}$ such that

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{i\in\bar{C}} \sum_{y_i\in\mathcal{Y}} \mathop{\mathbb{E}}_{\mathsf{P}_{R_i|\mathcal{E},1,y_i}} \mathsf{P}_{G_i|\mathcal{E},1}(y_i) \| |\varphi\rangle\langle\varphi|_{y_ir_i,1,y_i} - (\mathbb{1}\otimes U_{y_ir_i})|\varphi\rangle\langle\varphi|_{y^*r_i,1,y_i}(\mathbb{1}\otimes U^\dagger_{y_ir_i})\|_1 \\
&\leq \left(\frac{32\delta_1}{\zeta}\right)^{1/4} \qquad (16)
\end{aligned}
$$

Finally, we need to show that $\Pi_{x_ir_i}\otimes U_{y_ir_i}$ takes $|\varphi'\rangle_{y^*r_i}$ close to $|\varphi\rangle_{x_iy_ir_i}$. To do this, we shall first show that $U_{y_ir_i}$ in fact takes $|\varphi\rangle_{x_iy^*r_i}$ close to $|\varphi\rangle_{x_iy_ir_i}$. Consider the superoperator $\mathcal{O}_{X_i}$ that measures the register X_i and writes it in a different register.

$$
\begin{aligned}
\mathcal{O}_{X_i}(|\varphi\rangle\langle\varphi|_{y_ir_i,1,y_i}) &= \sum_{x_i} \mathsf{P}_{X_i|\mathcal{E},y_ir_i,D_i=1,G_i=y_i}(x_i)|x_i\rangle\langle x_i| \otimes |\varphi\rangle\langle\varphi|_{x_iy_ir_i,1,y_i} \\
&= \sum_{x_i} \mathsf{P}_{X_i|\mathcal{E},y_ir_i,D_i=1,G_i=y_i}(x_i)|x_i\rangle\langle x_i| \otimes |\varphi\rangle\langle\varphi|_{x_iy_ir_i} \\
\mathcal{O}_{X_i}(|\varphi\rangle\langle\varphi|_{y^*r_i,1,y_i}) &= \sum_{x_i} \mathsf{P}_{X_i|\mathcal{E},y^*r_i,D_i=1,G_i=y_i}(x_i)|x_i\rangle\langle x_i| \otimes |\varphi\rangle\langle\varphi|_{x_iy^*r_i}
\end{aligned}
$$

where we have made the observation that $|\varphi\rangle\langle\varphi|_{x_iy_ir_i,1,y_i}$ and $|\varphi\rangle\langle\varphi|_{x_iy^*r_i,1,y_i}$ are the same states as $|\varphi\rangle\langle\varphi|_{x_iy_ir_i}$ and $|\varphi\rangle\langle\varphi|_{x_iy^*r_i}$. By Fact 10 we can get,

$$
\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iG_iR_i|\mathcal{E},1} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|1,G_i}\|_1 \leq 2\sqrt{2\delta_1}.
$$

Hence, for any value $Y_i = y_i$,

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iG_iR_i|\mathcal{E},1} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|\mathcal{E},y_i,1,G_iR_i}\|_1 \\
&\leq \mathop{\mathbb{E}}_{i\in\bar{C}} \left[\|\mathsf{P}_{X_iG_iR_i|\mathcal{E},1} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|y_i,1,G_i})\|_1 + \|\mathsf{P}_{G_iR_i|\mathcal{E},1}(\mathsf{P}_{X_i|y_i,1,G_i} - \mathsf{P}_{X_i|\mathcal{E},y_i,1,G_iR_i})\|_1\right] \\
&\leq \mathop{\mathbb{E}}_{i\in\bar{C}} \left[\|\mathsf{P}_{X_iG_iR_i|\mathcal{E},1} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|1,G_i}\|_1 + \frac{2}{\frac{\zeta}{3}-\sqrt{2\delta}}\|\mathsf{P}_{X_iG_iR_i|\mathcal{E},1} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|1,G_i}\|_1\right] \\
&\leq \frac{8\sqrt{2\delta_1}}{\zeta}
\end{aligned}
$$

where we have used the fact that for any value $G_i = g_i$, we must have $\mathsf{P}_{Y_i|1,g_i}(y_i) \geq \zeta/3 - \sqrt{2\delta}$. Finally,

$$
\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iG_iR_i|\mathcal{E},1} - \mathsf{P}_{X_iY_iR_i|\mathcal{E},1}\|_1 \leq 2\mathsf{P}_{Y_iG_i|\mathcal{E},1}(Y_i \neq G_i) \leq \zeta/3 + \sqrt{2\delta}.
$$

