Electronic Colloquium on Computational Complexity, Report No. 78 (2021)

A direct product theorem for quantum communication complexity with applications to device-independent QKD

Rahul Jain [*] Srijita Kundu [†]

Abstract

We give a direct product theorem for the entanglement-assisted interactive quantum communication complexity of an l-player predicate V. In particular we show that for a distribution p that is product across the input sets of the l players, the success probability of any entanglement-assisted quantum communication protocol for computing n copies of V, whose communication is $o(\log(\mathrm{eff}^*(\mathsf{V}, p)) \cdot n)$, goes down exponentially in n. Here $\mathrm{eff}^*(\mathsf{V}, p)$ is a distributional version of the quantum efficiency or partition bound introduced by Laplante, Lerays and Roland (2014), which is a lower bound on the distributional quantum communication complexity of computing a single copy of V with respect to p. For a two-input boolean function f, the best result for interactive quantum communication complexity known so far is due to Sherstov (2012), who showed a direct product theorem in terms of the generalized discrepancy, which is a lower bound on communication. Our lower bound on non-distributional communication complexity is in terms of $\max_{\text{product } p} \mathrm{eff}^*(\mathsf{V}, p)$, and there is no known relationship between this and the generalized discrepancy. But we define a distributional version of the generalized discrepancy bound and can show that for a given p, $\mathrm{eff}^*(\mathsf{V}, p)$ upper bounds it. Moreover, unlike Sherstov's result, our result works for two-input functions or relations whose outputs are non-boolean as well, and is a strong direct product theorem for functions or relations whose quantum communication complexity is characterized by $\mathrm{eff}^*(\mathsf{V}_f, p)$ for a product p.

As an application of our result, we show that it is possible to do device-independent quantum key distribution (DIQKD) without the assumption that devices do not leak any information after inputs are provided to them. We analyze the DIQKD protocol given by Jain, Miller and Shi (2017), and show that when the protocol is carried out with devices that are compatible with n copies of the Magic Square game, it is possible to extract $\Omega(n)$ bits of key from it, even in the presence of $O(n)$ bits of leakage. Our security proof is parallel, i.e., the honest parties can enter all their inputs into their devices at once, and works for a leakage model that is arbitrarily interactive, i.e., the devices of the honest parties Alice and Bob can exchange information with each other and with the eavesdropper Eve in any number of rounds, as long as the total number of bits or qubits communicated is bounded.

1 Introduction

Communication complexity is an important model of computation with connections to many parts of theoretical computer science [KN96]. In this paper, we consider the communication com-

[*]Centre for Quantum Technologies and Department of Computer Science, National University of Singapore and MajuLab, UMI 3654, Singapore. Email: `rahul@comp.nus.edu.sg`

[†]Centre for Quantum Technologies, National University of Singapore, Singapore. Email: `srijita.kundu@u.nus.edu`

ISSN 1433-8092

plexity of computing a predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$ by $l(\geq 2)$ players who receive inputs $x^1 \ldots x^l \in \mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, and after communicating interactively, are required to produce outputs $a^1 \ldots a^l$ such that $\mathsf{V}(a^1 \ldots a^l, x^1 \ldots x^l)$ is satisfied. The l players cooperate and wish to minimize the total number of bits (in the classical model) or qubits (in the quantum model) communicated. The communication complexity of predicates generalizes the communication complexity of (total or partial) functions and relations that are most often considered in the literature.

In any model of computation, a fundamental question is: if we know how to do one copy of a task, what is the best way to do n independent copies of it? One possible way is to simply each copy independently; if we have an algorithm that successfully does a single copy of the task with probability $1-\varepsilon$, the success probability of this product strategy is $(1-\varepsilon)^n$ and its cost is n times the cost of doing a single copy. For many tasks, this is the best one can do, and a direct product theorem for the task proves so. That is, a direct product theorem proves that any protocol for doing n copies of the task that has cost at most cn, where c is some lower bound on the cost of doing one copy with success probability less than 1, has success probability exponentially small in n. When c is the exact cost of doing a single copy of the task, we call such a result a strong direct product theorem.

Direct product theorems are known in a number of computational models. In classical communication complexity, there is a long line of works showing direct product and weaker direct sum theorems (which show that the success probability of a protocol that uses cn resources is at most constant, instead of exponentially small) in the two-party setting [Raz92, CSWY01, BYJKS02, JRS05, KŠdW07, VW08, LSŠ08, HJMR10, BR11, JY12, BBCR13, BRWY13a, BRWY13b, JPY16].

For quantum communication, a direct sum theorem for one-way quantum communication for general functions was shown by [JRS05], and [JK20] showed a direct product theorem for the same. In the interactive quantum setting however, direct product theorems are known only for special classes of functions, for example [Kla10] showed a direct product theorem for symmetric functions. [She18] showed a direct product theorem for the generalized discrepancy method, which is one of the strongest lower bound techniques on quantum communication complexity — this gives a strong direct product theorem for functions whose quantum communication complexity is exactly characterized by the generalized discrepancy method.

Direct product theorems in communication are related to parallel repetition theorems for non-local games. A non-local game with l players is defined by a predicate V and a distribution p. The players are given inputs $x^1 \ldots x^l$ from distribution p on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, and they are required to produce outputs $a^1 \ldots a^l$ in $\mathcal{A}^1 \times \ldots \times \mathcal{A}^l$ so that $\mathsf{V}(a^1 \ldots a^l, x^1 \ldots x^l$ is satisfied, without communicating. In the classical model, the players are allowed to share randomness, and in the quantum model they are allowed to share entanglement. The maximum winning probability of the game over all strategies is called the value of the game, which may be quantum or classical. A parallel repetition theorem shows that the value of n independent instances of a non-local game is $(1-\varepsilon)^{\Omega(n)}$, if the value a single instance is $(1-\varepsilon)$.

A parallel repetition theorem for the classical value of general two-player non-local games was first shown by Raz [Raz95], and the proof was subsequently simplified by Holenstein [Hol07]. 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 two-player games such as XOR games [CSUU08], unique games [KRT10] and projection games [DSV15]. When the type of game is not restricted but the input distribution is, parallel repetition theorems have been shown under product distributions [JPY14] and anchored distributions [BVY17] — both of these results can be extended to l players. For general two-player games, the best current result is due to Yuen [Yue16], which shows that the quantum value of n parallel instances of a general game goes down polynomially in n, if the quantum value of the original game is strictly less than 1. The situation for more than 2 players is much less understood.

Device-independent cryptography. Quantum cryptography lets us do a number of tasks with information theoretic security, i.e., security without any computational assumptions, that are not possible classically. One such example is quantum key distribution (QKD) [BB84]. In a key distribution scenario, two honest parties Alice and Bob want to share a key, i.e., a uniformly random string of a given length, which is secret from a third party eavesdropper Eve. If Alice and Bob have access to secure private randomness and an authenticated classical channel, it is possible to do the key distribution task quantumly with information theoretic security, but not classically. In a conventional security proof for QKD (or any other quantum cryptographic protocol), one needs to have a complete description of the quantum devices, i.e., the states and measurements used by Alice and Bob. However, in practice quantum devices are often not fully characterized, and protocols that rely on complete characterization of quantum devices often have loopholes.

A way around this problem is the framework of device-independent cryptography, which tries to give quantum protocols for cryptographic tasks that are secure even when the devices used by the honest parties are not fully characterized, and in fact can be arbitrarily manipulated by dishonest parties. All known device-independent protocols with information theoretic security use non-local games and rely on the property of self-testing or rigidity displayed by some non-local games. Suppose we play a non-local game with devices implementing some unknown state and measurements, and in fact even the dimension of the systems are unspecified. If these state and measurements regardless achieve a winning probability for the game that is close to its optimal winning probability, then self-testing tells us that the state and measurements are close to the ideal state and measurements for that game, up to trivial isometries. For device-independent QKD (DIQKD), this means in particular that the measurement outputs of the devices given the inputs are random, i.e., they cannot be predicted by a third party even if they have access to the inputs used. This lets us use the outputs of the devices to produce a secret key.

A number of protocols and security proofs for DIQKD have been given over the years, in the sequential [PAB+09, AFDF+18, VV19] as well as parallel setting [JMS20, Vid17]. Aside from assuming that Alice and Bob's devices are modelled by quantum mechanics however, all these proofs require the assumpion that Alice and Bob's devices do not leak any information, i.e., do not communicate with each other or with Eve, unbeknownst to Alice and Bob. Although there have been some works studying non-local games in the presence of communication [TZCBB+20, TZCWP20], and and argument showing device-independent may be possible in the presence of a specific model of information leakage in [SPM13], none of these approaches have been developed into a full-fledged proof of security when there is leakage.

1.1 Our results

1.1.1 Direct product theorem

Let $\mathsf{V}(a^1 \ldots a^l, x^1 \ldots, x^l)$ be a predicate on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$. We shall use $\mathsf{V}^n(a_1^1 \ldots a_1^l \ldots a_n^1 \ldots a_n^l, x_1^1 \ldots x_1^l \ldots x_n^1 \ldots x_n^l)$ to denote n independent copies of V, i.e., the predicate which is satisfied when all n $(a_i^1 \ldots a_i^l, x_i^1 \ldots x_i^l)$-s satisfy V.

For a probability distribution p on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, a (quantum) communication protocol between l parties that takes inputs from $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$ and produces outputs in $\mathcal{A}_1 \times \ldots \times \mathcal{A}_n$, produces a conditional probability distribution on $\mathcal{A}^1 \times \ldots \times \mathcal{A}^l$ conditioned on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, and along with p there is an induced distribution on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \mathcal{X}^l)$. Let $\mathrm{suc}(p, \mathsf{V}, \mathcal{P})$ be the probability that the predicate V is satisfied according to this distribution.

Let $\mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$ denote the distributional quantum partition bound with error ε for V with respect to input distribution p, which we shall define formally in Section 4. $\mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$ is a lower bound on the quantum communication complexity of V. Let $\omega^*(G(p, \mathsf{V}))$ denote the quantum value of the non-local game $G = (p, \mathcal{X}^1 \times \ldots \times \mathcal{X}^l, \mathcal{A}^1 \times \ldots \times \mathcal{A}^l, \mathsf{V})$.

With this notation, our direct product theorem is stated below.

Theorem 1. *For any $\varepsilon, \zeta > 0$, any predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$ and any product probability distribution p on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, if $\mathcal{P}$ is an interactive entanglement-assisted quantum communication protocol between l parties which has total communication cn.*

(i) If $c < 1$, then

$$\mathrm{suc}(p^n, \mathsf{V}^n, \mathcal{P}) \le \left(1 - \frac{\nu}{2} + 4\sqrt{lc}\right)^{\Omega\left(\nu^2 n / \left(l^2 \cdot \log(|\mathcal{A}^1| \cdot \ldots \cdot |\mathcal{A}^l|)\right)\right)}$$

where $\nu = 1 - \omega^*(G(p, \mathsf{V}))$.

(ii) If $1 \le c = O\left(\frac{\zeta^2}{l^3} \mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V}, p)\right)$, then

$$\mathrm{suc}(p^n, \mathsf{V}^n, \mathcal{P}) \le (1 - \varepsilon)^{\Omega\left(n / \left(\log(|\mathcal{A}^1| \cdot \ldots \cdot |\mathcal{A}^l|)\right)\right)}.$$

The two cases in Theorem 1 should be interpreted as follows: $c < 1$ means there is less than one qubit of communication per copy of V, and we are close to the non-local game situation where there is no communication. Therefore we get an upper bound on the success probability for computing V^n in terms of the winning probability of the corresponding game. The theorem in this case is essentially saying that parallel-repeated non-local games under product distributions are resistant to communication, i.e., if the winning probability of n copies of the game goes does exponentially in n, then it also goes down exponentially in n if there is a small amount of communication. We also remark that in case (i), a corresponding theorem can also be proved if p is an anchored distribution, which was introduced in [BVY17], instead of a product distribution. We expand on this more in Section 2.

The case $c \ge 1$ means on average at least one qubit is communicated per copy of V. This corresponds to the true communication scenario, and thus if c is less than a lower bound on the per copy communication complexity of V, we get that the probability of success for computing V^n goes down exponentially in n. By Yao's Lemma, case (ii) of Theorem 1 has the following corollary for communication complexity.

Corollary 2. *For a predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$, let $\mathrm{Q}_\varepsilon(\mathsf{V})$ denote the interactive entanglement-assisted quantum communication complexity of computing it, and $\mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V}, p)$ be its distributional quantum partition bound for distribution p, and any $\varepsilon, \zeta > 0$. Then,*

$$\mathrm{Q}_{1-(1-\varepsilon)^{\Omega\left(n/\left(\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)\right)\right)}}(\mathsf{V}^n) = \Omega\left(\frac{\zeta^2 n}{l^3}\left(\max_{\text{product } p} \log \mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V}, p)\right)\right).$$

Corollary 2 is a strong direct product theorem for predicates whose interactive entanglement-assisted communication complexity is characterized by $\max_{\text{product } p} \mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$.

1.1.2 Applications in two-party communication complexity of functions

In the communication complexity setting for a two-input function or relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, we normally require that only one party gives an output. Nevertheless, we can define a predicate V_f for it in which one party has a singleton output set, say $\{\top\}$, and the other party's output set is $\mathcal{Z}$. We define

$$\mathsf{V}_f(\top z, xy) = 1 \quad \Longleftrightarrow \quad z \in f(x,y).$$

It is clear then that the two-party communication complexity of f is equal to the communication complexity of V_f.

In [ABJO21], it is shown that a large class of functions exists, whose quantum communication complexity is characterized by $\mathrm{eff}^*_\varepsilon(\mathsf{V}_f, p)$ for a product p. In particular, they show that a class of functions known as two-wise independent functions, $\mathrm{eff}^*_\varepsilon(\mathsf{V}_f, p)$ takes the maximum possible value of the uniform distribution, which is product.

Fact 1 ([ABJO21])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a two-wise independent function with $|\mathcal{X}| = |\mathcal{Y}|$, and let p_U be the uniform distribution on $\mathcal{X} \times \mathcal{Y}$. Then for any $\varepsilon > 0$,*

$$\mathrm{eff}^*_\varepsilon(\mathsf{V}_f, p_U) \geq \frac{|\mathcal{X}|}{|\mathcal{Z}|}\left(1 - \gamma - \frac{1}{|\mathcal{Z}|}\right)^2.$$

An example of a two-wise independent function is the generalized inner product $\mathrm{IP}^n_q : \mathbb{F}^n_q \times \mathbb{F}^n_q \to \mathbb{F}_q$ defined by:

$$\mathrm{IP}^n_q(x,y) = \sum_{i=1}^n x_i y_i \mod q.$$

This makes our result the first strong direct product theorem for generalized inner product that we are aware of. The direct product theorem in terms of the generalized discrepancy method by Sherstov [She18] works only for boolean-output functions, and gives a strong direct product theorem for quantum communication of IP^n_2.

For further comparison between our direct product theorem and Sherstov's, we prove Theorem 3. For a total function $f : \mathcal{X} \times \mathcal{Y} \to \{-1, +1\}$, let F denote the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix whose $[x, y]$-th entry is given by $f(x,y)$. The generalized discrepancy method lower bounds communication in terms of $\log \gamma^\alpha_2(F)$, where $\gamma^\alpha_2(M)$ is the α-approximate factorization norm of a matrix M. For a function f, $\gamma^\alpha_2(F)$ can be expressed as $\max_p \gamma^\alpha_2(F, p)$ where $\gamma^\alpha_2(F, p)$ is a distributional version of $\gamma^\alpha_2(F)$ with respect to p over $\mathcal{X} \times \mathcal{Y}$.

Theorem 3. *For a total function $f : \mathcal{X} \times \mathcal{Y} \to \{-1, +1\}$, let V_f denote the predicate on $(\{-1, +1\})^2 \times (\mathcal{X} \times \mathcal{Y})$ given by*

$$\mathsf{V}(ab, xy) = 1 \quad \Longleftrightarrow \quad a \cdot b = f(x,y).$$

Then for any distribution p on $\mathcal{X} \times \mathcal{Y}$,

$$\mathrm{eff}^*_\varepsilon(\mathsf{V}_f, p) \geq (1 - 2\varepsilon)\gamma^\alpha_2(F, p)$$

with $\alpha = \frac{1+2\varepsilon}{1-2\varepsilon}$.

This shows that $\mathrm{eff}^*_\varepsilon(\mathsf{V}_f, p)$ is a stronger lower bound technique than $\gamma_2(F, p)$ for boolean f. However, since our direct product theorem is in terms of $\max_{\text{product } p} \log \mathrm{eff}^*_\varepsilon(\mathsf{V}_f, p)$, and Sherstov's in terms of $\max_p \log \gamma_2(F, p)$, the two results cannot be directly compared. We also note that we are only able to show the relationship between $\mathrm{eff}^*_\varepsilon(\mathsf{V}_f, p)$ and $\gamma^\alpha_2(F, p)$ for total f, whereas Sherstov's direct product result works for partial functions as well.

1.1.3 DIQKD secure against leakage

Leakage model. In the device-independent setting, each honest party's device is modelled as a black box, into which the party provides inputs and from which they get outputs to play a non-local game. Ideally the boxes play n independent copies of the non-local game, although they may do so noisily, i.e., each game is won with probability δ-close to its optimal quantum value. For DIQKD, the honest parties are Alice and Bob and we assume their boxes are supplied by the eavesdropper Eve. The states and measurements implemented by these boxes may be very far from those corresponding to the two-player non-local game that each of Alice and Bob's boxes ideally play. In fact, instead of Alice and Bob sharing an entangled state that is uncorrelated with anything else, Eve may hold a purification of Alice and Bob's state, which we also model as a box.

As mentioned before, known DIQKD protocols rely on the assumption that Alice and Bob and Eve's boxes do not communicate with each other. We relax the assumption in a strong way: we assume Alice, Bob and Eve's boxes can all send classical messages to each other (since they share entanglement, this means they can also effectively exchange quantum states via teleportation) after Alice and Bob have entered their inputs into their boxes and before they receive their outputs. The communication between Alice, Bob and Eve's boxes may be arbitrarily interactive: we do not put any bound on the number of rounds of communication, only on the total number of bits communicated.

For the sake of concreteness, we analyze the parallel DIQKD protocol given by [JMS20] under this leakage model, but in principle the same analysis could be applied to any DIQKD protocol that is based on a non-local game that has: (i) a product input distribution, and (ii) a common bit that Alice and Bob can ideally both know given their outputs a and b, and both parties' inputs x and y (and this bit is their shared key). Using case (i) of Theorem 1, we prove the following theorem.

Theorem 4. *There are universal constants $0 < \delta_0 < 1$ and $0 < c_0 < 1$ such that for any $0 \leq \delta \leq \delta_0$, and $0 \leq c \leq c_0$, if the [JMS20] DIQKD protocol (given in Protocol 1) is carried out with boxes that play n copies of the Magic Square game δ-noisily, it is possible to extract $r(\delta, c)n$ bits of secret key in the interactive leakage model, with the total communication between Alice, Bob and Eve's boxes being cn bits, for some $r(\delta, c) > 0$.*

Remark 1. *In practice Alice and Bob's boxes can also continue sending messages after their outputs are produced (so can Eve's, but the keyrate depends on Eve's probability of guessing Alice and Bob's outputs, which cannot change due to her box sending messages to Alice and Bob's boxes after they have produced their outputs, so we ignore that at this time). But as far as security analysis is concerned, this communication is equivalent to communication between Alice and Bob over public channels after they have obtained their outputs, which is a standard part of QKD protocols and can be handled by standard DIQKD proof techniques. Using standard techniques, the amount of communication after the outputs are produced would just be subtracted from the key rate, and after a certain threshold of communication, key rate would just be zero. Communication before Alice and Bob's outputs are produced cannot be handled by standard techniques, however, and hence we focus on this in the above theorem.*

We also note that though we give a specific proof only for DIQKD with leakage, our proof technique can be seen as a general framework for making device-independent protocols that use parallel repetition theorems in their security proofs, secure against leakage. For example, this technique can also be applied to the device-independent protocol for encryption with certified deletion given by [KT20]. The security proof for that protocol uses a parallel repetition theorem for an anchored two-round game (where players receive two rounds of inputs and give two rounds of outputs). As we have already said, a version of Theorem 1 in case (i) also applies to anchored distributions for one-round games, and it is not difficult to generalize to two-round games by considering an appropriate round-by-round leakage model.

1.2 Organization of the paper

In Section 2 we give an overview of our proofs. In Section 3 we provide definitions and known results about the quantities used in our proofs. In Section 4, we introduce variants of the quantum partition bound, prove that they lower bound communication and also Theorem 3. In Section 5, we prove a lemma called the Substate Perturbation Lemma, which is a main tool for our direct product theorem. In Section 6, we give the proof of our main direct product theorem. Finally, in Section 7 we show the application of our direct product theorem to prove security of DIQKD with leakage.

2 Proof overview

2.1 Direct product theorem

We follow the information-theoretic framework for parallel repetition and direct product theorems introduced by [Raz95] and [Hol07]. The idea is this: take a protocol $\mathcal{P}$ for V^n that is "too good". We condition on the success in some t coordinates in this protocol, and show that either the probability of success in these coordinates is already small, or there is an i in the other $n-t$ coordinates such that the probability of success of i conditioned on success event $\mathcal{E}$ is bounded away from 1. This is done by showing that if the probability of $\mathcal{E}$ and the probability of success in i conditioned on $\mathcal{E}$ are both large, we can give a protocol $\mathcal{P}'$ for V that is "too efficient". Now our lower bound in the $c \geq 1$ case is in terms of $\mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$, which intuitively speaking, corresponds to the inverse of the maximum probability of not aborting in a zero-communication protocol in which the l parties either abort, or produce outputs that satisfy V with probability at least $1-\varepsilon$ (conditioned on not aborting). Therefore, $\mathcal{P}'$ for us will be a zero-communication protocol with aborts that computes V with high probability conditioned on not aborting, whose probability of not aborting is too high.

For simplicity, we shall give an overview of the proof with only two parties Alice and Bob; the proof for l parties follows similarly. When Alice and Bob's inputs are x_i and y_i respectively at the i-th coordinates in $\mathcal{P}$, we define a state $|\varphi\rangle_{x_iy_i}$ that represents the state at the end of $\mathcal{P}$ conditioned on $\mathcal{E}$. Considering the state at the end instead of round by round is the same approach as that taken in [JRS05], who use it to show a direct sum theorem. On input (x_i, y_i) in $\mathcal{P}'$, Alice and Bob will try to either abort, or get a shared state close to $|\varphi\rangle_{x_iy_i}$. Once they have this state, they can perform measurements on the i-th output registers to give their outputs (a_i, b_i). Their output distribution will be close to the output distribution in the i-th coordinate of $\mathcal{P}$ conditioned on $\mathcal{E}$; hence if the probability of success on i conditioned on $\mathcal{E}$ is too large, the probability of Alice and Bob correctly computing V in $\mathcal{P}'$ conditioned on not aborting is also large. Hence our proof mainly consists of showing how Alice and Bob can get the shared state close to $|\varphi\rangle_{x_iy_i}$ with probability of aborting $2^{-O(c)}$, where cn is the communication in $\mathcal{P}$. Since the probability of aborting in $\mathcal{P}'$ cannot be smaller than eff^*, this gives the desired lower bound on the communication of $\mathcal{P}$ in terms of eff^*.

In the $c < 1$ case, our proof is very similar to the proof of a parallel repetition theorem for non-local games with product distributions due to [JPY14]. The main difference between that $c \geq 1$ case and the parallel repetition of $c < 1$ case is that in the latter, we need to show that Alice and Bob can get the shared state $|\varphi\rangle$ by local unitaries (without aborting). We briefly describe their proof below.

Parallel repetition for games under product distribution. Let $|\varphi\rangle_{x_i}$ be the superposition of $|\varphi\rangle_{x_iy_i}$ over the distribution of Y_i, $|\varphi\rangle_{y_i}$ be the superposition over the distribution of X_i, and $|\varphi\rangle$ be the superposition over both. 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 V_{y_i} acting on Bob's registers that take $|\varphi\rangle$ close to $|\varphi\rangle_{y_i}$.
3. By applying the quantum operation that measures the X_i register and records the outcome, it can be shown that V_{y_i} also takes $|\varphi\rangle_{x_i}$ to $|\varphi\rangle_{x_iy_i}$.
4. Since U_{x_i} and V_{y_i} act on disjoint registers, $U_{x_i} \otimes V_{y_i}$ then takes $|\varphi\rangle$ close to $|\varphi\rangle_{x_iy_i}$.

Alice and Bob can thus share $|\varphi\rangle$ as entanglement, and get close to $|\varphi\rangle_{x_iy_i}$ by local unitariess U_{x_i} and V_{y_i}. In case (i) of our proof, everything is similar to this, except that the distance between $|\varphi\rangle$ and $|\varphi\rangle_{x_i}$ also accounts for c^{A}, $c^{\mathrm{A}}n$ being Alice's total communication to Bob, and the distance between $|\varphi\rangle$ and $|\varphi\rangle_{y_i}$ also accounts for c^{B}, $c^{\mathrm{B}}n$ being Bob's communication.

If we wish a give a proof for case (i) with anchored distributions instead of product distributions, we would need to follow the equivalent steps in the proof of the parallel repetition theorem for anchored games given in [BVY17] or the alternative proof given in [JK20] instead, and account for communication there.

Direct product for communication under product distribution. In case $c \geq 1$, we cannot use Uhlmann unitaries to go from $|\varphi\rangle$ to $|\varphi\rangle_{x_i}$ and $|\varphi\rangle_{y_i}$, as there is a lot of dependence between Alice's registers and Bob's registers due to communication. But we can use a compression scheme due to [JRS02, JRS05] which says that if the mutual information between X_i and Bob's registers is c, then there exist projectors Π_{x_i} acting on Alice's registers which succed on $|\varphi\rangle$ with probability 2^{-c}, and on success take it close to $|\varphi\rangle_{x_i}$. Following parallel repetition proof we can show:

1. If the total communication from Alice to Bob in $\mathcal{P}$ is $c^{\mathrm{A}}n$, then the mutual information between $X_1 \dots X_n$ and Bob's registers in $|\varphi\rangle$ is $O(c^{\mathrm{A}}n)$. By chain rule of mutual information, there exists an i such that the mutual information between X_i and Bob's registers is $O(c^{\mathrm{A}})$, and hence there exist projectors Π_{x_i} acting on Alice's registers which succeed with probability $2^{-O(c^{\mathrm{A}})}$ on $|\varphi\rangle$ and on success take $|\varphi\rangle$ close to $|\varphi\rangle_{x_i}$.
2. Similarly, if the total communication from Bob to Alice in $\mathcal{P}$ is $c^{\mathrm{B}}n$, then there exist projectors Π_{y_i} acting on Bob's registers which succeed with probability $2^{-O(c^{\mathrm{B}})}$ on $|\varphi\rangle$ and on success take $|\varphi\rangle$ close to $|\varphi\rangle_{y_i}$.

3. By applying the same argument with the operation measuring the X_i register and recording the outcome, it can be shown that Π_{y_i} succeeds on $|\varphi\rangle_{x_i}$ with probability $2^{-O(c^{\mathrm{B}})}$ and on success takes it close to $|\varphi\rangle_{x_i y_i}$.

However, unlike in the case of unitaries, even though Π_{x_i} and Π_{y_i} commute, there is a problem in combining items 2 and 3 above to say that $\Pi_{x_i} \otimes \Pi_{y_i}$ succeed on $|\varphi\rangle$ with probability $2^{-O(c^{\mathrm{A}}+c^{\mathrm{B}})}$ and on success take it close to $|\varphi\rangle_{x_i y_i}$. Since $\sqrt{\frac{1}{2^{-O(c^{\mathrm{A}})}}}\Pi_{x_i}|\varphi\rangle$ (i.e., the normalized state on success of Π_{x_i} on $|\varphi\rangle$) is only close to $|\varphi\rangle_{x_i}$ rather than exactly equal to it, acting Π_{y_i} on this state cannot take it close to $|\varphi\rangle_{x_i y_i}$, unless the distance between $\sqrt{\frac{1}{2^{-O(c^{\mathrm{A}})}}}\Pi_{x_i}|\varphi\rangle$ and $|\varphi\rangle_{x_i}$ is of the same order as the success probability of Π_{y_i} on $|\varphi\rangle_{x_i}$. This distance figures in the exponent in the success probability $2^{-O(c^{\mathrm{A}})}$, so we cannot afford to make it that small.

Instead we shall directly try to get projectors Π'_{y_i} that succeed with high probability on $|\rho\rangle$, which is we what we call the superposition over X_i of $\sqrt{\frac{1}{2^{-O(c^{\mathrm{A}})}}}\Pi_{x_i}|\varphi\rangle$, and on success take it close to $|\varphi\rangle_{y_i}$ (these will also take $|\rho\rangle_{x_i}$ close to $|\varphi\rangle_{x_i y_i}$). Since we do not have a bound on the mutual information between Y_i and Alice's registers in ρ, we prove what we call the Substate Perturbation Lemma in order to do this. The quantity that is actually of relevance in the [JRS05] compression scheme is the smoothed relative min-entropy $\mathsf{D}^{\varepsilon}_{\infty}$ between $\varphi_{Y_i A}$ and $\varphi_{Y_i} \otimes \varphi_A$ (A being Alice's registers), which is $O(c^{\mathrm{B}}/\varepsilon^2)$ if the mutual information between Y_i and A is $O(c^{\mathrm{B}})$, due to the Quantum Substate Theorem [JRS02, JRS09, JN12]. In the Substate Perturbation Lemma, which is one of our main technical contributions, we show that if $D^{\varepsilon}_{\infty}(\varphi_{Y_i A}\|\varphi_{Y_i} \otimes \varphi_A)$ is c' and ρ_A and φ_A are δ-close, then $\mathsf{D}^{3\varepsilon+\delta}_{\infty}(\varphi_{Y_i A}\|\varphi_{Y_i} \otimes \rho_A)$ is $O(c')$. Using the [JRS05] compression scheme, this lets us get projectors Π'_{y_i} on Bob's registers that succeed with probability $2^{-O(c^{\mathrm{B}})}$ on $|\rho\rangle$ and on success take it close to $|\varphi\rangle_{y_i}$.

The protocol $\mathcal{P}'$ will thus involve the following: Alice and Bob share $|\varphi\rangle$ as entanglement and on inputs (x_i, y_i), apply the measurements $\{\Pi_{x_i}, \mathbb{1} - \Pi_{x_i}\}$ and $\{\Pi'_{y_i}, \mathbb{1} - \Pi'_{y_i}\}$ on it. They abort if the Π_{x_i} or Π'_{y_i} projector does not succeed. Since $\Pi_{x_i} \otimes \Pi'_{y_i}$ succeeds on $|\varphi\rangle$ with probability $2^{-O(c^{\mathrm{A}}+c^{\mathrm{B}})} = 2^{-O(c)}$, $\mathcal{P}'$ does not abort with probability $2^{-O(c)}$ and on not aborting, gets a state close to $|\varphi\rangle_{x_i y_i}$.

2.2 Security of DIQKD with leakage

The [JMS20] protocol is based on the Magic Square non-local game. In a single copy of the Magic Square game, henceforth denoted by MS, Alice and Bob receive trits x and y and are required to output 3-bit strings a and b which respectively have even and odd parity; they win the game if their outputs satisfy the condition $a[y] = b[x]$. In the [JMS20] protocol, Alice and Bob have boxes which are compatible with n copies of MS. Using trusted private randomness, Alice and Bob generate i.i.d. inputs x_i, y_i for each game and generate outputs a_i, b_i. The inputs x_i, y_i are then publicly communicated. Alice and Bob select a small random subset of $[n]$ to test the MS winning condition on, i.e., they check if $a_i[y_i] = b_i[x_i]$ for i in that subset (up to error tolerance). If the test passes, they select $K^{\mathrm{A}} = (a_i[y_i])_i$ and $K^{\mathrm{B}} = (b_i[x_i])_i$ as their raw secret keys ; otherwise the protocol aborts. Due to error correction and privacy amplification, we can get a linear amount of secret key from this scheme if we can show

$$\mathsf{H}^{\varepsilon}_{\infty}(K^{\mathrm{A}}|\widetilde{E}')_{\rho} - \mathsf{H}^{\varepsilon}_{0}(K^{\mathrm{A}}|K^{\mathrm{B}})_{\rho} = \Omega(n),$$

where $\mathsf{H}^\varepsilon_\infty$ is the ε-smoothed conditional min-entropy and H^ε_0 is the ε-smoothed conditional Hartley entropy, ρ is the shared state of Alice, Bob and Eve conditioned on not aborting, and $\widetilde{E}'$ is everything Eve holds at the end of the protocol, including a quantum purification of Alice and Bob's systems and also the classical information X_iY_i that Alice and Bob have communicated publicly.

Challenges in a sequential security proof. Most security proofs for DIQKD work in the sequential setting, where Alice and Bob have to enter their inputs into their boxes and get their outputs one by one; in particular the sequential security proofs require the assumption that the $(i-1)$-th output is recorded before the box receives the i-th input. Sequential security proofs generally give better parameters than parallel ones, but we do not how to apply techniques for sequential proofs in the setting with leakage without fairly unnatural assumptions.

For example, one tool widely used in sequential security proofs is the Entropy Accumulation Theorem [DFR20, AFRV19]. Suppose the information released to Eve in the i-th round (in the sequential setting we call each time Alice and Bob enter inputs x_i, y_i into their box, a round) is T_i, and Eve's quantum register is $\widetilde{E}$. Then in order to apply the Entropy Accumulation Theorem to bound $\mathsf{H}^\varepsilon_\infty(K^{\mathrm{A}}|T_1 \ldots T_n\widetilde{E})_\rho$, we require the Markov condition $(A_1 \ldots A_{i-1}) - (T_1 \ldots T_{i-1}\widetilde{E}) - T_i$ for all i, i.e., the information leaked in the i-th round is independent of the Alice's outputs of the rounds before i, given Eve's side information before the i-th round. In the setting without leakage, T_i is just Alice and Bob's inputs X_iY_i for the i-th round, which are picked with trusted private randomness, and thus can be made independent of everything else. In the setting with leakage however, T_i would include the information leaked by Alice and Bob's boxes in the i-th round as well. Once we allow the boxes to leak information, there is nothing stopping them from leaking information about the outputs of the $(i-1)$-th round in the i-th round. Thus imposing the Markov condition here feels fairly unnatural, and closes off the possibility of using Entropy Accumulation in the model with leakage.

Parallel security proof. Instead we closely follow the approach of [JMS20] in giving a parallel security proof for their protocol. Here "parallel" means that their security proof works when Alice and Bob enter all their inputs into their boxes at once, and no Markov condition is required. The security proof of [JMS20] is based on the parallel repetition theorem for non-local games under product distributions [JPY14]. Since we are working in the setting with leakage, instead of a parallel repetition theorem for games, we use our direct product theorem for communication. The communication setting with 3 players exactly corresponds to the leakage model between the parties Alice, Bob and Eve in QKD. Case (i) of our direct product theorem says that if communication is cn for for sufficiently small $c < 1$, then the probability of computing n copies of a non-local game's predicate correctly goes down exponentially in n.

The game we consider is a three-player version of MS, which is a hybrid of the games considered by [JMS20] and [Vid17], which gives a simplified version of the [JMS20] proof. In this game which we call MSE, Alice and Bob play MS between them, and in addition Eve, who has no input, has to guess both their inputs x, y, and Alice's output bit $a[y]$.[1] The winning probability of this game is strictly smaller than $\frac{1}{9}$ (which is Eve's probability of correctly guessing x, y). Due to our direct product result, in the presence of a bounded amount of communication before the outputs are produced, the winning probability of n copies of this game is $\left(\frac{1}{9}(1-\nu)\right)^{\Omega(n)}$ for some $\nu > 0$.

Since Alice and Bob have performed the test to see that $a_i[y_i] = b_i[x_i]$ on a random subset,

[1] Due to technical reasons, we also need to include the following feature in the game: Alice gets an additional input bit z, and Alice and Bob's winning condition $a[y] = b[x]$ not being satisfied is forgiven if Eve is able to guess z.

this condition is satisfied in most locations with high probability conditioned on not aborting. Therefore, MSE is won if Eve can correctly guess $x_i, y_i, a_i[y_i]$. Now, suppose $\varphi_{K^A K^B X_1 \ldots X_n Y_1 \ldots Y_n \widetilde{E}}$ is the shared quantum state before $x_1 \ldots x_n, y_1 \ldots y_n$ are communicated, conditioned on not aborting[2], with $\widetilde{E}$ being Eve's quantum register. Operationally $\mathsf{H}^\varepsilon_\infty(X_1 \ldots X_n Y_1 \ldots Y_n K^A|\widetilde{E})_\varphi$ is the negative logarithm of Eve's probability of guessing $x_1 \ldots x_n y_1 \ldots y_n k^A$, which is the probability of winning n instances of MSE, since Alice and Bob's winning condition is satisfied with high probability. Hence by the direct product theorem, in the presence of a bounded amount of communication, $\mathsf{H}^\varepsilon_\infty(X_1 \ldots X_n Y_1 \ldots Y_n K^A|\widetilde{E})_\varphi$ is $\Omega(n(\log 9 + \log(1/(1-\nu))))$. By the chain rule of conditional min-entropy, this means that $\mathsf{H}^\varepsilon_\infty(K^A|XY\widetilde{E})_\varphi$ is $\Omega(n \log(1/(1-\nu)))$. We remark that since our direct product theorem is not "perfect", i.e., the exponent we have is $\Omega(n)$ instead of n, we can only have Alice and Bob communicate a subset of $x_1 \ldots x_n y_1 \ldots y_n$ here instead of all of them (and XY in the notation refers to the subset), and use those for key generation, so as not to make $\mathsf{H}^\varepsilon_\infty(K^A|XY\widetilde{E})_\varphi$ negative.

In the actual state ρ after xy is released, Eve can do some local operations on $XY\widetilde{E}$, but these do not change $\mathsf{H}^\varepsilon_\infty(K^A|XY\widetilde{E})_\varphi$, and hence we have the same lower bound for $\mathsf{H}^\varepsilon_\infty(K^A|XY\widetilde{E})_\rho$. In order to upper bound $\mathsf{H}^\varepsilon_0(K^B|K^A)_\rho$, we use the operational interpretation of $H^\varepsilon_0(K^B|K^A)_\rho$ as the maximum number of possible values of K^B given K^A. As mentioned before, conditioned on not aborting, K^A and K^B differ in very few locations with high probability, and hence we can bound this quantity.

Remark 2. *An alternate security proof of the [JMS20] protocol was given in [Vid17] by using the parallel repetition of anchored games instead of product games. A version of case (i) of Theorem 1 with anchored games could also be used to follow this proof instead, to prove security against leakage.*

3 Preliminaries

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

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

[2]Alice and Bob cannot actually check the abort condition before $x_1 \ldots x_n, y_1 \ldots y_n$ are communicated, but the aborting condition is a well-defined event on $K^A K^B XY$ and thus can be conditioned on before this.

Fact 3. *For two distributions P_X and $\mathsf{P}_{X'}$ on the same set and an event $\mathcal{E}$ on the set,*

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

The following result is a consequence of the well-known Serfling bound.

Fact 4 ([TL17])**.** *Let $Z = Z_1 \dots Z_n$ be n binary random variables with an arbitrary joint distribution, and let T be a random subset of size γn for $0 \leq \gamma \leq 1$, picked uniformly among all such subsets of $[n]$ and independently of Z. Then,*

$$\Pr\left[\left(\sum_{i\in T} Z_i \geq (1-\varepsilon)\gamma n\right) \wedge \left(\sum_{i\in[n]} Z_i < (1-2\varepsilon)n\right)\right] \leq 2^{-2\varepsilon^2\gamma n}.$$

3.2 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

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

The purified distance based on fidelity is given by

$$\Delta(\rho,\sigma) = \sqrt{1 - \mathsf{F}(\rho,\sigma)^2}.$$

ℓ_1 distance and Δ are both metrics that satisfy the triangle inequality.