Observing that $\mathsf{P}_{X_iY_iR_i|\mathcal{E},1}$ is the same as $\mathsf{P}_{X_iY_iR_i|\mathcal{E}}$ we get,

$$
\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|\mathcal{E},y_i,1,G_iR_i}\|_1 \leq \frac{8\sqrt{2\delta_1}}{\zeta} + \frac{\zeta}{3} + \sqrt{2\delta}.
$$

Using this and (16) we get,

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_iY_iR_i|\mathcal{E}}} \| |\varphi\rangle\langle\varphi|_{x_iy_ir_i} - (\mathbb{1}\otimes U_{y_ir_i})|\varphi\rangle\langle\varphi|_{x_iy^*r_i}(\mathbb{1}\otimes U^\dagger_{y_ir_i})\|_1 \\
&\leq \mathop{\mathbb{E}}_{i\in\bar{C}} \Big[\|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|\mathcal{E},y_i,1,G_iR_i}\|_1 + \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{G_iR_i|\mathcal{E},1}\mathsf{P}_{X_i|\mathcal{E},y^*,1,G_iR_i}\|_1
\end{aligned}
$$

$$+ \mathop{\mathbb{E}}_{\mathsf{P}_{G_i R_i|\mathcal{E},1}} \left\| \mathop{\mathbb{E}}_{\mathsf{P}_{X_i|\mathcal{E},y_i r_i,1,y_i}} |x_i\rangle\langle x_i| \otimes |\varphi\rangle\langle\varphi|_{x_i y_i r_i} - \mathop{\mathbb{E}}_{\mathsf{P}_{X_i|\mathcal{E},y^* r_i,1,y_i}} \mathbb{1} \otimes U_{y_i r_i} |\varphi\rangle\langle\varphi|_{x_i y^* r_i} \mathbb{1} \otimes U_{y_i r_i}^\dagger \right\|_1 \Bigg]$$

$$= \frac{16\sqrt{2\delta_1}}{\zeta} + \frac{2\zeta}{3} + 2\sqrt{2\delta} + \left(\frac{32\delta_1}{\zeta}\right)^{1/4} < \frac{7\zeta}{10} \tag{17}$$

where we have bounded the last term in the first inequality by applying Fact 15 on (16) with $\mathcal{O}_{X_i}$. Notice that we have also removed the conditioning $G_i \neq y^*$, since for $G_i = y^*$, the corresponding states are both $|\varphi\rangle_{x_i y^* r_i}$.

From (15) and (17) we get,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_i Y_i R_i|\mathcal{E}}} \left\| \frac{1}{\alpha} (\Pi_{x_i r_i} \otimes U_{y_i r_i}) |\varphi'\rangle\langle\varphi'|_{y^* r_i} (\Pi_{x_i r_i} \otimes U_{y_i r_i}^\dagger) - |\varphi\rangle\langle\varphi|_{x_i y_i r_i} \right\|_1 \\
&\leq \mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_i Y_i R_i|\mathcal{E}}} \Bigg[\left\| \frac{1}{\alpha} (\Pi_{x_i r_i} \otimes U_{y_i r_i}) |\varphi'\rangle\langle\varphi'|_{y^* r_i} (\Pi_{x_i r_i} \otimes U_{y_i r_i}^\dagger) \right. \\
&\quad \left. - (\mathbb{1} \otimes U_{y_i r_i}) |\varphi\rangle\langle\varphi|_{x_i y^* r_i} (\mathbb{1} \otimes U_{y_i r_i}^\dagger) \right\|_1 \\
&\quad + \left\| (\mathbb{1} \otimes U_{y_i r_i}) |\varphi\rangle\langle\varphi|_{x_i y^* r_i} (\mathbb{1} \otimes U_{y_i r_i}^\dagger) - |\varphi\rangle\langle\varphi|_{x_i y_i r_i} \right\|_1 \Bigg] \\
&= \mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_i Y_i R_i|\mathcal{E}}} \Bigg[\left\| \frac{1}{\alpha} (\Pi_{x_i r_i} \otimes \mathbb{1}) |\varphi'\rangle\langle\varphi'|_{y^* r_i} (\Pi_{x_i r_i} \otimes \mathbb{1}) - |\varphi\rangle\langle\varphi|_{x_i y^* r_i} \right\|_1 \\
&\quad + \left\| (\mathbb{1} \otimes U_{y_i r_i}) |\varphi\rangle\langle\varphi|_{x_i y^* r_i} (\mathbb{1} \otimes U_{y_i r_i}^\dagger) - |\varphi\rangle\langle\varphi|_{x_i y_i r_i} \right\|_1 \Bigg] \\
&\leq \frac{7\zeta}{2} + \frac{7\zeta}{10} = \frac{21\zeta}{5}.
\end{aligned} \tag{18}$$

Using Markov's inequality on (11), (13) and (18), we get an index $i \in \bar{C}$ such that the conditions (i)-(iii) for Lemma 32 hold.

4 Proof of parallel repetition theorem

In this section we prove Theorem 2, whose statement is recalled below.

▶ **Theorem 2.** *For a two-player non-local game $G = (q, \mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}, \mathsf{V})$ such that q is a distribution anchored on one side with anchoring probability ζ,*

$$\omega^*(G^k) = \left(1 - (1 - \omega^*(G))^5\right)^{\Omega\left(\frac{\zeta^2 k}{\log(|\mathcal{A}|\cdot|\mathcal{B}|)}\right)}.$$

4.1 Setup

The proof of this theorem is very similar to that of the direct product theorem, so we shall only highlight points of difference. Whereas in the communication case, we started with an arbitrary distribution p and defined distribution q anchored on one side close to p, here we start with an already anchored distribution. To preserve similarity with the direct product proof, we shall consider q to be anchored on the $\mathcal{Y}$ side here as well, but the proof goes through analogously for a distribution anchored on the $\mathcal{X}$ side. We define the correlation-breaking variables and the joint distribution P_{XYDG} exactly as before.[2]

[2] The definition of $\mathsf{P}_{X_i Y_i D_i G_i}$ in the previous section makes references to $p(x, y)$. Since there is no p in the present case, $p(x, y)$ can simply be replaced by $q(x, y|y \neq y^*)$.

We consider an entangled strategy $\mathcal{S}$ for G^k, where Alice and Bob, with input registers $X = X_1 \dots X_k$ and $Y = Y_1 \dots Y_k$, initially share an entangled state, and perform unitaries V^{A} and V^{A} respectively on their parts of the entangled state and and their input registers. As before, conditioned on any value $DG = dg$, we define the following pure state representing $\mathcal{S}$ after these unitaries:

$$|\theta\rangle_{X\tilde{X}Y\tilde{Y}ABE'^{\mathrm{A}}E'^{\mathrm{B}}|dg} = \sum_{xy} \sqrt{\mathsf{P}_{XY|dg}(xy)}|xxyy\rangle_{X\tilde{X}Y\tilde{Y}} \otimes |\theta\rangle_{ABE^{\mathrm{A}}E^{\mathrm{B}}|xy}$$

where AB are the answer registers which are measured in the computational basis by Alice and Bob to obtain their answers (a, b), and $E'^{\mathrm{A}}E'^{\mathrm{B}}$ are some additional registers which are discarded. We shall use $\mathsf{P}_{XYAB|dg}$ to denote the distribution of $XYAB$ in $|\theta\rangle_{dg}$; P_{XYDGAB} is obtained by averaging over dg.