Fact 5 (Uhlmann's theorem)**.** *Suppose ρ and σ are states on register X which are purified to $|\rho\rangle_{XY}$ and $|\sigma\rangle_{XY'}$ with Y ad Y' not necessarily being of the same dimension, then it holds that*

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

where the maximization is over isometries taking Y' to Y.

Fact 6 (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 7 ([Tom16])**.** *The square of the fidelity is jointly concave in both arguments, i.e.,*

$$\mathsf{F}(\varepsilon\rho + (1-\varepsilon)\rho', \varepsilon\sigma + (1-\varepsilon)\sigma')^2 \geq \varepsilon\mathsf{F}(\rho,\sigma)^2 + (1-\varepsilon)\mathsf{F}(\rho',\sigma')^2.$$

Fact 8 (Data-processing inequality)**.** *For a quantum channel $\mathcal{O}$ and states ρ and σ,*

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

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

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

We shall also denote this by $\mathsf{H}(Z)_\rho$. For a state ρ_{YZ} on registers YZ, the entropy of Y conditioned on Z is given by

$$\mathsf{H}(Y|Z)_\rho = \mathsf{H}(YZ)_\rho - \mathsf{H}(Z)_\rho$$

where $\mathsf{H}(Z)_\rho$ is calculated w.r.t. the reduced state ρ_Z. The relative entropy between two states ρ and σ of the same dimensions is given by

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

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

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

It is easy to see that for all ρ and σ,

$$0 \leq \mathsf{D}(\rho\|\sigma) \leq \mathsf{D}_\infty(\rho\|\sigma).$$

Fact 9 (Pinsker's inequality)**.** *For any two states ρ and σ,*

$$\|\rho - \sigma\|_1^2 \leq 2 \ln 2 \cdot \mathsf{D}(\rho\|\sigma) \quad \textit{and} \quad 1 - \mathsf{F}(\rho, \sigma) \leq \ln 2 \cdot \mathsf{D}(\rho\|\sigma).$$

Fact 10. *For any unitary U, and states ρ, σ, $\mathsf{D}(U\rho U^\dagger \| U\sigma U^\dagger) = \mathsf{D}(\rho\|\sigma)$, and $\mathsf{D}_\infty(U\rho U^\dagger \| U\sigma U^\dagger) = \mathsf{D}_\infty(\rho\|\sigma)$.*

Fact 11. *If $\sigma = \varepsilon\rho + (1-\varepsilon)\rho'$, then $\mathsf{D}_\infty(\rho\|\sigma) \leq \log(1/\varepsilon)$.*

Fact 12. *For any three quantum states ρ, σ, φ such that* $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\varphi) \subseteq \mathrm{supp}(\sigma)$*,*

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

The conditional min-entropy of Y given Z is defined as

$$\mathsf{H}_\infty(Y|Z)_\rho = \inf\{\lambda : \exists \sigma_Z \text{ s.t. } \rho_{YZ} \leq 2^{-\lambda} \mathbb{1}_Y \otimes \sigma_Z\}.$$

The conditional Hartley entropy of Y given Z is defined as

$$\mathsf{H}_0(Y|Z)_\rho = \log\left(\sup_{\sigma_Z} \mathrm{Tr}(\mathrm{supp}(\rho_{YZ})(\mathbb{1}_Y \otimes \sigma_Z))\right)$$

where $\mathrm{supp}(\rho_{YZ})$ is the projector on to the support of ρ_{YZ}. For a classical distribution P_{YZ}, this reduces to

$$\mathsf{H}_0(Y|Z)_{\mathsf{P}_{YZ}} = \log\left(\sup_z |\{y : \mathsf{P}_{YZ}(y,z) > 0\}|\right).$$

Fact 13. *If* $\rho_{YZ} = (\mathbb{1}_Y \otimes U)\sigma_{YZ}(\mathbb{1}_Y \otimes U)$, *then*

$$\mathsf{H}_\infty(Y|Z)_\sigma = \mathsf{H}_\infty(Y|Z)_\rho \quad \text{and} \quad \mathsf{H}_0(Y|Z)_\sigma = \mathsf{H}_0(Y|Z)_\rho.$$

For any distance measure (not necessarily a metric) d between states, the ε-smoothed relative min-entropy between ρ and σ w.r.t. d is defined as

$$\mathsf{D}_\infty^{\varepsilon,d}(\rho\|\sigma) = \inf_{\rho' : d(\rho,\rho') \leq \varepsilon} \mathsf{D}_\infty(\rho'\|\sigma).$$

When d is the ℓ_1 distance, we often omit the superscript.

Fact 14 (Quantum Substate Theorem, [JRS02, JRS09, JN12])**.** *For any two states ρ and σ such that the support of ρ is contained in the support of σ, and any $\varepsilon > 0$,*[3]

$$\mathsf{D}_\infty^{\varepsilon,\mathsf{F}}(\rho\|\sigma) \leq \frac{\mathsf{D}(\rho\|\sigma)+1}{\varepsilon} + \log\left(\frac{1}{1-\varepsilon}\right).$$

Consequently,

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

Fact 15 ([JRS02])**.** *For two states ρ_X and σ_X, if $\mathsf{D}_\infty^{\varepsilon,\Delta}(\rho_X\|\sigma_X) = c$, then for any purifications $|\rho\rangle_{XY}$ and $|\sigma\rangle_{XY'}$, there exists a measurement operator M taking Y' to Y, such that $\mathbb{1} \otimes M$ succeeds on $|\sigma\rangle_{XY'}$ with probability 2^{-c}, and*

$$\Delta\left(2^c(\mathbb{1}\otimes M)\,|\sigma\rangle\langle\sigma|_{XY'}\,(\mathbb{1}\otimes M^\dagger), |\rho\rangle\langle\rho|_{XY}\right) \leq \varepsilon.$$

Fact 16. *For any quantum state ρ_{YZ},*

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

The ε-smoothed versions of the conditional entropies are defined as

$$\mathsf{H}_\infty^\varepsilon(Y|Z)_\rho = \sup_{\rho' : \|\rho-\rho'\|_1 \leq \varepsilon} \mathsf{H}_\infty(Y|Z)_{\rho'} \quad \text{and} \quad \mathsf{H}_0^\varepsilon(Y|Z)_{\mathsf{P}_{YZ}} = \inf_{\rho' : \|\rho'-\rho\|_1 \leq \varepsilon} \mathsf{H}_0(Y|Z)_{\rho'}.$$

Fact 17. *For any state ρ_{XYZ},*

$$\mathsf{H}_\infty^\varepsilon(Y|Z)_\rho \geq \mathsf{H}_\infty^\varepsilon(Y|XZ)_\rho \geq \mathsf{H}_\infty^\varepsilon(YX|Z)_\rho - \log|\mathcal{X}|.$$

The mutual information between Y and Z with respect to a state ρ on YZ can be defined in the following equivalent ways:

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

The conditional mutual information between Y and Z conditioned on X is defined as

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

Mutual information can be seen to satisfy the chain rule

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

[3] Since $1-\mathsf{F}$ is the distance measure rather than F itself, the closeness condition for $\mathsf{D}_\infty^{\varepsilon,\mathsf{F}}(\rho\|\sigma)$ is $\mathsf{F}(\rho,\rho') \geq 1-\varepsilon$.

Fact 18 (Quantum Gibbs' inequality, see e.g. - [BVY17]). *For any three states $\rho_{XY}, \sigma_X, \varphi_Y$,*

$$\mathsf{D}(\rho_{XY} \| \sigma_X \otimes \varphi_Y) \geq \mathsf{D}(\rho_{XY} \| \sigma_X \otimes \rho_Y) \geq \mathsf{I}(X : Y)_\rho.$$

A state of the form

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

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 $\rho_{XY|x} = \left|x\right\rangle\!\left\langle x\right|_X \otimes \rho_{Y|x}$. When the registers are clear from context we shall often write simply ρ_x.

For CQ states where X is the classical register, relative entropy has the chain rule

$$\mathsf{D}(\rho_{XY} \| \sigma_{XY}) = \mathsf{D}(\rho_X \| \sigma_X) + \mathop{\mathbb{E}}_{\rho_X} \mathsf{D}(\rho_{Y|x} \| \sigma_{Y|x}).$$

Using this, the following fact follows by expanding out the relative entropies.

Fact 19. *For CQ states ρ_{XY} and σ_{XY},*

$$\mathop{\mathbb{E}}_{\rho_X} \mathsf{D}(\rho_{Y|x} \| \sigma_Y) - \mathsf{D}(\rho_Y \| \sigma_Y) = \mathop{\mathbb{E}}_{\rho_X} \mathsf{D}(\rho_{Y|x} \| \rho_Y) \geq 0.$$

Fact 20 ([KRS09]). *For a CQ state ρ_{XY} where X is the classical register, $\mathsf{H}_\infty(X|Y)_\rho$ is equal to the negative logarithm of the maximum probability of guessing X from the quantum system $\rho_{Y|x}$, i.e.,*

$$\mathsf{H}_\infty(X|Y)_\rho = -\log \left(\sup_{\{M_x\}_x} \sum_x \mathsf{P}_X(x) \mathrm{Tr}(M_x \rho_{Y|x}) \right)$$

where the maximization is over the set of POVMs with elements indexed by x.

3.3 Quantum communication & non-local games

An interactive entanglement-assisted quantum communication protocol $\mathcal{P}$ between l parties goes as follows: before the start of the protocol, the l parties share a joint entangled state, and at the start parties 1 through l receive inputs $x^1, \ldots, x^l$ respectively from $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$. We assume that only the j-th party communicates in rounds $\{j, j+l, j+2l, \ldots\}$, and sends messages to all the other parties. For $i \in \{j, j+l, \ldots, \}$, in the i-th round the j-th party has a memory register E_{i-l} from the previous round in which they communicated (when $i = j$, this is just the j-th party's part of the initial shared entangled state), as well as message registers $M^j_{i-l+1}, \ldots, M^j_{i-1}$ that they have received from all the other parties in the $(i-l+1)$-th to $(i-1)$-th rounds. The j-th party applies a unitary depending on their input x^j on all these registers, to generate a register E_i that they keep as memory, and a message $M_i = M^1_i \ldots M^{j-1}_i M^{j+1}_i \ldots M^l_i$, where $M^{j'}_i$ is sent to the j'-th party in this round. After all the communication rounds are done, the j-th party applies a final unitary on the memory and message registers they currently have, and then measures in the computational basis to produce their answer $a^j \in \mathcal{A}^j$. We shall denote the outputs of $\mathcal{P}$ on inputs $x^1 \ldots x^l$ to $\mathcal{P}$ by $\mathcal{P}(x^1 \ldots x^l)$ — this is a random variable, as $\mathcal{P}$'s outputs are not necessarily deterministic.

The following lemma about the final state of a quantum communication protocol is proved in Appendix A.

Lemma 5. *Let $|\sigma\rangle_{A^1\ldots A^l|x^1\ldots x^l}$ be the pure state shared by the l parties at the end of a quantum communication protocol, on inputs $x^1,\ldots x^l$, with party j holding register A^j. For any product input distribution $\mathsf{P}_{X^1\ldots X^l}$ on $\mathcal{X}^1\times\ldots\times\mathcal{X}^l$, define*

$$|\sigma\rangle_{X^1\widetilde{X}^1\ldots X^l\widetilde{X}^l A^1\ldots A^l} = \sum_{x_1\ldots x_l}\sqrt{\mathsf{P}_{X^1\ldots X^l}(x_1\ldots x_l)}\,|x^1x^1\ldots x^lx^l\rangle_{X^1\widetilde{X}^1\ldots X^l\widetilde{X}^l}\,|\sigma\rangle_{A^1\ldots A^l|x^1\ldots x^l}\,.$$

If c^j is the total communication from the j-th party in the protocol, then there for all $j\in[l]$, there exists a state $\rho^j_{X^j\widetilde{X}^jA^j}$ such that

$$\mathsf{D}_\infty\left(\sigma_{X^jX^{-j}\widetilde{X}^{-j}A^{-j}}\middle\|\sigma_{X^j}\otimes\rho_{X^{-j}\widetilde{X}^{-j}A^{-j}}\right)\leq 2c^j$$

where X^{-j} denotes $X^1\ldots X^{j-1}X^{j+1}\ldots X^l$, and $\widetilde{X}^{-j}$ and A^{-j} are defined analogously.

Definition 1. *For a predicate V on $(\mathcal{A}^1\times\ldots\times\mathcal{A}^l)\times(\mathcal{X}^1\times\ldots\times\mathcal{X}^l)$, its entanglement-assisted l-party quantum communication complexity with error $0<\varepsilon<1$, denoted by $\mathrm{Q}_\varepsilon(\mathsf{V})$, is the minimum total communication in an interactive entanglement-assisted quantum protocol such that for all $x^1\ldots x^l\in\mathcal{X}^1\times\ldots\times\mathcal{X}^l$,*

$$\Pr\left[\mathsf{V}\left(\mathcal{P}(x^1\ldots x^l),x^1\ldots x^l\right)=1\right]\geq 1-\varepsilon.$$

Definition 2. *For a predicate V on $(\mathcal{A}^1\times\ldots\times\mathcal{A}^l)\times(\mathcal{X}^1\times\ldots\times\mathcal{X}^l)$ and a distribution p on $\mathcal{X}^1\times\ldots\times\mathcal{X}^l$, the distributional entanglement-assisted l-party quantum communication complexity of V with error $0<\varepsilon<1$ w.r.t. distribution p, denoted by $\mathrm{Q}_\varepsilon(\mathsf{V},p)$, is the minimum total communication in an interactive entanglement-assisted quantum protocol such that,*

$$\Pr\left[\mathsf{V}\left(\mathcal{P}(x^1\ldots x^l),x^1\ldots x^l\right)=1\right]\geq 1-\varepsilon$$

where the probability is taken over the distribution p for $x^1\ldots x^l$, as well as the internal randomness of $\mathcal{P}$.

Fact 21 (Yao's Lemma, [Yao77])**.** *For any $0<\varepsilon<1$, and any predicate V, $\mathrm{Q}_\varepsilon(V)=\max_p\mathrm{Q}_\varepsilon(\mathsf{V},p)$.*

An l-player non-local game G is described as $(p,\mathcal{X}^1\times\ldots\times\mathcal{X}^l,\mathcal{A}^1\times\ldots\times\mathcal{A}^l,\mathsf{V})$ where p is a distribution over the input set $\mathcal{X}^1\times\ldots\times\mathcal{X}^l$, $\mathcal{A}^1\times\ldots\times\mathcal{A}^l$ is the output set, and V is a predicate on the outputs and inputs. In an entangled strategy for a non-local game, the players are allowed to share an l-partite entangled state. Player j gets input x^j and performs a measurement depending on their input on their part of the entangled state, to give their output a^j. The value achieved by a strategy on G is the probability over p and the internal randomness of the strategy that $\mathsf{V}(a^1\ldots a^l,x^1\ldots x^l)=1$.

Definition 3. *The entangled value of a game $G=(p,\mathcal{X}^1\times\ldots\times\mathcal{X}^l,\mathcal{A}^1\times\ldots\times\mathcal{A}^l,\mathsf{V})$, denoted by $\omega^*(G)$, is the maximum value achieved by any strategy for G.*

4 Quantum partition bound

For sets $\mathcal{X}^1,\ldots,\mathcal{X}^l$ and $\mathcal{A}^1,\ldots,\mathcal{A}^l$, let $\mathcal{Q}(\mathcal{A}^1\times\ldots\times\mathcal{A}^l,\mathcal{X}^1\times\ldots\times\mathcal{X}^l)$ denote the set of conditional probability distributions $q(a^1\ldots a^l|x^1\ldots x^l)$ that can be obtained by l parties who share an l-partite entangled state, receive inputs $x^j\in\mathcal{X}^j$ respectively, and perform measurements on their parts of the entangled state to obtain outputs a^j, without communicating. That is, $\mathcal{Q}(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l, \mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$ is the following set:

$$\left\{ \left(\langle\psi| M^1_{a^1|x^1} \otimes \ldots \otimes M^l_{a^l|x^l} |\psi\rangle \right)_{a^1\ldots a^l, x^1 \ldots x^l} \middle| |\psi\rangle \text{ is a state}, \forall a^j, x^j, j, \sum_{a^j \in \mathcal{A}^j} M^j_{a^j|x^j} = \mathbb{1}, M^j_{a^j|x^j} \geq 0 \right\}.$$

We state definitions for three variants of the quantum partition bound, the first of which is non-distributional and was given by [LLR12]. The second two are distributional modifications which we shall use.

Definition 4. *For a predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$, and $0 < \varepsilon < 1$, let $\perp$ be a special symbol not in any $\mathcal{A}^j$. The quantum partition bound for V with ε error, denoted by $\mathrm{eff}^*_\varepsilon(\mathsf{V})$, is defined as the optimal value of the following optimization problem:*

$$\begin{aligned}
\min \quad & \frac{1}{\eta} \\
\text{s.t.} \quad & \sum_{a^1\ldots a^l : \mathsf{V}(a^1\ldots a^l, x^1 \ldots x^l)=1} q(a^1 \ldots a^l | x^1 \ldots x^l) \geq (1-\varepsilon)\eta \quad \forall x^1 \ldots x^l \in \mathcal{X}^1 \times \ldots \times \mathcal{X}^l \\
& \sum_{a^1 \ldots a_l \in \mathcal{A}^1 \times \ldots \times \mathcal{A}^l} q(a^1 \ldots a^l | x^1 \ldots x^l) = \eta \quad \forall\, x^1 \ldots x^l \in \mathcal{X}^1 \times \ldots \times \mathcal{X}^l \\
& q(a'^1 \ldots a'^l | x^1 \ldots x^l) \in \mathcal{Q}\left((\mathcal{A}^1 \cup \{\perp\}) \times \ldots \times (\mathcal{A}^l \cup \{\perp\}), \mathcal{X}^1 \times \ldots \times \mathcal{X}^l \right).
\end{aligned}$$

Definition 5. *For a predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$, a distribution $p(x^1 \ldots x^l)$ on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, and $0 < \varepsilon < 1$, let $\perp$ be a special symbol not in any $\mathcal{A}^j$. The quantum partition bound for V with ε error with respect to p, denoted by $\widetilde{\mathrm{eff}^*}_\varepsilon(\mathsf{V}, p)$, is defined as the optimal value of the following optimization problem:*

$$\begin{aligned}
\min \quad & \frac{1}{\eta} \\
\text{s.t.} \quad & \sum_{x^1 \ldots x^l \in \mathcal{X}^1 \times \ldots \times \mathcal{X}^l} p(x^1 \ldots x^l) \sum_{a^1\ldots a^l : \mathsf{V}(a^1\ldots a^l, x^1 \ldots x^l)=1} q(a^1 \ldots a^l | x^1 \ldots x^l) \geq (1-\varepsilon)\eta \\
& \sum_{a^1 \ldots a_l \in \mathcal{A}^1 \times \ldots \times \mathcal{A}^l} q(a^1 \ldots a^l | x^1 \ldots x^l) = \eta \quad \forall\, x^1 \ldots x^l \in \mathcal{X}^1 \times \ldots \times \mathcal{X}^l \\
& q(a'^1 \ldots a'^l | x^1 \ldots x^l) \in \mathcal{Q}\left((\mathcal{A}^1 \cup \{\perp\}) \times \ldots \times (\mathcal{A}^l \cup \{\perp\}), \mathcal{X}^1 \times \ldots \times \mathcal{X}^l \right).
\end{aligned}$$

Definition 6. *For a predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$, a distribution $p(x^1 \ldots x^l)$ on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, and $0 < \varepsilon < 1$, let $\perp$ be a special symbol not in any $\mathcal{A}^j$. The average quantum partition bound for V with ε error with respect to p, denoted by $\mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$, is defined as the optimal value of the following optimization problem:*

$$\begin{aligned}
\min \quad & \frac{1}{\eta} \\
\text{s.t.} \quad & \sum_{x^1 \ldots x^l \in \mathcal{X}^1 \times \ldots \times \mathcal{X}^l} p(x^1 \ldots x^l) \sum_{a^1\ldots a^l : \mathsf{V}(a^1\ldots a^l, x^1 \ldots x^l)=1} q(a^1 \ldots a^l | x^1 \ldots x^l) \geq (1-\varepsilon)\eta \\
& \sum_{x^1 \ldots x^l \in \mathcal{X}^1 \times \ldots \times \mathcal{X}^l} p(x^1 \ldots x^l) \sum_{a^1 \ldots a^l \in \mathcal{A}^1 \times \ldots \times \mathcal{A}^l} q(a^1 \ldots a^l | x^1 \ldots x^l) = \eta \\
& q(a'^1 \ldots a'^l | x^1 \ldots x^l) \in \mathcal{Q}\left((\mathcal{A}^1 \cup \{\perp\}) \times \ldots \times (\mathcal{A}^l \cup \{\perp\}), \mathcal{X}^1 \times \ldots \times \mathcal{X}^l \right).
\end{aligned}$$

Operationally, $\mathrm{eff}^*_\varepsilon(\mathsf{V})$, $\widetilde{\mathrm{eff}^*}_\varepsilon(\mathsf{V}, p)$ and $\mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$ are connected to zero-communication protocol (with aborts) to compute V. A zero-communication protocol is one which any player is allowed to abort (indicated by them outputting the $\perp$ symbol), but if nobody aborts they need to compute V correctly. A zero-communication protocol for V is basically a strategy for a non-local game version of V, with the output alphabet extended to $\mathcal{A}^1 \cup \{\perp\}) \times \ldots \times (\mathcal{A}^l \cup \{\perp\})$. Now we can have different conditions on the abort and success probability conditioned on not aborting for such protocols.