Let the winning probability of of $\omega^*(G)$ be $1 - 5\varepsilon$ for an appropriate ε. We shall prove the following lemma, which is analogous to the direct product case. It is clear that the lemma implies

$$\omega^*(G^k) \leq (1-\varepsilon)^{\frac{\zeta^2\varepsilon^4 k}{\log(|\mathcal{A}|\cdot|\mathcal{B}|)}} = \left(1 - (1-\omega^*(G))^5\right)^{\Omega\left(\frac{\zeta^2 k}{\log(|\mathcal{A}|\cdot|\mathcal{B}|)}\right)}.$$

▶ **Lemma 33.** *Let $\delta = \frac{\zeta^2\varepsilon^4}{1440000}$ and $\delta' = \frac{\zeta^2\varepsilon^4}{1440000\log(|\mathcal{A}|\cdot|\mathcal{B}|)}$. For $i \in [k]$, let T_i denote the random variable $\mathsf{V}(X_i, Y_i, A_i, B_i)$, where $X_iY_iA_iB_i$ are according to P_{XYAB}. Then there exist $\lfloor \delta' k \rfloor$ coordinates $\{i_1, \dots, i_{\lfloor \delta' k \rfloor}\} \subseteq [k]$, such that for all $1 \leq r \leq \lfloor \delta' k \rfloor - 1$, at least one of the conditions holds*

- **(i)** $\Pr\left[\prod_{j=1}^{r} T_{i_j} = 1\right] \leq (1-\varepsilon)^{\delta k}$
- **(ii)** $\Pr\left[T_{i_{r+1}} = 1 \middle| \prod_{j=1}^{r} T_{i_j} = 1\right] \leq 1 - \varepsilon.$

As before, we shall consider that we have identified a set of coordinates $C = \{i_1, \dots, i_t\}$ such that for all $1 \leq r \leq t-1$, $\Pr\left[T_{i_{r+1}} = 1 | \prod_{j=1}^{r} T_{i_j} = 1\right] \leq 1 - \varepsilon$ and $\Pr[\mathcal{E}] = \Pr\left[\prod_{j=1}^{t} T_{i_j} = 1\right] \geq (1-\varepsilon)^{\delta k}$, and identify a $(t+1)$-th coordinate i. Let E^{A} and E^{B} to denote $A_{\bar{C}}E'^{\mathrm{A}}$ and $B_{\bar{C}}E'^{\mathrm{B}}$ respectively. We define the following state, which is $|\theta\rangle_{dg}$ conditioned on success in C:

$$|\varphi\rangle_{X\tilde{X}Y\tilde{Y}A_CB_CBE^{\mathrm{A}}E^{\mathrm{B}}|dg} = \frac{1}{\sqrt{\gamma_{dg}}} \sum_{xy} \sqrt{\mathsf{P}_{XY|dg}(xy)}|xxyy\rangle_{X\tilde{X}Y\tilde{Y}} \otimes \sum_{a_Cb_C:\mathsf{V}^t(x_C,y_C,a_C,b_C)=1} |a_Cb_C\rangle_{A_CB_C}|\tilde{\varphi}\rangle_{E^{\mathrm{A}}E^{\mathrm{B}}|xya_Cb_C}.$$

Here $|\tilde{\varphi}\rangle_{E^{\mathrm{A}}E^{\mathrm{B}}|xya_Cb_C}$ is a subnormalized state satisfying $\||\tilde{\varphi}\rangle_{E^{\mathrm{A}}E^{\mathrm{B}}|xya_Cb_C}\|_2^2 = \mathsf{P}_{A_CB_C|xy}(a_Cb_C)$.

The following lemma is the analog of Lemma 32, which we shall use to prove Lemma 33.

▶ **Lemma 34.** *If $\Pr[\mathcal{E}] \geq (1-\varepsilon)^{\delta k}$, then there exist a coordinate $i \in \bar{C}$, a random variable $R_i = X_CY_CA_CB_CD_{-i}G_{-i}$, such that the following conditions hold:*