- Suppose the protocol is required to not abort on every input $x^1 \ldots x^l$ with the same probability η, and conditioned on not aborting, every input is required to compute V correctly with probability $(1 - \varepsilon)$. $\mathrm{eff}^*_\varepsilon(\mathsf{V})$ corresponds to the efficiency, i.e., the inverse of the maximum probability of not aborting, in such a protocol.
- Suppose the protocol is required to not abort with the same probability η on every input $x^1 \ldots x^l$, but conditioned on not aborting, the probability of computing V correctly, averaged over the inputs from p, is at least $(1 - \varepsilon)$. $\widetilde{\mathrm{eff}^*}_\varepsilon(\mathsf{V})$ is the inverse of the maximum probability of not aborting in such a protocol.
- Suppose the protocol aborts on input $x^1 \ldots x^l$ with probability $\eta_{x^1 \ldots x^l}$, and we require that the average over $x^1 \ldots x^l$ from p is η. Moreover, we require that the average probability of computing V correctly is at least $(1 - \varepsilon)\eta$, i.e., the average probability of correctness conditioned on not aborting is at least $(1 - \varepsilon)$. $\mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$ is the inverse of the maximum probability of not aborting in such a protocol.

Because the requirements from the protocols are successively relaxed, it is easy to see that for any p,

$$\mathrm{eff}^*_\varepsilon(\mathsf{V}) \geq \widetilde{\mathrm{eff}^*}_\varepsilon(\mathsf{V}, p) \geq \mathrm{eff}^*_\varepsilon(\mathsf{V}, p).$$

The following lemma shows that $\widetilde{\mathrm{eff}^*}(\mathsf{V}, p)$, and hence $\mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$ lower bounds communication. The proof of this is a slight modification the proof in [LLR12] which lower bounded $\mathrm{Q}_\varepsilon(f)$ by $\mathrm{eff}^*_\varepsilon(\mathsf{V})$. We provide the proof in Appendix B for completeness.

Lemma 6. *For any predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$, any distribution p on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$ and error ε,*

$$\mathrm{Q}_\varepsilon(\mathsf{V}, p) \geq \frac{1}{2} \log \widetilde{\mathrm{eff}^*}_\varepsilon(\mathsf{V}, p).$$

Yao's Lemma and Lemma 6 imply that for any p, $\log \widetilde{\mathrm{eff}^*}_\varepsilon(\mathsf{V}, p)$ and therefore $\log \mathrm{eff}^*_\varepsilon(\mathsf{V}, p)$ are lower bounds on $\mathrm{Q}_\varepsilon(\mathsf{V})$.

4.1 Relationship between eff^* and the generalized discrepancy method

In this section we shall prove Theorem 3, recalled below.

Theorem 3. *For a total function $f : \mathcal{X} \times \mathcal{Y} \to \{-1, +1\}$, let V_f denote the predicate on $(\{-1, +1\})^2 \times (\mathcal{X} \times \mathcal{Y})$ given by*

$$\mathsf{V}(ab, xy) = 1 \iff a \cdot b = f(x, y).$$

Then for any distribution p on $\mathcal{X} \times \mathcal{Y}$,

$$\mathsf{eff}^*_\varepsilon(\mathsf{V}_f, p) \geq (1-2\varepsilon)\gamma_2^\alpha(F, p)$$

with $\alpha = \frac{1+2\varepsilon}{1-2\varepsilon}$.

We shall not define γ_2^α and its dual norm γ_2^* for general matrices. Instead, we shall use an exact characterization of $\gamma_2^*(F)$ for a boolean f in terms of non-local games given by Tsirelson, and then use a duality relation to express γ_2^α in terms of γ_2^*.

Fact 22 ([Tsi87])**.** *For total $f : \mathcal{X} \times \mathcal{Y} \to \{-1, +1\}$, let V_f denote its corresponding predicate as given in the statement of Theorem 3, and let p be any distribution on $\mathcal{X} \times \mathcal{Y}$. Then,*

$$\omega^*(G(p, \mathsf{V}_f)) = \frac{1}{2}(1 + \gamma_2^*(F \circ p)).$$

Fact 23 (see e.g. - Theorem 64 in [LS09])**.** *For any matrix A, $\alpha \geq 1$, $\gamma_2^\alpha(A)$ and $\gamma_2^*(A)$ are related as*

$$\gamma_2^\alpha(A) = \max_M \frac{(\alpha+1)\langle A, M\rangle - (\alpha-1)\|M\|_1}{2\gamma_2^*(M)}.$$

When A is the matrix corresponding to a boolean function f, this can also be expressed as

$$\gamma_2^\alpha(F) = \max_{F', p} \frac{(\alpha+1)\langle F, F' \circ p\rangle - (\alpha-1)}{2\gamma_2^*(F' \circ p)}$$

where the maximization is taken over matrices F' with ± 1 entries, and distributions p.

Using this characterization, we give the following definition of $\gamma_2^\alpha(F, p)$.

Definition 7. *For matrix F with ± 1 entries, $\gamma_2^\alpha(F, p)$ with respect to distribution p is defined as*

$$\gamma_2^\alpha(F, p) = \max_{F'} \frac{(\alpha+1)\langle F, F' \circ p\rangle - (\alpha-1)}{2\gamma_2^*(F' \circ p)}.$$

Proof of Theorem 3. Our proof closely follows the lower bound for $\mathsf{Q}_\varepsilon(f)$ in terms of $\log \gamma_2^\alpha$ as described in Section 5.3.2 of [LS09], which is credited to Harry Buhrman.

Suppose $\mathsf{eff}^*_\varepsilon(\mathsf{V}_f, p) = \frac{1}{\eta}$ for some η. Let $\mathcal{P}$ be a zero-communication protocol for V_f with constraints as required in the definition of $\mathsf{eff}^*_\varepsilon(\mathsf{V}_f, p)$. Let η_{xy} denote the probability of the protocol aborts on input (x, y). Let $O(x, y)$ denote the average (over internal randomness) output given by $\mathcal{P}$ conditioned on not aborting on inputs x, y. Here we are calling $a \cdot b$ the output of the protocol, if Alice outputs a and Bob outputs b, which means $O(x, y)$ is some number in $[-1, 1]$. Note that $O(x, y)$ is defined conditioned on not aborting, so it is in fact normalized by the quantity η. From the definition of $\mathsf{eff}^*_\varepsilon(\mathsf{V}_f, p)$, the following condition holds

$$\sum_{x,y} p(x, y) f(x, y) O(x, y) \geq 1 - 2\varepsilon.$$

The above expression is actually the difference between the probability of computing f correctly and the probability of computing it incorrectly, which is why we get $1 - 2\varepsilon$.

Now let $f' : \mathcal{X} \times \mathcal{Y} \to \{-1, +1\}$ be an arbitrary boolean function, and define $\mathsf{V}_{f'}$ the same way as V_f. We shall give a strategy $\mathcal{S}$ for the game $G(p, \mathsf{V}_{f'})$ using the zero-communication protocol $\mathcal{P}$. $\mathcal{S}$ works as follows:

- On inputs x, y for $G(p, \mathsf{V}_{f'})$, Alice and Bob run the protocol $\mathcal{P}$ on x, y.
- If $\mathcal{P}$ gives output $\perp$ for either player, they output ± 1 uniformly at random.
- If $\mathcal{P}$ does not abort, then Alice and Bob both output according to $\mathcal{P}$.

Note that conditioned on $\mathcal{P}$ not aborting, the average output produced by Alice and Bob on inputs x, y is also $O(x, y)$. Strategy $\mathcal{S}$ thus wins with probability $\frac{1}{2}(1+\delta)$ (δ may be negative), where

$$\delta = \eta \sum_{x,y} p(x,y) f'(x,y) O(x,y).$$

$f(x,y), f'(x,y)$ are in $\{-1,+1\}$, and $O(x,y)$ is in $[-1,1]$. For three numbers $\alpha, \beta \in \{-1,+1\}$, $\theta \in [-1,1]$, the following condition is true, and can be checked by putting in the four possible values of (α, β):

$$\beta\theta \geq \alpha\beta + \alpha\theta - 1.$$

Using the above on $f(x,y), f'(x,y), O(x,y)$ we get,

$$\begin{aligned}
\sum_{x,y} p(x,y) f'(x,y) O(x,y) &\geq \sum_{x,y} p(x,y) \left(f(x,y) f'(x,y) + f(x,y) O(x,y) - 1 \right) \\
&\geq \sum_{x,y} p(x,y) f(x,y) f'(x,y) + (1-2\varepsilon) - 1 \\
&= \langle F, F' \circ p \rangle - 2\varepsilon.
\end{aligned}$$

By Fact 22 we have,

$$\gamma_2^*(F' \circ p) \geq \delta \geq \eta(\langle F, F' \circ p \rangle - 2\varepsilon)$$

which gives us

$$\frac{1}{\eta} \geq \max_F \frac{\langle F, F' \circ p \rangle - 2\varepsilon}{\gamma_2^*(F' \circ p)} = (1-2\varepsilon)\gamma_2^\alpha(F, p)$$

with $\alpha = \frac{1+2\varepsilon}{1-2\varepsilon}$. □

5 Substate Perturbation Lemma

To prove the Substate Perturbation Lemma, we use the following result due to [ABJT20]. This result is stated in terms of $\mathsf{I}_{\max}$ for general states in [ABJT20], where some of the states involved are optimized over. However, for the purposes of the proof this does not matter, so we state in the form below. Our proof of the Substate Perturbation Lemma is also heavily inspired by their proof of this result.

Fact 24 ([ABJT20], Theorem 2)**.** *Suppose there are states $\sigma_{XB}, \sigma'_{XB}$ and ψ_X satisfying $\Delta(\sigma_{XB}, \sigma'_{XB}) \leq \varepsilon$ and*

$$\sigma'_{XB} \leq 2^c(\psi_X \otimes \sigma_B).$$

Then for any $\delta > 0$, there exists a state σ''_{XB} satisfying $\Delta(\sigma_{XB}, \sigma''_{XB}) \leq 2\varepsilon + \delta$, $\sigma''_B = \sigma_B$, and

$$\sigma''_{XB} \leq 2^c \left(1 + \frac{8}{\delta^2}\right) \psi_X \otimes \sigma_B.$$

Lemma 7 (Substate Perturbation Lemma)**.** *Suppose there are three states $\sigma_{XB}, \sigma'_{XB}$ and ψ_X satisfying $\Delta(\sigma_{XB}, \sigma'_{XB}) \leq \varepsilon$,*

$$\sigma'_{XB} \leq 2^c(\psi_X \otimes \sigma_B)$$

and a state ρ_B satisfying $\Delta(\sigma_B, \rho_B) \leq \delta_1$. Then for any $\delta_0 < 0$, there exists state ρ'_{XB} satisfying $\Delta(\rho'_{XB}, \sigma_{XB}) \leq 2\varepsilon + \delta_0 + \delta_1$, and

$$\rho'_{XB} \leq 2^{c+1}\left(1 + \frac{4}{\delta_0^2}\right)\psi_X \otimes \rho_B.$$

Proof. First we use Fact 24 to get a state σ''_{XB} satisfying

$$\sigma''_{XB} \leq 2^c\left(1 + \frac{8}{\delta_0^2}\right)\psi_X \otimes \sigma_B$$

such that $\Delta(\sigma_{XB}, \sigma''_{XB}) \leq 2\varepsilon + \delta_0$ and $\sigma''_B = \sigma_B$.

Let U be the unitary such that

$$\mathsf{F}\left(\rho_B, \sigma_B\right) = \mathrm{Tr}\left(U\rho_B^{1/2}\sigma_B^{1/2}\right).$$

Define

$$\rho'_{XB} = \underbrace{(\mathbb{1} \otimes \rho_B^{1/2}U\sigma_B^{-1/2})\sigma''_{XB}(\mathbb{1} \otimes \sigma_B^{-1/2}U^\dagger\rho_B^{1/2})}_{\tilde{\varphi}_{XB}} + \underbrace{\sigma_X \otimes \rho_B^{1/2}(\mathbb{1} - U\Pi U^\dagger)\rho_B^{1/2}}_{\tilde{\psi}_{XB}}$$

where all the inverses are generalized and Π is the projector onto the support of σ_B. Note that

$$(\mathbb{1} \otimes \rho_B^{1/2}U\sigma_B^{-1/2})\sigma''_{XB}(\mathbb{1} \otimes \sigma_B^{-1/2}U^\dagger\rho_B^{1/2}) \leq 2^c\left(1 + \frac{8}{\delta_0^2}\right)\psi_X \otimes \rho_B^{1/2}U\sigma_B^{-1/2}\sigma_B\sigma_B^{-1/2}U^\dagger\rho_B^{1/2},$$

and hence

$$\begin{aligned}\rho'_{XB} &\leq 2^c\left(1 + \frac{8}{\delta_0^2}\right)\psi_X \otimes \rho_B^{1/2}U\Pi U^\dagger\rho_B^{1/2} + \psi_X \otimes \rho_B^{1/2}(\mathbb{1} - U\Pi U^\dagger)\rho_B^{1/2} \\ &\leq 2^{c+1}\left(1 + \frac{4}{\delta_0^2}\right)\psi_X \otimes \rho_B.\end{aligned}$$

Now we only have to show that $\Delta(\rho'_{XB}, \sigma_{XB}) \leq 2\varepsilon + \delta_0 + \delta_1$. In order to do this, we note that

$$\Delta(\rho'_{XB}, \sigma_{XB}) \leq \Delta\left(\rho'_{XB}, \sigma''_{XB}\right) + \Delta(\sigma''_{XB}, \sigma_{XB}). \tag{1}$$

Using Fact 7,

$$\begin{aligned}\mathsf{F}\left(\rho'_{XB}, \sigma''_{XB}\right)^2 &\geq \mathrm{Tr}(\tilde{\varphi}_{XB}) \cdot \mathsf{F}\left(\frac{\tilde{\varphi}_{XB}}{\mathrm{Tr}(\tilde{\varphi}_{XB})}, \sigma''_{XB}\right)^2 + \mathrm{Tr}(\tilde{\psi}_{XB}) \cdot \mathsf{F}\left(\frac{\tilde{\psi}_{XB}}{\mathrm{Tr}(\tilde{\psi}_{XB})}, \sigma''_{XB}\right)^2 \\ &\geq \mathrm{Tr}(\tilde{\varphi}_{XB}) \cdot \mathsf{F}\left(\frac{\tilde{\varphi}_{XB}}{\mathrm{Tr}(\tilde{\varphi}_{XB})}, \sigma''_{XB}\right)^2 \\ &\geq \mathrm{Tr}(\tilde{\varphi}_{XB}) \cdot \mathsf{F}\left(|\varphi\rangle\langle\varphi|_{XBC}, |\sigma''\rangle\langle\sigma''|_{XBC}\right)^2\end{aligned} \tag{2}$$

where in the last step $|\varphi\rangle_{XBC}$ and $|\sigma''\rangle_{XBC}$ are arbitrary purifications of $\varphi_{XB} = \tilde{\varphi}_{XB}/\mathrm{Tr}(\tilde{\varphi}_{XB})$ and σ''_{XB}, and we have used Fact 8 with the tracing out operation. Note that φ_{BC} is obtained from σ''_{BC}

by doing an operation $\mathcal{O}_B$ only on B, which is akin to applying a measurement and conditioning on success. In particular this operation preserves purity of states. We let $|\varphi\rangle_{XBC}$ be the state we get by applying $\mathcal{O}_B$ on $|\sigma''\rangle_{XBC}$. Now let $|\sigma_1''\rangle_{B\widetilde{B}}$ be the canonical purification of σ_B'' and $|\varphi_1\rangle_{B\widetilde{B}}$ be the state we get by applying $\mathcal{O}_B$ on $|\sigma_1''\rangle_{B\widetilde{B}}$. These are given by

$$|\sigma_1''\rangle_{B\widetilde{B}} = ((\sigma_B'')^{1/2} \otimes \mathbb{1}) \sum_i |i\rangle_B |i\rangle_{\widetilde{B}} = (\sigma_B^{1/2} \otimes \mathbb{1}) \sum_i |i\rangle_B |i\rangle_{\widetilde{B}}$$

$$|\varphi_1\rangle_{B\widetilde{B}} = \frac{\rho_B^{1/2} U \sigma_B^{-1/2} \otimes \mathbb{1}}{\mathrm{Tr}(\tilde{\varphi}_{XB})^{1/2}} |\sigma''\rangle_{B\widetilde{B}} = \frac{\rho_B^{1/2} U \Pi \otimes \mathbb{1}}{\mathrm{Tr}(\tilde{\varphi}_{XB})^{1/2}} \sum_i |i\rangle_B |i\rangle_{\widetilde{B}}.$$

Since $|\sigma''\rangle_{XBC}$ is also a purification of σ_B'', there exists an isometry V acting only on $\widetilde{B}$ such that $\mathbb{1}_B \otimes V |\sigma''\rangle_{XBC} = |\sigma_1''\rangle_{B\widetilde{B}}$. Hence,

$$\begin{aligned}
\mathsf{F}\left(|\varphi\rangle\langle\varphi|_{XBC}, |\sigma''\rangle\langle\sigma''|_{XBC}\right) &= \mathsf{F}\left(\mathcal{O}_B(|\sigma''\rangle\langle\sigma''|_{XBC}), |\sigma''\rangle\langle\sigma''|_{XBC}\right) \\
&= \mathsf{F}\left(\mathbb{1}_B \otimes V \left(\mathcal{O}_B(|\sigma''\rangle\langle\sigma''|_{XBC})\right) \mathbb{1}_B \otimes V^\dagger, \mathbb{1}_B \otimes V |\sigma''\rangle\langle\sigma''|_{XBC} \mathbb{1}_B \otimes V^\dagger\right) \\
&= \mathsf{F}\left(\mathcal{O}_B\left(\mathbb{1}_B \otimes V |\sigma''\rangle\langle\sigma''|_{XBC}) \mathbb{1}_B \otimes V^\dagger\right), \mathbb{1}_B \otimes V |\sigma''\rangle\langle\sigma''|_{XBC} \mathbb{1}_B \otimes V^\dagger\right) \\
&= \mathsf{F}\left(|\sigma_1''\rangle\langle\sigma_1''|_{B\widetilde{B}}, |\varphi_1\rangle\langle\varphi_1|_{B\widetilde{B}}\right).
\end{aligned}$$

Putting this in (2) gives us

$$\begin{aligned}
\mathsf{F}\left(\rho'_{XB}, \sigma''_{XB}\right)^2 &\geq \left| \sum_i \sum_j \left(\langle ii| \left(\Pi U \rho_B^{1/2} \otimes \mathbb{1}\right)\right) \left(\left(\sigma_B^{1/2} \otimes \mathbb{1}\right) |jj\rangle\right) \right|^2 \\
&= \left| \sum_i \langle i| \Pi U \rho_B^{1/2} \sigma_B^{1/2} |i\rangle \right|^2 \\
&= \left| \mathrm{Tr}(\Pi U \rho_B^{1/2} \sigma_B^{1/2}) \right|^2 \\
&= \left| \mathrm{Tr}(U \rho_B^{1/2} \sigma_B^{1/2}) \right|^2 = \mathsf{F}\left(\rho_B, \sigma_B\right)^2
\end{aligned}$$

where we have used the fact that $\sigma_B^{1/2}\Pi = \sigma_B^{1/2}$, and the definition of U. Putting this in (1) we get,

$$\Delta(\rho'_{XB}, \sigma_{XB}) \leq \Delta\left(\rho_B, \sigma_B\right) + \Delta(\sigma''_{XB}, \sigma_{XB}) \leq \delta_1 + 2\varepsilon + \delta_0. \qquad \square$$

6 Proof of the direct product theorem

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

Theorem 1. *For any $\varepsilon, \zeta > 0$, any predicate V on $(\mathcal{A}^1 \times \ldots \times \mathcal{A}^l) \times (\mathcal{X}^1 \times \ldots \times \mathcal{X}^l)$ and any product probability distribution p on $\mathcal{X}^1 \times \ldots \times \mathcal{X}^l$, if $\mathcal{P}$ is an interactive entanglement-assisted quantum communication protocol between l parties which has total communication cn.*

(i) If $c < 1$, then

$$\mathrm{suc}(p^n, \mathsf{V}^n, \mathcal{P}) \leq \left(1 - \frac{\nu}{2} + 4\sqrt{lc}\right)^{\Omega\left(\nu^2 n / \left(l^2 \cdot \log(|\mathcal{A}^1| \cdot \ldots \cdot |\mathcal{A}^l|)\right)\right)}$$

where $\nu = 1 - \omega^*(G(p, \mathsf{V}))$.