- **(i)** $\|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},X_i}\|_1 \leq \frac{7\varepsilon}{150}$
- **(ii)** $\|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},Y_i}\|_1 \leq \frac{7\varepsilon}{150}$
- **(iii)** *There exist unitaries $\{U_{x_ir_i}\}_{x_ir_i}$ and $\{U_{y_ir_i}\}_{y_ir_i}$ respectively acting only on $X_{\bar{C}}\tilde{X}_{\bar{C}}E^{\mathrm{A}}$ and $Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}$, such that*

$$\mathop{\mathbb{E}}_{\mathsf{P}_{X_iY_iR_i|\mathcal{E}}} \left\|(U_{x_ir_i} \otimes U_{y_ir_i})|\varphi\rangle\langle\varphi|_{y^*r_i}(U^\dagger_{x_ir_i} \otimes U^\dagger_{y_ir_i}) - |\varphi\rangle\langle\varphi|_{x_iy_ir_i}\right\|_1 \leq \frac{36\varepsilon}{5}.$$

It is easy to see how this lemma implies Lemma 33. As in the direct product case, Alice and Bob share $|\varphi\rangle_{y^*r_i}$ as entanglement – though in this case only one copy, as well as classical randomness with which they can produce $R_i^{\mathrm{A}}R_i^{\mathrm{B}}$ satisfying

$$\|\mathsf{P}_{X_iY_iR_i^{\mathrm{A}}R_i^{\mathrm{B}}} - \mathsf{P}_{X_iY_iR_iR_i|\mathcal{E}}\|_1 \leq \frac{7\varepsilon}{30}.$$

Alice and Bob apply $U_{x_ir_i^{\mathrm{A}}}$ and $U_{y_ir_i^{\mathrm{B}}}$ according to their inputs and R_i^{A} and R_i^{B} respectively, on their registers E^{A} and E^{B} of $|\varphi\rangle_{y^*r_i}$. They then measure in the computational basis on the A_iB_i registers of resulting state, to give their outcomes (a_i, b_i). $\Pr[T_i = 1|\mathcal{E}] \geq 1 - \varepsilon$ implies that the resulting strategy for G has success probability $> (1 - 5\varepsilon)$, a contradiction which lets us identify i as the $(t + 1)$-th coordinate.

4.2 Proof of Lemma 34

We can prove

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},X_i}\|_1 \leq \frac{7\varepsilon}{600} \tag{19}$$

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iY_iR_i|\mathcal{E}} - \mathsf{P}_{X_iY_i}\mathsf{P}_{R_i|\mathcal{E},Y_i}\|_1 \leq \frac{7\varepsilon}{600} \tag{20}$$

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_iY_iR_i|\mathcal{E}}} \||\varphi\rangle\langle\varphi|_{x_iy_ir_i} - (\mathbb{1}\otimes U_{y_ir_i})|\varphi\rangle\langle\varphi|_{x_iy^*r_i}(\mathbb{1}\otimes U_{y_ir_i}^\dagger)\|_1 \leq \frac{4\varepsilon}{5} \tag{21}$$

exactly the same way as in the direct product case, except conditioning on z_C is replaced by conditioning on a_Cb_C, which leads to the factor of $\log(|\mathcal{A}|\cdot|\mathcal{B}|)$. The rest of the proof will hence be spent getting Alice's unitaries $U_{x_ir_i}$.

Letting $\delta_1 = \delta + \delta'\log(|\mathcal{A}|\cdot|\mathcal{B}|)$, the following is derived analogously to the direct product case, except for the extra factor in the mutual information bound due to communication:

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{R_i|\mathcal{E},D_i=1,G_i=y^*} \mathsf{I}(X_i : Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}})_{\varphi_{r_i,D_i=1,G_i=y^*}} \leq \frac{10\delta_1}{\zeta} \tag{22}$$

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iR_i|\mathcal{E},y^*} - \mathsf{P}_{X_iR_i|\mathcal{E},1,y^*}\|_1 \leq \frac{7\sqrt{2\delta_1}}{\zeta} \tag{23}$$

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iR_i|\mathcal{E}} - \mathsf{P}_{X_iR_i|\mathcal{E},1,y^*}\|_1 \leq \frac{40\sqrt{2\delta_1}}{\zeta}. \tag{24}$$

From (22), by applying Pinsker's inequality, we get,

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_iR_i|\mathcal{E},1,y^*}} \|\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|x_ir_i,1,y^*} - \varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|r_i,1,y^*}\|_1 \leq \left(\frac{10\delta_1}{\zeta}\right)^{1/2}$$

Note that $\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|x_ir_i,1,y^*}$ is the same state as $\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|x_iy^*r_i}$. But $\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|r_i,1,y^*}$ is not the same state as $\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|y^*r_i}$, due to the averaging over X_i being done with respect to $\mathsf{P}_{X_i|\mathcal{E},r_i,1,y^*}$ in one, and with respect to $\mathsf{P}_{X_i|\mathcal{E},y^*r_i}$ in the other. However, due to (23) we can say,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_iR_i|\mathcal{E},1,y^*}} \|\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|x_iy^*r_i} - \varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|y^*r_i}\|_1 \\
&\leq \left(\frac{10\delta_1}{\zeta}\right)^{1/2} + \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iR_i|\mathcal{E},1,y^*} - \mathsf{P}_{R_i|\mathcal{E},1,y^*}\mathsf{P}_{X_i|\mathcal{E},R_i,y^*}\|_1 \\
&\leq \left(\frac{10\delta_1}{\zeta}\right)^{1/2} + \mathop{\mathbb{E}}_{i\in\bar{C}} \|\mathsf{P}_{X_iR_i|\mathcal{E},y^*} - \mathsf{P}_{X_iR_i|\mathcal{E},1,y^*}\|_1
\end{aligned}$$

$$\leq \frac{2\sqrt{108\delta_1}}{\zeta}.$$

Since $|\varphi\rangle_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{A}}E^{\mathrm{B}}|y^*r_i}$ is a purification of $\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|y^*r_i}$ and $|\varphi\rangle_{X_{\bar{C}}\tilde{X}_{\bar{C}}Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{A}}E^{\mathrm{B}}|x_iy^*r_i}$ is a purification of $\varphi_{Y_{\bar{C}}\tilde{Y}_{\bar{C}}E^{\mathrm{B}}|x_iy^*r_i}$, by the Fuchs-van de Graaf inequality and Uhlmann's theorem we can say that there exist unitaries $U_{x_ir_i}$ on $X_{\bar{C}}\tilde{X}_{\bar{C}}E^{\mathrm{A}}$ such that

$$\mathop{\mathbb{E}}_{i\in\bar{C}}\mathsf{P}_{X_iR_i|\mathcal{E},1,y^*}\mathop{\mathbb{E}}\left\||\varphi\rangle\langle\varphi|_{x_iy^*r_i}-(U_{x_ir_i}\otimes\mathbb{1})|\varphi\rangle\langle\varphi|_{y^*r_i}(U^\dagger_{x_ir_i}\otimes\mathbb{1})\right\|_1\leq\left(\frac{2\sqrt{108\delta_1}}{\zeta}\right)^{1/2}$$

and by (24) again,

$$\begin{aligned}\mathop{\mathbb{E}}_{i\in\bar{C}}\mathsf{P}_{X_iR_i|\mathcal{E}}\mathop{\mathbb{E}}\left\||\varphi\rangle\langle\varphi|_{x_iy^*r_i}-(U_{x_ir_i}\otimes\mathbb{1})|\varphi\rangle\langle\varphi|_{y^*r_i}(U^\dagger_{x_ir_i}\otimes\mathbb{1})\right\|_1&\leq\left(\frac{2\sqrt{108\delta_1}}{\zeta}\right)^{1/2}+\frac{40\sqrt{2\delta_1}}{\zeta}\\&\leq 2\left(\frac{10800\delta_1}{\zeta^2}\right)^{1/4}\\&\leq\varepsilon.\end{aligned}\tag{25}$$

Combining (25) and (21) we get,

$$\mathop{\mathbb{E}}_{i\in\bar{C}}\mathsf{P}_{X_iY_iR_i|\mathcal{E}}\mathop{\mathbb{E}}\left\|(U_{x_ir_i}\otimes U_{y_ir_i})|\varphi\rangle\langle\varphi|_{y^*r_i}(U^\dagger_{x_ir_i}\otimes U^\dagger_{y_ir_i})-|\varphi\rangle\langle\varphi|_{x_iy_ir_i}\right\|_1\leq\frac{9\varepsilon}{5}.$$

The result then follows by Markov's inequality.