(ii) If $1 \leq c = O\left(\frac{\zeta^2}{l^3}\mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V},p)\right)$*, then*

$$\mathrm{suc}(p^n, \mathsf{V}^n, \mathcal{P}) \leq (1-\varepsilon)^{\Omega\left(n/\left(\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)\right)\right)}.$$

6.1 Setup

We consider an interactive quantum protocol $\mathcal{P}$ for n copies of V with player j having input registers $X^j = X^j_1 \ldots X^j_n$, and communicating $c^j n$ bits. The total communication of the protocol is cn, where $c = \sum_{j=1}^{l} c^j$. In the case $c \geq 1$, we shall also assume each $c^j \geq 1$; if some c^j is smaller than 1, we can pad extra bits to it, and this increases total communication by a factor of at most l. Hence we have, $\sum_{j=1}^{l} c^j \leq cl$.

We define the following pure state

$$|\psi\rangle_{X^1\widetilde{X}^1\ldots X^l\widetilde{X}^l E^1\ldots E^l A^1\ldots A^l} = \sum_{xy}\sqrt{\mathsf{P}_{X^1\ldots X^l}(x^1\ldots x^l)}\,|x^1x^1\ldots x^lx^l\rangle_{X^1\widetilde{X}^1\ldots X^l\widetilde{X}^l}\,|\psi\rangle_{E^1\ldots E^l A^1\ldots A^l|x^1\ldots x^l}$$

where $\mathsf{P}_{X^1\ldots X^l}$ is the distribution p^n on $(\mathcal{X}^1\times\ldots\times\mathcal{X}^l)^n$, and $|\psi\rangle_{E^1\ldots E^l A^1\ldots A^l|x^1\ldots x^l}$ being the state at the end of the protocol on inputs $x^1,\ldots,x^l$. In $|\psi\rangle_{E^1\ldots E^l A^1\ldots A^l|x^1\ldots x^l}$, $A^j = A^j_1\ldots A^j_n$ are the output registers of player j, and E^j is some quantum register they have that they don't measure. We use $\mathsf{P}_{X^1\ldots X^l A^1\ldots A^l}$ to denote the distribution of $X^1\ldots X^l A^1\ldots A^l$ in $|\psi\rangle$. We shall use X to denote $X^1\ldots X^l$, X_i to denote $X^1_i\ldots X^l_i$, X^{-j} to denote $X^1\ldots X^{j-1}X^{j+1}\ldots X^l$, and $X^{\leq j}$ to denote $X^1\ldots X^j$. Similar notation will be used for $\widetilde{X}^j, E^j, A^j$. Also for a subset $C\subseteq[n]$, we shall use use X_C to denote $(X_i)_{i\in C}$.

We shall show the following lemma, which can be applied inductively to get Theorem 1.

Lemma 8. *For* $i\in[k]$*, let* $T_i = \mathsf{V}(A^1_i\ldots A^l_i, X^1_i\ldots X^l_i)$ *in* $\mathcal{P}$*, and let* $\mathcal{E}$ *denote the event* $\prod_{i\in C} T_i = 1$ *for some* $C\subseteq[n]$ *such that* $|C|\leq n/2$*,*

(i) If $c<1$*,*

$$\mathop{\mathbb{E}}_{i\in\bar{C}}\Pr[T_i=1|\mathcal{E}] \leq \omega^*(G(p,\mathsf{V})) + 4\sqrt{lc} + \frac{7l+1}{2}\sqrt{2\delta},$$

(ii) If $1\leq c < \frac{\zeta^2}{270 l^3}\mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V},p)$*, and if* $\delta<1$*, there exists an* $i\in\bar{C}$ *such that*

$$\Pr[T_i=1|\mathcal{E}]\leq 1-\varepsilon,$$

where

$$\delta = \frac{|C|\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)+\log(1/\Pr[\mathcal{E}])}{n}.$$

In order to get the statement of case (i) of Theorem 1 from case (i) of Lemma 8, we start with $C=\emptyset$, and find some $i\in[n]$ such that $\Pr[T_i=1]\leq 1-\nu+4\sqrt{lc}+\frac{\nu}{2}$. As long as $\frac{7l+1}{2}\sqrt{2\delta}$ is at most $\frac{\nu}{2}$ we can do this. When we have built up a non-empty set C this way, if either $|C| = \Omega\left(\frac{\nu^2 n}{l^2\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)}\right)$, or $\Pr[\Pi_{i\in C}T_i=1]\leq\exp\left(-\Omega\left(\frac{\nu^2 n}{l^2\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)}\right)\right)$, we are already done. Otherwise, $\frac{7l+1}{2}\sqrt{2\delta}<\frac{\nu}{2}$, and we can continue the process.

The bound on $\Pr[T_i = 1|\mathcal{E}]$ in case (ii) of Lemma 8 does not depend on δ, but it requires $\delta < 1$ as a precondition. Hence following the same process there, we can go up to C of size $|C| = \Theta\left(\frac{n}{\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)}\right)$, or $\Pr[\Pi_{i\in C} T_i = 1] = \exp\left(-\frac{n}{\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)}\right)$.

Since in case (i) Lemma 8 gives us a bound on $\mathbb{E}_{i\in\bar{C}} \Pr[T_i = 1|\mathcal{E}]$ rather than showing just that there exists an i for which $\Pr[T_i = 1|\mathcal{E}]$ is bounded, we can use it to show the following corollary, which we shall later use in our DIQKD application. See Appendix C of [JMS20] for a proof of how this follows from the lemma.

Corollary 9. *Let $\mathsf{V}^{t/n}_{\text{rand}}$ be the randomized predicate which is satisfied if V is satisfied on a random subset of size t of $[n]$. If the communication cost of $\mathcal{P}$ is $cn < n$, then*[4]

$$\mathrm{suc}(p^n, \mathsf{V}^{t/n}_{\text{rand}}, \mathcal{P}) \leq \left(\omega^*(G(p,\mathsf{V})) + O\left(\sqrt{lc} + l\sqrt{\frac{t\cdot\log(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|)}{n}}\right)\right)^t.$$

6.2 Proof of Lemma 8

We define the following state which is $|\psi\rangle$ conditioned on success event $\mathcal{E}$ in C:

$$|\varphi\rangle_{X\widetilde{X}EA} = \frac{1}{\sqrt{\gamma}}\sum_{x_C x_{\bar{C}}}\sqrt{\mathsf{P}_X(x_C x_{\bar{C}})}\,|x_C x_{\bar{C}} x_C x_{\bar{C}}\rangle_{X\widetilde{X}} \otimes \sum_{a_C: V^{|C|}(a_C, x_C)=1} |a_C\rangle_{A_C}\,|\tilde{\varphi}\rangle_{EA_{\bar{C}}|x_C x_{\bar{C}} a_C}$$

where $|\tilde{\varphi}\rangle_{EA_{\bar{C}}|x_C x_{\bar{C}} a_C}$ is a subnormalized state satisfying $\|\,|\tilde{\varphi}\rangle_{EA_{\bar{C}}|x_C x_{\bar{C}} a_C}\|_2^2 = \mathsf{P}_{A_C|x_C x_{\bar{C}}}(a_C)$, and $\gamma = \Pr[\mathcal{E}]$.

We shall use the following lemma, whose proof we give later.

Lemma 10. *Letting $R = X_C A_C$, the following conditions hold:*

1. $\mathbb{E}_{i\in\bar{C}}\left\|\mathsf{P}_{X_i R|\mathcal{E}} - \mathsf{P}_{X_i}\mathsf{P}_{R|\mathcal{E}}\right\|_1 \leq \sqrt{2\delta}$.
2. *In case (i): $c < 1$, for every $i \in \bar{C}$ and $j \in [l]$, there exist unitaries $\left\{U^j_{i,x^j_i r}\right\}_{i,x^j_i r}$ acting only on the registers $X^j_{\bar{C}}\widetilde{X}^j_{\bar{C}}E^j A^j_{\bar{C}}$ such that*
$$\mathop{\mathbb{E}}_{i\in\bar{C}}\mathop{\mathbb{E}}_{\mathsf{P}_{X_i R|\mathcal{E}}}\left\|\Big(\bigotimes_{j\in[l]} U^j_{i,x^j_i r}\Big)|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|r}\Big(\bigotimes_{j\in[l]}(U^j_{i,x^j_i r})^\dagger\Big) - |\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i r}\right\|_1 \leq 8\sqrt{lc} + 7l\sqrt{2\delta}.$$
3. *In case (ii): $1 \leq c < \frac{\zeta^2}{270 l^3}\mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V}, p)$ and $\delta < 1$, there exists an $i \in \bar{C}$ such that for every $j \in [l]$, there exist measurement operators M^j_i taking registers $X^j_i\widetilde{X}^j_{\bar{C}}E^j A^j_{\bar{C}}$ to $\widetilde{X}^j_{\bar{C}}E^j A^j_{\bar{C}}$ (with $M^j_i(M^j_i)^\dagger$ being the POVM element), such that each $\bigotimes_{j\in[l]} M^j_i$ succeeds on $|\psi\rangle_{X'_i X_i}\otimes|\varphi\rangle_{\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}$ with probability $\alpha_i \geq 2^{-\frac{270 l^3 c}{\zeta^2}}$, and*
$$\left\|\frac{1}{\alpha_i}\Big(\bigotimes_{j\in[l]} M^j_i\Big)\Big(|\psi\rangle\langle\psi|_{X'_i X_i}\otimes|\varphi\rangle\langle\varphi|_{\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\Big)\Big(\bigotimes_{j\in[l]}(M^j_i)^\dagger\Big) - |\varphi\rangle\langle\varphi|_{X'_i\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\right\|_1 \leq 2\zeta$$
where $|\psi\rangle_{X'_i X_i} = \sum_{x_i}\sqrt{\mathsf{P}_{X_i}(x_i)}\,|x_i x_i\rangle_{X'_i X_i}$, $|\varphi\rangle_{X'_i\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}$ is the same state as $|\varphi\rangle_{X'_i\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}$ with the X_i register replaced by the X'_i register.

[4] Note that $\mathrm{suc}(p^n, \mathsf{V}^{t/n}_{\text{rand}}, \mathcal{P})$ accounts for the randomness inherent in $\mathsf{V}^{t/n}_{\text{rand}}$ in addition to p^n and the protocol.

6.2.1 Case (i): $c < 1$

Using conditions 1 and 2 of Lemma 10, we can give a quantum strategy $\mathcal{S}$ for the non-local game $G(p, \mathsf{V})$ whose winning probability is at least

$$\mathop{\mathbb{E}}_{i \in \bar{C}} \Pr[T_i = 1|\mathcal{E}] - 4\sqrt{lc} - \frac{7l+1}{2}\sqrt{2\delta}.$$

By the definition of $\omega^*(G(p, \mathsf{V}))$, $\mathcal{S}$ cannot have success probability more than $\omega^*(G(p, \mathsf{V}))$. This gives the required upper bound on $\mathbb{E}_{i \in \bar{C}} \Pr[T_i = 1|\mathcal{E}]$.

On input $x_i^1 \ldots x_i^l$, $\mathcal{P}'$ works as follows:

- The l players share $\log(|\bar{C}|)$ uniformly random bits and r according to the distribution $\mathsf{P}_{R|\mathcal{E}}$.
- For every r, the players also share $|\varphi\rangle_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|r}$ as entanglement, with player j holding registers $X^j_{\bar{C}}\widetilde{X}^j_{\bar{C}}E^j A^j_{\bar{C}}$.
- The players jointly select a uniform $i \in \bar{C}$ and r from $\mathsf{P}_{R|\mathcal{E}}$.
- Player j applies the $U^j_{i,x^j_i r}$ unitary according to their input x^j_i and the shared randomness, on their part of the shared entangled state $|\varphi\rangle_{\widetilde{X}_{\bar{C}}EA_{\bar{C}}|r}$. Then they measure the A^j_i register of the resulting state to give their output.

Due to 2, the players produce an output distribution $(8\sqrt{lc} + 7l\sqrt{2\delta})/2$-close to that of $|\varphi\rangle_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i r}$, when averaged over i and (x_i, r) from $\mathsf{P}_{X_i R|\mathcal{E}}$. $|\varphi\rangle_{x_i r}$ gives the correct answer with probability $\Pr[T_i = 1|\mathcal{E}]$ over $\mathsf{P}_{X_i R|\mathcal{E}}$. Hence $\mathcal{S}$ gives the correct answer with probability at least

$$\mathop{\mathbb{E}}_{i \in \bar{C}} \Pr[T_i = 1|\mathcal{E}] - 4\sqrt{lc} - \frac{7l+1}{2}\sqrt{2\delta}$$

when averaged over i and (x_i, r) from $\mathsf{P}_{X_i}\mathsf{P}_{R|\mathcal{E}}$.

6.2.2 Case (ii): $c \geq 1$

Using condition 3 of Lemma 10, we can give a zero-communication protocol $\mathcal{P}'$ for V whose average probability of not aborting is at least $2^{-\frac{270l^3 c}{\zeta^2}} > 1/\mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V}, p)$ (by the condition on cn), and conditioned on not aborting, is correct with probability at least

$$\Pr[T_i = 1|\mathcal{E}] - \zeta$$

(with the i provided by this condition) averaged on inputs from p. By the definition of $\mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V}, p)$, $\mathcal{P}'$ cannot be correct conditioned on not aborting with probability more than $1 - (\varepsilon + \zeta)$ when inputs come from p. This gives the required upper bound on $\Pr[T_i = 1|\mathcal{E}]$.

For this case, it will be helpful to think of the joint state of the inputs and entangled state in a zero-communication protocol quantumly. If the player receive inputs from a distribution $\mathsf{P}_Y = \mathsf{P}_{Y^1 \ldots Y^l}$, we can think of them as receiving registers $Y^1, \ldots, Y^l$ respectively of a pure state

$$|\sigma\rangle_{Y'Y} = \sum_y \sqrt{\mathsf{P}_Y(y)}\,|yy\rangle_{Y'Y}$$

with say a referee holding the Y' registers. The players hold a shared entangled state $|\rho\rangle_{EA} = |\rho\rangle_{E^1\ldots E'^l A^1 \ldots A^l}$, with player j holding $E^j A^j$, A^l being the answer register. Player j now applies some measurement on registers $Y^j E^j A^j$ to determine their output. Strictly speaking, this measurement should only use Y^j as a control register, since it is classical. But player j can always copy over Y^j to a different register $\widetilde{Y}^j$ and apply a general measurement on $\widetilde{Y}^j E^j A^j$ — the effect of this will be the same as applying a general measurement on $Y^j E^j A^j$ that does not use Y^j as a control register. So we shall assume that player j can in fact apply a general measurement on $Y^j E^j A^j$.

We shall also assume that in the protocol, the players first apply a measurement to decide whether they will abort or not abort, and conditioned on not aborting, do another measurement to give outputs in $\mathcal{A}^1 \times \ldots \times \mathcal{A}^l$ (in general they can do a single measurement to decide their output, which may be abort, or some element of $\mathcal{A}^j$, but the protocol $\mathcal{P}'$ we describe will have two measurements). In fact they do not need to actually do this last measurement in order for us to determine the average success probability: we can assume that the state conditioned on not aborting already has the correlations they want between the registers Y' and A (the Y^j registers may have been modified by the measurement), and the average success probability is determined by computing V on $Y'A$ of the state conditioned on not aborting. That is, suppose the measurement operator corresponding to not abort for player j is M^j. Then the average probability of not aborting in the protocol is the success probability α of $\bigotimes_{j\in[l]} M^j$ on $|\sigma\rangle_{Y'Y} \otimes |\rho\rangle_{EA}$. And the average success probability of the protocol conditioned on not aborting is determined by computing V on the $Y'A$ registers of $\frac{1}{\sqrt{\alpha}}\left(\bigotimes_{j\in[l]} M^j\right) |\sigma\rangle_{Y'Y} \otimes |\rho\rangle_{EA}$.

Now we shall describe the actual protocol $\mathcal{P}'$. In $\mathcal{P}'$:

- The players share $|\varphi\rangle_{\widetilde{X}_{\bar{C}} E A_{\bar{C}} R}$ as shared entanglement, with player j holding the registers $\widetilde{X}^j_{\bar{C}} E^j A^j_{\bar{C}}$ (the extra R register can go to any player, say the first, but they won't need to do anything on it).
- The players receive inputs as the X^j_i register of $|\psi\rangle_{X'_i X_i}$ (note that the distribution in this state is the correct one, p).
- Player j applies measurements $\{M^j_i (M^j_i)^\dagger, \mathbb{1} - M^j_i (M^j_i)^\dagger\}$ on the registers $X^j_i \widetilde{X}^j_{\bar{C}} E^j A^j_{\bar{C}}$ and declares not abort if the Π^j_i measurement succeeds.
- Conditioned on not aborting, player j provides A^j_i as their answer register.

By our description above, and condition 3, the average probability of not aborting in this protocol is $\alpha_i \geq 2^{-\frac{270 l^3 c}{\zeta^2}} > \frac{1}{\mathrm{eff}^*_{\varepsilon+\zeta}(\mathsf{V},p)}$ by the condition on c. Now note that if V is computed in the $X'_i A_i$ register of $|\varphi\rangle_{X'_i \widetilde{X}_{\bar{C}} E A_{\bar{C}}}$, the average success probability is by definition $\Pr[T_i = 1|\mathcal{E}]$. Since by condition 3,

$$\left\| \frac{1}{\alpha_i}\Big(\bigotimes_{j\in[l]} M^j_i\Big) \Big(|\psi\rangle\langle\psi|_{X'_i X_i} \otimes |\varphi\rangle\langle\varphi|_{\widetilde{X}_{\bar{C}} E A_{\bar{C}} R}\Big) \Big(\bigotimes_{j\in[l]} (M^j_i)^\dagger\Big) - |\varphi\rangle\langle\varphi|_{X'_i \widetilde{X}_{\bar{C}} E A_{\bar{C}} R} \right\|_1 \leq 2\zeta$$

the average success probability on $\frac{1}{\sqrt{\alpha_i}}\Big(\bigotimes_{j\in[l]} M^j_i\Big) |\psi\rangle_{X'_i X_i} \otimes |\varphi\rangle_{\widetilde{X}_{\bar{C}} E A_{\bar{C}} R}$, that is, the average success probability of $\mathcal{P}'$ conditioned on not aborting, is at least $\Pr[T_i = 1|\mathcal{E}] - \zeta$.

6.3 Proof of Lemma 10

The first part of the proof goes the same way for both cases (i) and (ii). We shall proceed with a common proof and then diverge when required.

Since player j's communication in $\mathcal{P}$ is $c^j n$ bits, by Lemma 5 for the final state $|\psi\rangle$ of $\mathcal{P}$, there exists a state $\rho^j_{X^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}$ such that

$$\mathsf{D}_\infty\left(\psi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\middle\|\psi_{X^j}\otimes\rho^j_{X^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\right)\leq 2c^jn.$$

Using Facts 11 and 12, this gives us

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathsf{P}_{R|\mathcal{E}}}\mathsf{D}\left(\varphi_{X^j_{\bar{C}}X^{-j}_{\bar{C}}\tilde{X}^{-j}_{\bar{C}}E^{-j}A^{-j}_{\bar{C}}|r}\middle\|\psi_{X^j_{\bar{C}}}\otimes\rho^j_{X^{-j}_{\bar{C}}\tilde{X}^{-j}_{\bar{C}}E^{-j}A^{-j}_{\bar{C}}}\right)\\
&=\mathop{\mathbb{E}}_{\mathsf{P}_{X_CA_C|\mathcal{E}}}\mathsf{D}\left(\varphi_{X^j_{\bar{C}}X^{-j}_{\bar{C}}\tilde{X}^{-j}_{\bar{C}}E^{-j}A^{-j}_{\bar{C}}|x_Ca_C}\middle\|\psi_{X^j_{\bar{C}}}\otimes\rho^j_{X^{-j}_{\bar{C}}\tilde{X}^{-j}_{\bar{C}}E^{-j}A^{-j}_{\bar{C}}}\right)\\
&\leq\mathop{\mathbb{E}}_{\mathsf{P}_{A_C|\mathcal{E}}}\mathsf{D}\left(\varphi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}|a_C}\middle\|\psi_{X^j}\otimes\rho^j_{X^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\right)\\
&\leq\mathop{\mathbb{E}}_{\mathsf{P}_{A_C|\mathcal{E}}}\mathsf{D}_\infty\left(\varphi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}|a_C}\middle\|\psi_{X^j}\otimes\rho^j_{X^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\right)\\
&\leq\mathop{\mathbb{E}}_{\mathsf{P}_{X_CA_C|\mathcal{E}}}\left[\mathsf{D}_\infty\left(\varphi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}|a_C}\middle\|\varphi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\right)\right.\\
&\qquad+\mathsf{D}_\infty\left(\varphi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\middle\|\psi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\right)\\
&\qquad\left.+\mathsf{D}_\infty\left(\psi_{X^jX^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\middle\|\psi_{X^j}\otimes\rho^j_{X^{-j}\tilde{X}^{-j}E^{-j}A^{-j}}\right)\right]\\
&\leq\mathop{\mathbb{E}}_{\mathsf{P}_{X_CA_C|\mathcal{E}}}\left[\log(1/\mathsf{P}_{A_C|\mathcal{E}}(a_C))+\log(1/\Pr[\mathcal{E}])+2c^jn\right]\\
&\leq\mathop{\mathbb{E}}_{\mathsf{P}_{X_CA_C|\mathcal{E}}}\left[|C|\cdot\log\left(|\mathcal{A}^1|\cdot\ldots\cdot|\mathcal{A}^l|\right)+\log(1/\Pr[\mathcal{E}])+2c^jn\right]\\
&=(\delta+2c^j)n. \qquad (3)
\end{aligned}$$

Similarly we also have,

$$\mathsf{D}\left(\varphi_{X_{\bar{C}}R}\middle\|\psi_{X_{\bar{C}}}\otimes\varphi_R\right)=\mathop{\mathbb{E}}_{\mathsf{P}_{R|\mathcal{E}}}\mathsf{D}\left(\varphi_{X_{\bar{C}}|r}\middle\|\psi_{X_{\bar{C}}}\right)\leq\delta n. \qquad (4)$$

Now using Pinsker's inequality on this, and Jensen's inequality along with the convexity of the square function,

$$\mathop{\mathbb{E}}_{i\in\bar{C}}\left\|\mathsf{P}_{X_iR|\mathcal{E}}-\mathsf{P}_{X_i}\mathsf{P}_{R|\mathcal{E}}\right\|_1\leq\sqrt{\mathop{\mathbb{E}}_{i\in\bar{C}}\left\|\mathsf{P}_{X_iR}-\mathsf{P}_{X_i}\mathsf{P}_{R|\mathcal{E}}\right\|_1^2}\leq\sqrt{\frac{1}{n-|C|}n\cdot\delta}\leq\sqrt{2\delta}. \qquad (5)$$

This already shows item 1 of the lemma. For further calculations, we shall also upper bound for any $j\in[l]$

$$\mathop{\mathbb{E}}_{i\in\bar{C}}\left\|\mathsf{P}_{X_i^{\leq j}R|\mathcal{E}}-\mathsf{P}_{X_i^jR|\mathcal{E}}\mathsf{P}_{X_i^{<j}|\mathcal{E},R}\right\|_1$$

$$
\begin{aligned}
&\leq \mathop{\mathbb{E}}_{i \in \bar{C}} \left(\left\| \mathsf{P}_{X_i^{\leq j} R|\mathcal{E}} - \mathsf{P}_{X_i^{\leq j}} \mathsf{P}_{R|\mathcal{E}} \right\|_1 + \left\| \mathsf{P}_{X_i^{\leq j}} \mathsf{P}_{R|\mathcal{E}} - \mathsf{P}_{X_i^j} \mathsf{P}_{X_i^{<j} R|\mathcal{E}} \right\|_1 + \left\| \mathsf{P}_{X_i^j} \mathsf{P}_{X_i^{<j} R|\mathcal{E}} - \mathsf{P}_{X_i^j R|\mathcal{E}} \mathsf{P}_{X_i^{<j}|\mathcal{E},R} \right\|_1 \right) \\
&= \mathop{\mathbb{E}}_{i \in \bar{C}} \left(\left\| \mathsf{P}_{X_i^{\leq j} R|\mathcal{E}} - \mathsf{P}_{X_i^{\leq j}} \mathsf{P}_{R|\mathcal{E}} \right\|_1 + \left\| \mathsf{P}_{X_i^j} \left(\mathsf{P}_{X_i^{<j}} \mathsf{P}_{R|\mathcal{E}} - \mathsf{P}_{X_i^{<j} R|\mathcal{E}} \right) \right\|_1 + \left\| \left(\mathsf{P}_{X_i^j} \mathsf{P}_{R|\mathcal{E}} - \mathsf{P}_{X_i^j R|\mathcal{E}} \right) \mathsf{P}_{X_i^{<j}|\mathcal{E},R} \right\|_1 \right) \\
&= \mathop{\mathbb{E}}_{i \in \bar{C}} \left(\left\| \mathsf{P}_{X_i^{\leq j} R|\mathcal{E}} - \mathsf{P}_{X_i^{\leq j}} \mathsf{P}_{R|\mathcal{E}} \right\|_1 + \left\| \mathsf{P}_{X_i^{<j}} \mathsf{P}_{R|\mathcal{E}} - \mathsf{P}_{X_i^{<j} R|\mathcal{E}} \right\|_1 + \left\| \mathsf{P}_{X_i^j} \mathsf{P}_{R|\mathcal{E}} - \mathsf{P}_{X_i^j R|\mathcal{E}} \right\|_1 \right) \\
&\leq 3\sqrt{2\delta} \qquad (6)
\end{aligned}
$$

where in the last step we have used (5), tracing out the $X_i^{>j}$ $X_i^{\geq j}$ and X_i^{-j} registers respectively in the three terms.

6.3.1 Case (i): $c < 1$

This case follows the proof in [JPY14] closely, so we shall only give a brief sketch. Let $\bar{C}_{<i}$ denote the set of coordinates in $\bar{C}$ which are less than i. By Quantum Gibbs' inequality on (3) and chain rule of relative entropy, we get for all $j \in [l]$,

$$
\begin{aligned}
2c^j + \delta &\geq \mathop{\mathbb{E}}_{\mathsf{P}_{R|\mathcal{E}}} \mathsf{D}\left(\varphi_{X_{\bar{C}}^j X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \| \psi_{X_{\bar{C}}^j} \otimes \varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \right) \\
&= \sum_{i \in \bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{R|\mathcal{E}}} \mathop{\mathbb{E}}_{\mathsf{P}_{X^j_{\bar{C}_{<i}}|r}} \mathsf{D}\left(\varphi_{X_i^j X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|x^j_{\bar{C}_{<i}} r} \| \psi_{X_i^j} \otimes \varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \right) \\
&\overset{(a)}{\geq} \sum_{i \in \bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{R|\mathcal{E}}} \mathsf{D}\left(\varphi_{X_i^j X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \| \psi_{X_i^j} \otimes \varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \right) \\
&\geq \sum_{i \in \bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{R|\mathcal{E}}} \mathsf{D}\left(\varphi_{X_i^j X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \| \varphi_{X_i^j} \otimes \varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \right)
\end{aligned}
$$

where in (a) we have used Fact 19. Using Pinsker's inequality on this, it follows for any $j \in [l]$ that

$$
1 - \mathop{\mathbb{E}}_{i \in \bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_i R|\mathcal{E}}} \mathsf{F}\left(\varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|x_i^j r}, \varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \right) \leq 4c^j + 2\delta.
$$

Let $U^j_{i,x_i^j r}$ be the unitary from Uhlmann's theorem such that

$$
\begin{aligned}
&\mathsf{F}\left(|\varphi\rangle\langle\varphi|_{X_{\bar{C}} \widetilde{X}_{\bar{C}} E A_{\bar{C}}|x_i^j r}, \left(U^j_{i,x_i^j r} \otimes \mathbb{1} \right) |\varphi\rangle\langle\varphi|_{X_{\bar{C}} \widetilde{X}_{\bar{C}} E A_{\bar{C}}|r} \left((U^j_{i,x_i^j r})^\dagger \otimes \mathbb{1} \right) \right) \\
&= \mathsf{F}\left(\varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|x_i^j r}, \varphi_{X_{\bar{C}}^{-j} \widetilde{X}_{\bar{C}}^{-j} E^{-j} A_{\bar{C}}^{-j}|r} \right).
\end{aligned}
$$

Using the Fuchs-van de Graaf inequality and Jensen's inequality for the square root function, then we have,

$$
\begin{aligned}
\mathop{\mathbb{E}}_{i \in \bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X_i R|\mathcal{E}}} \left\| \left(U^j_{i,x_i^j r} \otimes \mathbb{1} \right) |\varphi\rangle\langle\varphi|_{X_{\bar{C}} \widetilde{X}_{\bar{C}} E A_{\bar{C}}|r} \left((U^j_{i,x_i^j r})^\dagger \otimes \mathbb{1} \right) - |\varphi\rangle\langle\varphi|_{X_{\bar{C}} \widetilde{X}_{\bar{C}} E A_{\bar{C}}|x_i^j r} \right\|_1 &\leq 4\sqrt{4c^j + 2\delta} \\
&\leq 8\sqrt{c^j} + 4\sqrt{2\delta}. \qquad (7)
\end{aligned}
$$

Defining $\mathcal{O}_{X_i^{<k}}$ as the quantum channel that measures the $X_i^{<k}$ registers and records the outcome, this gives us

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in\bar{C}}\mathop{\mathbb{E}}_{\mathsf{P}_{X_iR|\mathcal{E}}}\left\|\bigotimes_{j\in[l]}U^j_{i,x_i^jr}\,|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|r}\bigotimes_{j\in[l]}(U^j_{i,x_i^jr})^\dagger-|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_ir}\right\|_1\\
&\le\sum_{k=1}^{l}\mathop{\mathbb{E}}_{i\in\bar{C}}\mathop{\mathbb{E}}_{\mathsf{P}_{X_iR|\mathcal{E}}}\left\|\bigotimes_{j>k}U^j_{i,x_i^jr}\left(U^k_{i,x_i^kr}\,|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i^{<k}r}\,(U^k_{i,x_i^kr})^\dagger-|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i^{\le k}r}\right)\bigotimes_{j>k}U^j_{i,x_i^jr}\right\|_1\\
&\le\sum_{k=1}^{l}\mathop{\mathbb{E}}_{i\in\bar{C}}\mathop{\mathbb{E}}_{\mathsf{P}_{X_i^{\le k}R|\mathcal{E}}}\left\|U^k_{i,x_i^kr}\,|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i^{<k}r}\,(U^k_{i,x_i^kr})^\dagger-|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i^{\le k}r}\right\|_1\\
&\le\sum_{k=1}^{l}\mathop{\mathbb{E}}_{i\in\bar{C}}\mathop{\mathbb{E}}_{\mathsf{P}_{X_iR|\mathcal{E}}}\left(\left\|\mathcal{O}_{X_i^{<k}}\left(U^k_{i,x_i^kr}\,|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|r}\,(U^k_{i,x_i^kr})^\dagger-|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i^kr}\right)\right\|_1\right.\\
&\qquad\left.+\left\|\mathsf{P}_{X_i^{<k}|\mathcal{E},x_i^kr}-\mathsf{P}_{X_i^{<k}|\mathcal{E},r}\right\|_1\right)\\
&\le\sum_{k=1}^{l}\mathop{\mathbb{E}}_{i\in\bar{C}}\left(\mathop{\mathbb{E}}_{\mathsf{P}_{X_iR|\mathcal{E}}}\left\|U^k_{i,x_i^kr}\,|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|r}\,(U^k_{i,x_i^kr})^\dagger-|\varphi\rangle\langle\varphi|_{X_{\bar{C}}\widetilde{X}_{\bar{C}}EA_{\bar{C}}|x_i^kr}\right\|_1\right.\\
&\qquad\left.+\left\|\mathsf{P}_{X_i^{\le k}R|\mathcal{E}}-\mathsf{P}_{X_i^kR|\mathcal{E}}\mathsf{P}_{X_i^{<k}|\mathcal{E},R}\right\|_1\right)\\
&\overset{(b)}{\le}\sum_{k=1}^{l}(8\sqrt{c^k}+4\sqrt{2\delta}+3\sqrt{2\delta})\\
&\overset{(c)}{\le}8\sqrt{l\sum_{k=1}^{l}c^k}+7l\sqrt{2\delta}=8\sqrt{lc}+7l\sqrt{2\delta}
\end{aligned}$$

where for (b) we have used (7) and (6), and for (c) we have used the Cauchy-Schwarz inequality. This proves condition 2 of Lemma 10.

6.3.2 Case (ii): $c \ge 1$

Using the Quantum Gibb's inequality on (3) we have,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathsf{P}_{X_{\bar{C}}^jR|\mathcal{E}}}\mathsf{D}\left(\varphi_{X_{\bar{C}}^{-j}\widetilde{X}_{\bar{C}}^{-j}E^{-j}A_{\bar{C}}^{-j}|x_{\bar{C}}^jr}\middle\|\varphi_{X_{\bar{C}}^{-j}\widetilde{X}_{\bar{C}}^{-j}E^{-j}A_{\bar{C}}^{-j}|r}\right)&\le\mathop{\mathbb{E}}_{\mathsf{P}_{R|\mathcal{E}}}\mathsf{D}\left(\varphi_{X_{\bar{C}}^jX_{\bar{C}}^{-j}\widetilde{X}_{\bar{C}}^{-j}E^{-j}A_{\bar{C}}^{-j}|r}\middle\|\psi_{X_{\bar{C}}^j}\otimes\varphi_{X_{\bar{C}}^{-j}\widetilde{X}_{\bar{C}}^{-j}E^{-j}A_{\bar{C}}^{-j}|r}\right)\\
&\le(2c^j+\delta)n.
\end{aligned}$$

From (4) we have,

$$\mathsf{D}\left(\varphi_{X_{\bar{C}}^jR}\middle\|\psi_{X_{\bar{C}}^j}\otimes\varphi_R\right)\le\delta n.$$

Hence by the chain rule of relative entropy,

$$\mathsf{D}\left(\varphi_{X_{\bar{C}}^jX_{\bar{C}}^{-j}\widetilde{X}_{\bar{C}}^{-j}E^{-j}A_{\bar{C}}^{-j}R}\middle\|\psi_{X_{\bar{C}}^j}\otimes\varphi_{X_{\bar{C}}^{-j}\widetilde{X}_{\bar{C}}^{-j}E^{-j}A_{\bar{C}}^{-j}R}\right)\le 2(c^j+\delta)n.$$

By the chain rule of relative entropy again,

$$
\begin{aligned}
4(c^j+\delta) &\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \mathop{\mathbb{E}}_{\mathsf{P}_{X^j_{\bar{C}_{<i}}}} \mathsf{D}\left(\varphi_{X^j_i X^{-j}_{\bar{C}} \widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R | x_{\bar{C}_{<i}}} \middle\| \psi_{X^j_i} \otimes \varphi_{X^{-j}_{\bar{C}} \widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R}\right) \\
&\geq \mathop{\mathbb{E}}_{i\in\bar{C}} \mathsf{D}\left(\varphi_{X^j_i X^{-j}_{\bar{C}} \widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R} \middle\| \psi_{X^j_i} \otimes \varphi_{X^{-j}_{\bar{C}} \widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R}\right)
\end{aligned}
$$

where we have used Fact 19. Using the Quantum Substate Theorem on the above and tracing out $X^{-j}_{\bar{C}}$ we get for all $j \in [l]$,

$$
\mathop{\mathbb{E}}_{i\in\bar{C}} \mathsf{D}_\infty^{\sqrt{2\zeta'},\Delta}\left(\varphi_{X^j_i \widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R} \middle\| \psi_{X^j_i} \otimes \varphi_{\widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R}\right) \leq \frac{4c^j+4\delta+1}{\zeta'} + \log\left(\frac{1}{1-\zeta'}\right)
$$

for some ζ' to be fixed later. Now since X'^j_i as used in $|\psi\rangle_{X'^j_i X^j_i}$ and $|\varphi\rangle_{X'^j_i \widetilde{X}_{\bar{C}} E A_{\bar{C}} R}$ in the statement of item 3 in Lemma 10, is identical to X^j_i, we also have,

$$
\mathop{\mathbb{E}}_{i\in\bar{C}} \mathsf{D}_\infty^{\sqrt{2\zeta'},\Delta}\left(\varphi_{X'^j_i \widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R} \middle\| \psi_{X'^j_i} \otimes \varphi_{\widetilde{X}^{-j}_{\bar{C}} E^{-j} A^{-j}_{\bar{C}} R}\right) \leq \frac{4c^j+4\delta+1}{\zeta'} + \log\left(\frac{1}{1-\zeta'}\right) \tag{8}
$$

To find the measurement operators M^j_i, we shall do induction on the number of players. In particular we shall prove the following lemma.