References

1 Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An Information Statistics Approach to Data Stream and Communication Complexity. In *Proceedings of the 43th Annual IEEE Symposium on Foundations of Computer Science, FOCS '02*, pages 209–218, 2002. doi:10.1109/SFCS.2002.1181944.

2 Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to Compress Interactive Communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013. doi:10.1137/100811969.

3 Mohammad Bavarian, Thomas Vidick, and Henry Yuen. Anchoring Games for Parallel Repetition, 2015. arXiv:1509.07466.

4 Mohammad Bavarian, Thomas Vidick, and Henry Yuen. Hardness Amplification for Entangled Games via Anchoring. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC '17, page 303–316, 2017. doi:10.1145/3055399.3055433.

5 Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A Hypercontractive Inequality for Matrix-Valued Functions with Applications to Quantum Computing and LDCs. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS '08*, pages 477–486, 2008. doi:10.1109/FOCS.2008.45.

6 Mario Berta, Matthias Christandl, and Renato Renner. The Quantum Reverse Shannon Theorem Based on One-Shot Information Theory. *Communications in Mathematical Physics*, 306(3):579–615, 2011. doi:10.1007/s00220-011-1309-7.

7 Mark Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015. doi:10.1137/130938517.

8 Mark Braverman, Ankit Garg, Young Kun Ko, Jieming Mao, and Dave Touchette. Near-Optimal Bounds on Bounded-Round Quantum Communication Complexity of Disjointness. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 773–791, 2015. doi:10.1109/FOCS.2015.53.

9 Mark Braverman and Gillat Kol. Interactive Compression to External Information. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC '18, page 964–977, 2018. doi:10.1145/3188745.3188956.

10 Mark Braverman and Anup Rao. Information Equals Amortized Communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014. doi:10.1109/TIT.2014.2347282.

11 Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct Product via Round-Preserving Compression. In *Automata, Languages, and Programming*, volume 7965 of *Lecture Notes in Computer Science*, pages 232–243. Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-39206-1_20.

12 Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct Products in Communication Complexity. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS '13*, pages 746–755, 2013. doi:10.1109/FOCS.2013.85.

13 Mark Braverman and Omri Weinstein. An Interactive Information Odometer and Applications. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, page 341–350, 2015. doi:10.1145/2746539.2746548.

14 Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science, FOCS '01*, pages 270–278, 2001. doi:10.1109/SFCS.2001.959901.

15 Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect Parallel Repetition Theorem for Quantum XOR Proof Systems. *Computational Complexity*, 17(2):282–299, 2008. doi:10.1007/s00037-008-0250-4.

16 Irit Dinur. The PCP Theorem by Gap Amplification. *J. ACM*, 54(3):12–es, 2007. doi:10.1145/1236457.1236459.

17 Irit Dinur, David Steurer, and Thomas Vidick. A Parallel Repetition Theorem for Entangled Projection Games. *Computational Complexity*, 24(2):201–254, 2015. doi:10.1007/s00037-015-0098-3.

18 Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The Communication Complexity of Correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010.

19 Thomas Holenstein. Parallel Repetition: Simplifications and the No-Signaling Case. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, page 411–419, 2007. doi:10.1145/1250790.1250852.

20 Rahul Jain. New Strong Direct Product Results in Communication Complexity. *Journal of the ACM*, 62(3), 2015. doi:10.1145/2699432.

21 Rahul Jain and Hartmut Klauck. New Results in the Simultaneous Message Passing Model via Information Theoretic Techniques. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC '09*, pages 369–378, 2009. doi:10.1109/CCC.2009.28.

22 Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct Product Theorems for Classical Communication Complexity via Subdistribution Bounds: Extended Abstract. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, STOC '08, pages 599–608, 2008. doi:10.1145/1374376.1374462.

23 Rahul Jain and Ashwin Nayak. Short Proofs of the Quantum Substate Theorem. *IEEE Transactions on Information Theory*, 58(6):3664–3669, 2012.