Lemma 11. *Suppose we have measurement operators $\left\{M^j_i\right\}_i$ for $j \in [k], i \in \bar{C}, 0 \leq k < l$, taking registers $X^j_i \widetilde{X}^j_{\bar{C}} E^j A^j_{\bar{C}}$ to $\widetilde{X}^j_{\bar{C}} E^j A^j_{\bar{C}}$ respectively, such that $\bigotimes_{j\in[k]} M^j_{i,r}$ succeeds on $\left(\bigotimes_{j\in[k]} |\psi\rangle_{X'^j_i X^j_i}\right) \otimes |\varphi\rangle_{\widetilde{X}_{\bar{C}} E A_{\bar{C}} R}$ with probability $\alpha^{\leq k}_i = 2^{-\sum_{j=1}^k \tilde{c}^j_i}$ where*

$$
\mathop{\mathbb{E}}_{i\in\bar{C}} \tilde{c}^j_i \leq \frac{15c^j}{\zeta'},
$$

and for all $i \in \bar{C}$,

$$
\begin{aligned}
\Delta\Bigg(&\frac{1}{\alpha^{\leq k}_i}\left(\bigotimes_{j\in[k]} M^j_i \otimes \mathbb{1}\right)\left(\bigotimes_{j\in[k]} |\psi\rangle\langle\psi|_{X'^j_i X^j_i} \otimes |\varphi\rangle\langle\varphi|_{\widetilde{X}_{\bar{C}} E A_{\bar{C}} R}\right)\left(\bigotimes_{j\in[k]} (M^j_i)^\dagger \otimes \mathbb{1}\right), \\
&|\psi\rangle\langle\psi|_{X'^{>k}_i X^{>k}_i} \otimes |\varphi\rangle\langle\varphi|_{X'^{\leq k}_i \widetilde{X}_{\bar{C}} E A_{\bar{C}} R}\Bigg) \leq (3k-2)\sqrt{2\zeta'}.
\end{aligned} \tag{9}
$$

Then there are measurement operators $\left\{M^{k+1}_i\right\}_i$ taking registers $X^{k+1}_i \widetilde{X}^{k+1}_{\bar{C}} E^{k+1} A^{k+1}_{\bar{C}}$ to $\widetilde{X}^{k+1}_{\bar{C}} E^{k+1} A^{k+1}_{\bar{C}}$, such that $\bigotimes_{j\in[k+1]} M^j_i$ succeeds on $\left(\bigotimes_{j\in[k+1]} |\psi\rangle_{X'^j_i X^j_i}\right) \otimes |\varphi\rangle_{\widetilde{X}_{\bar{C}} E A_{\bar{C}} R}$ with probability $\alpha^{\leq(k+1)}_i = \alpha^{k+1}_i \alpha^{\leq k}_i$ where $\alpha^{k+1}_i = 2^{-\tilde{c}^{k+1}_i}$, with

$$
\mathop{\mathbb{E}}_{i\in\bar{C}} \tilde{c}^{k+1}_i \leq \frac{15c^{k+1}}{\zeta'},
$$

and for all $i \in \bar{C}$

$$
\Delta\left(\frac{1}{\alpha^{\leq(k+1)}_i}\left(\bigotimes_{j\in[k+1]} M^j_i \otimes \mathbb{1}\right)\left(\bigotimes_{j\in[k+1]} |\psi\rangle\langle\psi|_{X'^j_i X^j_i} \otimes |\varphi\rangle\langle\varphi|_{\widetilde{X}_{\bar{C}} E A_{\bar{C}} R}\right)\left(\bigotimes_{j\in[k+1]} (M^j_i)^\dagger \otimes \mathbb{1}\right),\right.
$$

$$|\psi\rangle\langle\psi|_{X_i'^{>(k+1)}X_i^{>(k+1)}} \otimes |\varphi\rangle\langle\varphi|_{X_i'^{\leq(k+1)}\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\Bigg) \leq (3k+1)\sqrt{2\zeta'}.$$

We also clarify that if the distance in (9) is $\Delta[k]$, then the way we pick our parameters in the proof of the lemma gives us $\Delta[k+1] = \Delta[k] + 3\sqrt{2\zeta'}$. The expression $(3k-2)\sqrt{2\zeta'}$ is obtained by setting $\Delta[1] = \sqrt{2\zeta'}$.

Proof of Lemma 11. Let

$$|\rho\rangle_{X_i'X_i^{>k}\widetilde{X}_{\bar{C}}EA_{\bar{C}}R} = \frac{1}{\sqrt{\alpha_i^{\leq k}}}\left(\bigotimes_{j\in[k]} M_i^j \otimes \mathbb{1}\right)\left(\bigotimes_{j\in[k]} |\psi\rangle_{X_i^jX_i'^j} \otimes |\varphi\rangle_{\widetilde{X}_{\bar{C}}EA_{\bar{C}}|r}\right).$$

Note that $|\rho\rangle$ has an i dependence, but we are not writing it explicitly. By (9),

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \Delta\left(\rho_{\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R}, \varphi_{\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R}\right) \leq \Delta[k].$$

Moreover, since none of the operators M_i^j for $j \in [k]$ act on the X_i^{k+1} register,

$$\rho_{X_i'^{k+1}\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R} = \rho_{X_i'^{k+1}} \otimes \rho_{\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R} = \psi_{X_i'^{k+1}} \otimes \rho_{\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R}.$$

Using the Substate Perturbation Lemma on the above and (8) with $j = k+1$, picking parameters $\varepsilon = \delta_0 = \sqrt{2\zeta'}, \delta_1 = \Delta[k]$ we get,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in\bar{C}} \mathsf{D}_\infty^{\Delta[k+1],\Delta}\left(\varphi_{X_i'^{k+1}\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R}\middle\|\rho_{X_i'^{k+1}\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R}\right)\\
&= \mathop{\mathbb{E}}_{i\in\bar{C}} \mathsf{D}_\infty^{3\sqrt{2\zeta'}+\Delta[k],\Delta}\left(\varphi_{X_i'^{k+1}\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R}\middle\|\psi_{X_i'^{k+1}} \otimes \rho_{X_i'^{k+1}\widetilde{X}_{\bar{C}}^{-(k+1)}E^{-(k+1)}A_{\bar{C}}^{-(k+1)}R}\right)\\
&\leq \frac{4c^{k+1}+4\delta+1}{\zeta'} + \log\left(\frac{1}{1-\zeta'}\right) + 1 + \log\left(1+\frac{2}{\zeta'}\right)\\
&\leq \frac{4c^{k+1}+4\delta+1}{\zeta'} + 3\zeta' + 1 + \frac{2}{\zeta'} \leq \frac{15c^{k+1}}{\cdot\zeta'}.
\end{aligned}$$

Now note that $|\psi\rangle_{X_i'^{>(k+1)}X_i^{>(k+1)}} \otimes |\varphi\rangle_{X_i'^{\leq(k+1)}\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}$ is a purification of the state in the first argument in the above smoothed entropy, and $|\rho\rangle_{X_i'X_i\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}$ is obviously a purification of the state in the second. Therefore, by Fact 15, there exist measurement operators $\left\{M_i^{k+1}\right\}_i$ taking registers $X_i^{k+1}\widetilde{X}_{\bar{C}}^{k+1}E^{k+1}A_{\bar{C}}^{k+1}$ to $\widetilde{X}_{\bar{C}}^{k+1}E^{k+1}A_{\bar{C}}^{k+1}$, that succeed on $|\rho\rangle_{X_i'X_i^{>k}\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}$ with probability $\alpha_i^{k+1} = 2^{-\tilde{c}_i^{k+1}}$, where

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \tilde{c}_i^{k+1} \leq \frac{15c^{k+1}}{\zeta'},$$

and for all i,

$$\Delta\Bigg(\frac{1}{\alpha_i^{k+1}\alpha_i^{\leq k}}\left(\bigotimes_{j\in[k+1]} M_i^j \otimes \mathbb{1}\right)\left(\bigotimes_{j\in[k+1]} |\psi\rangle\langle\psi|_{X_i'^jX_i^j} \otimes |\varphi\rangle\langle\varphi|_{\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\right)\left(\bigotimes_{j\in[k+1]} (M_i^j)^\dagger \otimes \mathbb{1}\right),$$

$$|\psi\rangle\langle\psi|_{X_i'^{>(k+1)}X_i^{>(k+1)}} \otimes |\varphi\rangle\langle\varphi|_{X_i'^{\leq(k+1)}\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\Bigg)$$

$$= \Delta\left(\frac{1}{\alpha_i^{k+1}} M_i^{k+1} \otimes \mathbb{1}\left(|\rho\rangle\langle\rho|_{X_i'X_i^{>k}\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\right)(M_i^{k+1})^\dagger \otimes \mathbb{1}, |\psi\rangle\langle\psi|_{X_i'^{>(k+1)}X_i^{>(k+1)}} \otimes |\varphi\rangle\langle\varphi|_{X_i'^{\leq(k+1)}\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\right)$$

$$\leq \Delta[k+1].$$

This proves the lemma. □

After the induction process, we have measurement operators $\left\{M_i^j\right\}_i$ for $j \in [l]$ and the conditions in the statement of Lemma 11 hold with $k = l$. Therefore, by the Fuchs-van de Graaf inequality,

$$\left\|\frac{1}{\alpha_i}\Big(\bigotimes_{j\in[l]} M_i^j\Big)\left(|\psi\rangle\langle\psi|_{X_i'X_i} \otimes |\varphi\rangle\langle\varphi|_{\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\right)\Big(\bigotimes_{j\in[l]}(M_i^j)^\dagger\Big) - |\varphi\rangle\langle\varphi|_{X_i'\widetilde{X}_{\bar{C}}EA_{\bar{C}}R}\right\|_1$$
$$\leq 2(3l-2)\sqrt{2\zeta'}.$$

Setting $(3l-2)\sqrt{2\zeta'} = \zeta$ we get, $\zeta' \geq \frac{\zeta^2}{18l^2}$. This gives us

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \sum_{j=1}^{l} \tilde{c}_i^j \leq \frac{270l^2}{\zeta^2}\sum_{j=1}^{l} c^j \leq \frac{270l^3c}{\zeta^2}.$$

Since 2^{-x} is a convex function, by Jensen's inequality we have,

$$\mathop{\mathbb{E}}_{i\in\bar{C}} \alpha_i = \mathop{\mathbb{E}}_{i\in\bar{C}} 2^{-\sum_{j=1}^{l}\tilde{c}_i^j} \geq 2^{-\mathbb{E}_{i\in\bar{C}}\sum_{j=1}^{l}\tilde{c}_i^j} \geq 2^{-270l^3c/\zeta^2}.$$

Therefore there exists an $i \in \bar{C}$ such that $\alpha_i \geq 2^{-270l^3c/\zeta^2}$. This proves condition 3 in Lemma 10.

7 DIQKD with leakage

In this section, we prove Theorem 4, whose statement is recalled below.

Theorem 4. *There are universal constants $0 < \delta_0 < 1$ and $0 < c_0 < 1$ such that for any $0 \leq \delta \leq \delta_0$, and $0 \leq c \leq c_0$, if the [JMS20] DIQKD protocol (given in Protocol 1) is carried out with boxes that play n copies of the Magic Square game δ-noisily, it is possible to extract $r(\delta, c)n$ bits of secret key in the interactive leakage model, with the total communication between Alice, Bob and Eve's boxes being cn bits, for some $r(\delta, c) > 0$.*

Protocol 1 is given below. It makes use of the following equipment:

(i) Boxes $(\mathcal{B}^A, \mathcal{B}^B)$ with Alice and Bob respectively, whose honest behaviour is to play n i.i.d. instances of MS δ-noisily, i.e., each copy of MS is won with probability $1-\delta$;

(ii) Private sources of randomness for both Alice and Bob;

(iii) A public authenticated channel between Alice and Bob.

Protocol 1 DIQKD protocol (with parameters α, γ, δ)

Alice chooses $x_1 \dots x_n \in \{0,1,2\}^n$ uniformly at random from private randomness, inputs it into her box $\mathcal{B}^{\mathrm{A}}$, and records the output $a_1 \dots a_n$
Bob chooses $y_1 \dots y_n \in \{0,1,2\}^n$ uniformly at random from private randomness, inputs it into his box $\mathcal{B}^{\mathrm{B}}$, and records the output $b_1 \dots b_n$
Alice chooses $S \subseteq [n]$ of size αn, $T \subseteq S$ of size $\gamma|S|$ uniformly at random from private randomness
Alice sends (S, T, x_S, a_T) to Bob using the public channel
Bob sends y_S to Alice using the public channel
Bob tests if $a_i[y_i] = b_i[x_i]$ for at least $(1-2\delta)|T|$ many i-s in T
if the test fails **then**
    Bob aborts the protocol
else
    Alice sets $(K^{\mathrm{A}})_{i \in S} = a_i[y_i]$ and Bob sets $(K^{\mathrm{B}})_{i \in S} = b_i[x_i]$ as their respective keys

We shall prove the following theorem about Protocol 1, which implies Theorem 4.

Theorem 12. *Let $\rho_{K^{\mathrm{A}}K^{\mathrm{B}}X_SY_SA_TST\widetilde{E}}$ be the state of Alice's and Bob's raw keys and Eve's side information conditioned on not aborting in Protocol 1 carried out with parameters α, γ, δ (where $\widetilde{E}$ is Eve's quantum register and $X_SY_SA_TST$ is the communication through the public channel which she also has access to). If the total communication in the interactive leakage model is cn for some $c < 1$, then the state ρ satisfies*

$$\mathsf{H}^{\varepsilon}_{\infty}(K^{\mathrm{A}}|X_SY_SA_TST\widetilde{E})_\rho - \mathsf{H}^{\varepsilon}_0(K^{\mathrm{A}}|K^{\mathrm{B}})_\rho \geq \alpha\left(\nu - \beta(\sqrt{c} + \sqrt{\alpha}) - 2h_2(4\delta) - \gamma\right)n - \log(1/\Pr[\mathcal{E}]),$$

where $\mathcal{E}$ is the event that the protocol does not abort, $\varepsilon' = \frac{2 \cdot 2^{-8\delta^2 \alpha n}}{\Pr[\mathcal{E}]}$, β, ν are constants in $(0,1)$, and h_2 is the binary entropy function. Moreover, when $(\mathcal{B}^{\mathrm{A}}, \mathcal{B}^{\mathrm{B}})$ have their honest δ-noisy behaviour, then $\Pr[\mathcal{E}] \geq 1 - 2^{-2\delta^2\gamma\alpha n}$.

We are free to pick the parameters α, γ in Protocol 1; δ is also a parameter in the Protocol, which is picked according to the noise level expected in honest boxes. The constant ν is the one provided by Fact 25. For c, δ such that $\nu > \beta\sqrt{c} + 2h_2(4\delta)$, there exist choices of α, γ and values of $\Pr[\mathcal{E}]$ for which the above quantity is positive. Hence we get a positive key rate for c, δ in this region.

7.1 Properties of the Magic Square game

Definition 8. *The 2-player Magic Square game, denoted by* MS, *is as follows:*

- *Alice and Bob receive respective inputs $x \in \{0,1,2\}$ and $y \in \{0,1,2\}$ independently and uniformly at random.*
- *Alice outputs $a \in \{0,1\}^3$ such that $a[0] \oplus a[1] \oplus a[2] = 0$ and Bob outputs $b \in \{0,1\}^3$ such that $b[0] \oplus b[1] \oplus b[2] = 1$.*
- *Alice and Bob win the game iff $a[y] = b[x]$.*

The classical value of the magic square game is $\omega(\mathrm{MS}) = 8/9$, whereas the quantum value is $\omega^*(\mathrm{MS}) = 1$.

Definition 9. *The 3-player variant of the Magic Square game, denoted by* MSE, *is as follows:*

- *Alice receives inputs $x \in \{0,1,2\}, z \in \{0,1\}$ and Bob receives input $y \in \{0,1,2\}$ independently and uniformly at random; Eve receives no input.*
- *Alice outputs $a \in \{0,1\}^3$ such that $a[0] \oplus a[1] \oplus a[2] = 0$, Bob outputs $b \in \{0,1\}^3$ such that $b[0] \oplus b[1] \oplus b[2] = 1$, and Eve outputs $x' \in \{0,1,2\}, y' \in \{0,1,2\}, z' \in \{0,1\}$ and $c \in \{0,1\}$.*
- *Alice, Bob and Eve win the game iff*

$$(x = x') \wedge (y = y') \wedge (a[y] = c) \wedge ((a[y] = b[x]) \vee (z = z')).$$

Fact 25 ([JMS20])**.** *There is a constant $0 < \nu < 1$ such that $\omega^*(\mathrm{MSE}) = \frac{1}{9}(1-\nu)$.*

The above fact is a consequence of Proposition 4.1 in [JMS20]. The game considered in the statement of this proposition in [JMS20] is different: they consider a 6-player game between Alice, Bob, Alice$'$, Bob$'$, Charlie and Charlie$'$. Here we have given Charlie's role to Alice, and merged Alice$'$, Bob$'$ and Charlie$'$ into Eve (this is later done in the analysis in [JMS20] anyway). Doing this makes no difference in the proof of the game's winning probability as given in [JMS20].[5] Alternatively, the fact can be seen as a consequence of Lemma 2 in [Vid17]. The game considered in [Vid17] does not include Eve having to produce guesses x', y', z' for x, y, z. Suppose the probability of winning Vidick's game is $(1-\nu')$. Since by no-signalling Eve's best probability of guessing z is $\frac{1}{2}$, the probability of winning the version of the game where Eve has to produce z' but not x', y' is $(1-\frac{\nu'}{2})$. Further, since Eve's probability of guessing x and y are both $\frac{1}{3}$ the probability of winning MSE where she has to produce x', y' is $\frac{1}{9}(1-\frac{\nu'}{2})$.

Now Corollary 9 has the following consequence for the parallel-repeated MSE game in the interactive leakage model.

Corollary 13. *There exists a constant $\beta > 0$ such that if the total communication in the interactive leakage model is at most cn for some $c < 1$, with ν being the constant from Fact 25, then the probability of winning* MSE *in a random subset of size t out of n instances is at most*

$$\left(\frac{1-\nu+\beta(\sqrt{c}+\sqrt{t/n})}{9}\right)^t.$$

7.2 Security proof with leakage

We introduce some notation for states. Note that we have defined $\mathcal{E}$ to be the abort event, but we can equivalently define it to be the event that $a_i[y_i] = b_i[x_i]$ for at least $(1-2\delta)|T|$ many i-s in T. This way we can condition states of the protocol before Alice and Bob have communicated on $\mathcal{E}$ as well, even though they cannot abort at this point. For the variable K^{V} that is defined in Lemma 14, we use:

$\rho_{K^{\mathrm{A}}K^{\mathrm{B}}X_SY_SA_TST\widetilde{E}}$	:	state conditioned on $\mathcal{E}$ at the end of Protocol 1
$\sigma_{K^{\mathrm{A}}K^{\mathrm{B}}K^{\mathrm{V}}X_SY_SA_TST\widetilde{E}}$	:	state after step 3 in Protocol 1
$\varphi_{K^{\mathrm{A}}K^{\mathrm{B}}K^{\mathrm{V}}X_SY_SA_TST\widetilde{E}}$	:	state after step 3 in Protocol 1 conditioned on $\mathcal{E}$.

[5] In [JMS20], z' is the input of Charlie$'$ rather than an output. For the application perspective, we think it makes more sense to make it an output, since we consider the probability of Eve guessing z. However, due to no-signalling the probability that $z = z'$ is $\frac{1}{2}$ regardless, and this change makes no difference.

First we shall prove some lemmas about the states σ and φ, and then use them to get the final min-entropy bound on ρ.

Lemma 14. *Define the variable*

$$(K^{\mathrm{V}})_{i\in S} = \begin{cases} 0 & \text{if } a_i[y_i] = b_i[x_i] \\ 0/1 \text{ w.p. } \frac{1}{2} & \text{otherwise.} \end{cases}$$

If the total communication in the interactive leakage model is at most cn for some $c < 1$, then

$$\mathsf{H}_\infty(K^{\mathrm{A}}K^{\mathrm{V}}|X_SY_SS\widetilde{E})_\sigma \geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha})\right)n$$

where β, ν are the constants from Corollary 13.

The extra bit K_i^{V} in the statement of this lemma takes nontrivial value when $a_i[y_i] \neq b_i[y_i]$, and Eve can potentially guess this. This will take the role of Alice's extra input bit z in the definition of MSE, so that it is possible to win MSE on all coordinates in S, even if $a_i[y_i] \neq b_i[x_i]$. In order to use Corollary 13 for our security proof, it is important that it is possible to win MSE on all these coordinates. Using Corollary 13 on Protocol 1 will give us a min-entropy bound including the extra K_i^{V} bits, but these can be taken away later as conditioned on the not-aborting event, K_i^{V} takes non-trivial value on very few coordinates.

Proof of Lemma 14. Consider the $\mathrm{MSE}^{\alpha n/n}_{\mathrm{rand}}$ game being played on the state shared by Alice, Bob and Eve (with S being the random subset of size αn, and $\mathrm{MSE}^{\alpha n/n}_{\mathrm{rand}}$ being won if the instances in the random subset S are won) in Protocol 1. Here K_i^{V} is being interpreted as Alice's input Z_i when $A_i[Y_i] \neq B_i[X_i]$; when $A_i[Y_i] = B_i[X_i]$, Z_i is irrelevant to the winning condition of MSE, so it does not matter that K_i^{V} is trivial here. Let U_i be the indicator variable of the event that Eve guesses $X_iY_iA_i[Y_i]$ correctly, V_i be the indicator variable for the event Eve guesses K_i^{V} correctly and W_i be the indicator variable for the event that $A_i[Y_i] = B_i[X_i]$ for $i \in S$. From Fact 20,

$$\begin{aligned}
\mathsf{H}_\infty(K^{\mathrm{A}}K^{\mathrm{V}}X_SY_S|S\widetilde{E})_\sigma &\geq \log\left(\frac{1}{\Pr[\prod_{i\in S}U_i \wedge (\neg W_i \implies V_i)]}\right) \\
&= \log\left(\frac{1}{\Pr[\text{Win } \mathrm{MSE}^{\alpha n/n}_{\mathrm{rand}}]}\right) \\
&\geq \alpha n \cdot \log\left(\frac{9}{1-\nu+\beta(\sqrt{c}+\sqrt{\alpha})}\right)
\end{aligned}$$

where we have used Corollary 13 along with the upper bound on communication in the last line.