24 Rahul Jain, Attila Pereszlényi, and Penghui Yao. A Parallel Repetition Theorem for Entangled Two-Player One-Round Games under Product Distributions. In *2014 IEEE 29th Conference on Computational Complexity (CCC '14)*, pages 209–216, 2014.

25 Rahul Jain, Attila Pereszlényi, and Penghui Yao. A Direct Product Theorem for Two-Party Bounded-Round Public-Coin Communication Complexity. *Algorithmica*, 76(3):720–748, 2016. doi:10.1007/s00453-015-0100-0.

26 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. The Quantum Communication Complexity of the Pointer Chasing Problem: The Bit Version. In *FSTTCS 2002: Foundations of Software Technology and Theoretical Computer Science*, volume 2556 of *Lecture Notes in Computer Science*, pages 218–229, 2002. doi:10.1007/3-540-36206-1_20.

27 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A Direct Sum Theorem in Communication Complexity via Message Compression. In *Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 300–315. Springer, 2003. doi:10.1007/3-540-45061-0_26.

28 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A Lower Bound for the Bounded Round Quantum Communication Complexity of Set Disjointness. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science, FOCS '03*, pages 220–229. IEEE Computer Society, 2003. doi:10.1109/SFCS.2003.1238196.

29 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior Entanglement, Message Compression and Privacy in Quantum Communication. In *20th Annual IEEE Conference on Computational Complexity (CCC '05)*, pages 285–296, 2005.

30 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Optimal Direct Sum and Privacy Tradeoff Results for Quantum and Classical Communication Complexity, 2008. arXiv:0807.1267.

31 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A Property of Quantum Relative Entropy with an Application to Privacy in Quantum Communication. *Journal of the ACM*, 56(6), 2009. doi:10.1145/1568318.1568323.

32 Rahul Jain and Penghui Yao. A Strong Direct Product Theorem in Terms of the Smooth Rectangle Bound, 2012. arXiv:1209.0263.

33 Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP*=RE, 2020. arXiv:2001.04383.

34 Julia Kempe, Oded Regev, and Ben Toner. Unique Games with Entangled Provers are Easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010. doi:10.1137/090772885.

35 Hartmut Klauck. A Strong Direct Product Theorem for Disjointness. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, pages 77–86, 2010. doi:10.1145/1806689.1806702.

36 Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and Classical Strong Direct Product Theorems and Optimal Time-Space Tradeoffs. *SIAM Journal on Computing*, 36(5):1472–1493, 2007. doi:10.1137/05063235X.

37 Gillat Kol. Interactive Compression for Product Distributions. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 987–998, 2016. doi:10.1145/2897518.2897537.

38 Troy Lee, Adi Shraibman, and Robert Špalek. A Direct Product Theorem for Discrepancy. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC '08*, pages 71–80, 2008. doi:10.1109/CCC.2008.25.

39 Ran Raz. A Parallel Repetition Theorem. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, page 447–456, 1995. doi:10.1145/225058.225181.

40 Alexander A. Razborov. On the Distributional Complexity of Disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992. doi:10.1016/0304-3975(92)90260-M.

41 Ronen Shaltiel. Towards Proving Strong direct Product Theorems. *Computational Complexity*, 12(1-2):1–22, 2003. doi:10.1007/s00037-003-0175-x.

42 Alexander A. Sherstov. Strong Direct Product Theorems for Quantum Communication and Query Complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012. doi:10.1137/110842661.

43 Alexander A. Sherstov. Compressing Interactive Communication Under Product Distributions. *SIAM Journal on Computing*, 47(2):367–419, 2018. doi:10.1137/16M109380X.

44 Emanuele Viola and Avi Wigderson. Norms, XOR Lemmas, and Lower Bounds for Polynomials and Protocols. *Theory of Computing*, 4(7):137–168, 2008. doi:10.4086/toc.2008.v004a007.

45 Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (SFCS 1977)*, pages 222–227, 1977. doi:10.1109/SFCS.1977.24.

46 Henry Yuen. A Parallel Repetition Theorem for All Entangled Games. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP '16)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 77:1–77:13, 2016. doi:10.4230/LIPIcs.ICALP.2016.77.