Since X_iY_i are uniformly random on a set of support size 9 we then have by Fact 17,

$$\begin{aligned}
\mathsf{H}_\infty(K^{\mathrm{A}}K^{\mathrm{V}}|X_SY_SS\widetilde{E})_\sigma &\geq \alpha n \cdot \log\left(\frac{9}{1-\nu+\beta(\sqrt{c}+\sqrt{\alpha})}\right) - \log|X_SY_S| \\
&\geq \alpha n \cdot \log\left(\frac{9}{1-\nu+\beta(\sqrt{c}+\sqrt{\alpha})}\right) - \alpha n \cdot \log 9 \\
&= \alpha n \cdot \log\left(\frac{1}{1-\nu+\beta(\sqrt{c}+\sqrt{\alpha})}\right) \\
&\geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha})\right)n.
\end{aligned}$$

□

Lemma 15. *f the total communication in the interactive leakage model is at most cn for some $c < 1$, then*

$$\mathsf{H}^{\varepsilon}_{\infty}(K^{\mathrm{A}}|X_S Y_S S\widetilde{E})_{\varphi} \geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha}) - 4\delta\right) n - \log(1/\Pr[\mathcal{E}])$$

for $\varepsilon = 2\cdot 2^{-8\delta^2\alpha\gamma n}/\Pr[\mathcal{E}]$.

Proof. Firstly, since φ is σ conditioned on an event of probability $\Pr[\mathcal{E}]$, by Fact 11 and the previous lemma we have,

$$\mathsf{H}_{\infty}(K^{\mathrm{A}}K^{\mathrm{V}}|X_S Y_S S\widetilde{E})_{\varphi} \geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha})\right) n - \log(1/\Pr[\mathcal{E}]).$$

Let W_i denote the indicator variable for the event $A_i[Y_i] = B_i[X_i]$ and let φ' denote σ conditioned on the following event which we call $\mathcal{E}'$:

$$\left(\sum_{i\in T} W_i \geq (1-2\delta)|T|\right) \wedge \left(\sum_{i\in S} W_i \geq (1-4\delta)|S|\right).$$

By Fact 4, $\Pr[\mathcal{E}'] \geq 1 - 2^{-8\delta^2\alpha\gamma n}$, which gives us $\|\varphi - \varphi'\|_1 \leq \frac{2\cdot 2^{-8\delta^2\alpha\gamma n}}{\Pr[\mathcal{E}]}$. In φ', $A_i[Y_i]$ and $B_i[X_i]$ differ in at most $4\delta|S|$ many places in S, and K^{V} is a uniformly random bit only in these places. Hence by Fact 17,

$$\mathsf{H}_{\infty}(K^{\mathrm{A}}|X_S Y_S S\widetilde{E})_{\varphi'} \geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha})\right) n - \log(1/\Pr[\mathcal{E}]) - 4\delta\alpha n,$$

which gives us the ε-smoothed bound for φ from ℓ_1 bound between φ and φ'. □

Proof of Theorem 12. First we shall condition the conditional min-entropy bound from Lemma 15 further on (T, A_T). Among these, T is independent of K^{A}, so conditioning on them makes no difference. A_T is contained in K^{A}, and uniformly random in $\{0,1,2\}^{|T|}$. Hence,

$$\mathsf{H}^{\varepsilon}_{\infty}(K^{\mathrm{A}}|X_S Y_S A_T S T\widetilde{E})_{\varphi} \geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha}) - 4\delta\right) n - \log(1/\Pr[\mathcal{E}]) - \alpha\gamma n.$$

Now notice that in ρ, $X_S Y_S S T A_T$ is revelaed to Eve, so she may do some operations on her side depending on these. ρ is thus related to φ by some local operations on the registers $X_S Y_S S T A_T \widetilde{E}$. Hence by Fact 13,

$$\mathsf{H}^{\varepsilon}_{\infty}(K^{\mathrm{A}}|X_S Y_S A_T S T\widetilde{E})_{\rho} \geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha}) - 4\delta - \gamma\right) n - \log(1/\Pr[\mathcal{E}]).$$

Finally, to bound $\mathsf{H}^{\varepsilon}_0(K^{\mathrm{A}}|K^{\mathrm{B}})_{\rho}$, we consider the state ρ', which is conditioned on the event $\mathcal{E}'$ as defined in the proof of Lemma 15 instead of $\mathcal{E}$ like ρ. They satisfy $\|\rho - \rho'\|_1 \leq \frac{2\cdot 2^{-8\delta^2\alpha\gamma n}}{\Pr[\mathcal{E}]}$. The number of strings K^{B} of length that can differ from a given value of K^{A} in at most $4\delta|S|$ places is at most $2^{h_2(4\delta)|S|}$, which gives us $\mathsf{H}_0(K^{\mathrm{B}}|K^{\mathrm{A}})_{\rho'} \leq h_2(4\delta)\alpha n$. Putting everything together we get,

$$\begin{aligned}\mathsf{H}^{\varepsilon}_{\infty}(K^{\mathrm{A}}|X_S Y_S A_T S T\widetilde{E})_{\rho} - \mathsf{H}^{\varepsilon}_0(K^{\mathrm{B}}|K^{\mathrm{A}})_{\rho} &\geq \alpha\left(\nu - 2\beta(\sqrt{c}+\sqrt{\alpha}) - 4\delta - \gamma - h_2(4\delta)\right) n - \log(1/\Pr[\mathcal{E}]) \\ &\geq \alpha\left(\nu - \beta(\sqrt{c}+\sqrt{\alpha}) - 2h_2(4\delta) - \gamma\right) n - 1/\Pr([\mathcal{E}])\end{aligned}$$

for $\delta \leq \frac{1}{2}$.

To lower bound $\Pr[\mathcal{E}]$ in the honest case when each instance of MS is won with probability $1-\delta$, we use the Chernoff bound. Letting W_i denote the indicator variable for $A_i[Y_i] = B_i[X_i]$, the W_i-s are i.i.d. in this case, and the expected value of each W_i is $1-\delta$. Hence

$$\Pr[\neg\mathcal{E}] = \Pr\left[\sum_{i\in T} W_i < (1-2\varepsilon)|T|\right] \leq 2^{-2\delta^2\gamma\alpha n}.$$ □

Acknowledgements

We thank Ernest Tan and Tony Metger for helpful discussions on the security of DIQKD with leakage, in particular on the difficulties of applying a sequential proof technique, and for pointing out references [SPM13, TZCBB+20, TZCWP20].

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.

References

[ABJO21] Divesh Aggarwal, Naresh Boddu, Rahul Jain, and Maciej Obremski. Quantum Measurement Adversary. https://arxiv.org/abs/2106.02766, 2021.

[ABJT20] Anurag Anshu, Mario Berta, Rahul Jain, and Marco Tomamichel. Partially Smoothed Information Measures. *IEEE Transactions on Information Theory*, 66(8):5022–5036, 2020.

[AFDF+18] Rotem Arnon-Friedman, Frédéric Dupuis, Omar Fawzi, Renato Renner, and Thomas Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nature Communications*, 9(1):459, 2018.

[AFRV19] Rotem Arnon-Friedman, Renato Renner, and Thomas Vidick. Simple and Tight Device-Independent Security Proofs. *SIAM Journal on Computing*, 48(1):181–225, 2019.

[BB84] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of International Conference on Computers, Systems and Signal Processing*, page 175, 1984.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to Compress Interactive Communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[BR11] Mark Braverman and Anup Rao. Information Equals Amortized Communication. In *Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS '11)*, page 748–757, 2011.

[BRWY13a] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct Product via Round-Preserving Compression. In *Automata, Languages, and Programming*, pages 232–243, 2013.

[BRWY13b] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct Products in Communication Complexity. In *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS '13)*, page 746–755, 2013.

[BVY17] 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.

[BYJKS02] 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.

[CSUU08] 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.

[CSWY01] 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.

[DFR20] Frédéric Dupuis, Omar Fawzi, and Renato Renner. Entropy Accumulation. *Communications in Mathematical Physics*, 379(3):867–913, 2020.

[DSV15] Irit Dinur, David Steurer, and Thomas Vidick. A Parallel Repetition Theorem for Entangled Projection Games. *Computational Complexity*, 24(2):201–254, 2015.

[HJMR10] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The Communication Complexity of Correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010.

[Hol07] 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.

[JK20] Rahul Jain and Srijita Kundu. A Direct Product Theorem for One-Way Quantum Communication. https://arxiv.org/abs/2008.08963, 2020.

[JMS20] Rahul Jain, Carl A. Miller, and Yaoyun Shi. Parallel Device-Independent Quantum Key Distribution. *IEEE Transactions on Information Theory*, 66(9):5567–5584, 2020.

[JN12] Rahul Jain and Ashwin Nayak. Short Proofs of the Quantum Substate Theorem. *IEEE Transactions on Information Theory*, 58(6):3664–3669, 2012.

[JPY14] 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.

[JPY16] 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.

[JRS02] 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.

[JRS05] 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.

[JRS09] 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.

[JY12] Rahul Jain and Penghui Yao. A Strong Direct Product Theorem in Terms of the Smooth Rectangle Bound. http://arxiv.org/abs/1209.0263, 2012.

[Kla10] 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.

[KN96] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996.

[KRS09] Robert Konig, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, 2009.

[KRT10] Julia Kempe, Oded Regev, and Ben Toner. Unique Games with Entangled Provers are Easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010.

[KŠdW07] 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.

[KT20] Srijita Kundu and Ernest Y.-Z. Tan. Composably secure device-independent encryption with certified deletion. https://arxiv.org/abs/2011.12704, 2020.

[LLR12] Sophie Laplante, Virginie Lerays, and Jérémie Roland. Classical and Quantum Partition Bound and Detector Inefficiency. In *Automata, Languages, and Programming*, pages 617–628, 2012.

[LS09] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends® in Theoretical Computer Science*, 3(4):263–399, 2009.

[LSŠ08] 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.

[PAB+09] Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009.

[Raz92] Alexander A. Razborov. On the Distributional Complexity of Disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[Raz95] Ran Raz. A Parallel Repetition Theorem. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, page 447–456, 1995.

[She18] Alexander A. Sherstov. Compressing Interactive Communication Under Product Distributions. *SIAM Journal on Computing*, 47(2):367–419, 2018.

[SPM13] Jonathan Silman, Stefano Pironio, and Serge Massar. Device-independent randomness generation in the presence of weak cross-talk. *Phys. Rev. Lett.*, 110:100504, 2013.

[TL17] Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, 2017.

[Tom16] Marco Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016.

[Tsi87] B. S. Tsirelson. Quantum analogues of the bell inequalities. the case of two spatially separated domains. *Journal of Soviet Mathematics*, 36(4):557–570, 1987.

[TZCBB+20] Armin Tavakoli, Emmanuel Zambrini Cruzeiro, Jonatan Bohr Brask, Nicolas Gisin, and Nicolas Brunner. Informationally restricted quantum correlations. *Quantum*, 4:332, 2020.

[TZCWP20] Armin Tavakoli, Emmanuel Zambrini Cruzeiro, Erik Woodhead, and Stefano Pironio. Informationally restricted correlations: a general framework for classical and quantum systems, 2020.

[Vid17] Thomas Vidick. Parallel DIQKD from parallel repetition. https://arxiv.org/abs/1703.08508, 2017.

[VV19] Umesh Vazirani and Thomas Vidick. Fully device independent quantum key distribution. *Communications of the ACM*, 62(4):133, 2019.

[VW08] Emanuele Viola and Avi Wigderson. Norms, XOR Lemmas, and Lower Bounds for Polynomials and Protocols. *Theory of Computing*, 4(7):137–168, 2008.

[Yao77] Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (SFCS 1977)*, pages 222–227, 1977.

[Yue16] 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.

A Proof of Lemma 5

We shall do induction on the number of rounds. Let c_i be the communication in the i-th round and $\mathcal{R}^j = \{j, j+l, \ldots\}$ denote the set of rounds in which the j-th player communicates, so that $\sum_{i\in\mathcal{R}^j} c_i = c^j$. Let M_i be the message register of the i-th round, E_i be the memory register the party who communicates in the i-th round holds after sending their message. For $i \in \mathcal{R}^j$, the registers held by the j-th party at the beginning of the i-th round are messages $M^j_{i-l+1} \ldots M^j_{i-1}$ from other parties in the $(i-l+1)$-th to $(i-1)$-th rounds, which we shall jointly denote by N^j_{i-1}, and their memory register E_{i-l} which they have retained from the $(i-l)$-th round. We shall denote all other (non-input) registers held by parties other than the j-th party at the beginning of the I-th round by F^{-j}_{i-1}. Since $i \in \mathcal{R}^j$, clearly $F^{-j}_i = F^{-j}_{i-1}M_i$. Using X to denote $X^1 \ldots X^l$ and similar notation for $\widetilde{X}$, we shall call the shared state including the input purifications at the beginning of the the i-th round

$$|\sigma^i\rangle_{X\widetilde{X}N^j_{i-1}E_{i-l}F^{-j}_{i-1}} = \sum_x \sqrt{\mathsf{P}_X(x)}\, |xx\rangle_{X\widetilde{X}}\, |\sigma^i\rangle_{N^j_{i-1}E_{i-l}F^{-j}_{i-1}}\,.$$

For the base case $i=1$, communication is zero. Since $\mathsf{P}_{X^1\dots X^l}$ is a product distribution, $\sigma^1_{X^jX^{-j}\widetilde{X}^{-j}F_0^{-j}}$ is product between X^j and the other registers, F_0^{-j} being simply the other parties' parts of the initial shared entangled state, which is independent of the inputs. So the condition trivially holds. For the induction step, we shall assume the condition

$$\mathsf{D}_\infty\left(\sigma^i_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\middle\|\sigma^i_{X^j}\otimes\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\right)\le 2\sum_{\substack{i'\in\mathcal{R}^j,\\ i'<i}}c_{i'}$$

holds at the beginning of the i-th round, where $i\in\mathcal{R}^j$, for some state $\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}$, and see how it changes in the i-th to $(i+l-1)$-th rounds.

In the i-th round, the j-th party applies a unitary on the $X^jN^j_{i-1}E_{i-l}$ registers, getting registers $X^jM_iE_i$. By Fact 16, there exists a state $\tilde{\rho}^{i+1}_{M_i}$ such that

$$\mathsf{D}_\infty\left(\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}M_i}\middle\|\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\right)\le 2c_i.$$

Now note that the marginal states $\sigma^i_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}$ and $\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}$ are exactly the same, since the unitary relating $|\sigma^i\rangle$ and $|\sigma^{i+1}\rangle$ does not act on $X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}$ at all, and only uses X^j as a control register. Hence we have,

$$\begin{aligned}
&\mathsf{D}_\infty\left(\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\middle\|\sigma^{i+1}_{X^j}\otimes\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\right)\\
&=\mathsf{D}_\infty\left(\sigma^{i}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\middle\|\sigma^{i}_{X^j}\otimes\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\right)\\
&=\mathsf{D}_\infty\left(\sigma^{i}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\middle\|\sigma^{i}_{X^j}\otimes\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\right)\\
&\le 2\sum_{\substack{i'\in\mathcal{R}^j,\\ i'<i}}c_{i'}.
\end{aligned}$$

Now using Fact 12 we can say,

$$\begin{aligned}
&\mathsf{D}_\infty\left(\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}M_i}\middle\|\sigma^{i+1}_{X^j}\otimes\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\right)\\
&\le\mathsf{D}_\infty\left(\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}M_i}\middle\|\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\right)\\
&\quad+\mathsf{D}_\infty\left(\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\middle\|\sigma^{i+1}_{X^j}\otimes\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}\right)\\
&\le 2c_i+2\sum_{\substack{i'\in\mathcal{R}^j,\\ i'<i}}c_{i'}\\
&=2\sum_{\substack{i'\in\mathcal{R}^j,\\ i'\le i}}c_{i'}.
\end{aligned}$$

Hence the condition holds at the beginning of the $(i+1)$-th round with $\rho^{i+1}_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}M_i}=\rho^i_{X^{-j}\widetilde{X}^{-j}F_{i-1}^{-j}}\otimes\tilde{\rho}^{i+1}_{M_i}$.

In the $(i+1)$-th round, the $(j+1)$-th player applies a unitary on the $X^{j+1}N_i^{j+1}E_{i-l+1}$ registers, getting registers $X^j M_{i+1}^1 \ldots M_{i+1}^j \ldots M_{i+1}^l E_{i+1}$, of which they send M_{i+1}^j to the j-th player. So after this round, the registers held by the j-th player are $E_i M_{i+1}^j$, and F_{i+1}^{-j} does not include M_{i+1}^j. By Fact 10 we have that,

$$\begin{aligned}\mathsf{D}_\infty\left(\sigma^{i+2}_{X^jX^{-j}\widetilde{X}^{-j}M^j_{i+1}F_i^{-j}}\middle\|\sigma^{i+2}_{X^j}\otimes\rho^{i+2}_{X^{-j}\widetilde{X}^{-j}M^j_{i+1}F_i^{-j}}\right) &= \mathsf{D}_\infty\left(\sigma^{i+1}_{X^jX^{-j}\widetilde{X}^{-j}F^{-j}_{i-1}M_i}\middle\|\sigma^{i+1}_{X^j}\otimes\rho^{i+1}_{X^{-j}\widetilde{X}^{-j}F^{-j}_{i-1}M_i}\right)\\ &\leq 2\sum_{\substack{i'\in\mathcal{R}^j,\\ i'\leq i}} c_{i'}\end{aligned}$$

where ρ^{i+2} is the state obtained by applying the $(j+1)$-th player's unitary in the $(i+1)$-th round to ρ^{i+1}. From this we can trace out the M_{i+1}^j-th register to show that

$$\mathsf{D}_\infty\left(\sigma^{i+2}_{X^jX^{-j}\widetilde{X}^{-j}F^{-j}_{i+1}}\middle\|\sigma^{i+2}_{X^j}\otimes\rho^{i+2}_{X^{-j}\widetilde{X}^{-j}F^{-j}_{i+1}}\right)\leq 2\sum_{\substack{i'\in\mathcal{R}^j,\\ i'\leq i}} c_{i'}.$$

The bound is similarly unchanged in the rounds $i+2,\ldots,i+l-1$. Hence we can say that at the beginning of the next round $i+l$ in which the j-th party communicates, it holds that

$$\mathsf{D}_\infty\left(\sigma^{i+l}_{X^jX^{-j}\widetilde{X}^{-j}F^{-j}_{i+l-1}}\middle\|\sigma^{i+l}_{X^j}\otimes\rho^{i+l}_{X^{-j}\widetilde{X}^{-j}F^{-j}_{i+l-1}}\right)\leq 2\sum_{\substack{i'\in\mathcal{R}^j,\\ i'<i+l}} c_{i'}.$$

B Proof of Lemma 6

We shall show that if there is a quantum interactive protocol $\mathcal{P}$ for V with c qubits of communication and error probability at most ε, over input distribution p, then there is a zero-communication quantum protocol $\mathcal{P}''$ which does not abort with probability 2^{-2c} worst case over all inputs, and when it does not abort it computes V with the same error probability over p.

Firstly, we can use entanglement and teleportation to get a protocol $\mathcal{P}'$ from $\mathcal{P}$, which only involves at most $2c$ bits of classical communication (with the players doing measurements according to the classical messages they receive and their inputs, on their parts of a shared entangled state). We assume that the number of bits communicated in $\mathcal{P}'$ is of some fixed length every round for every input, with the total communication being $2c$ (this can be done by padding dummy bits if necessary).

Now in the zero-communication protocol $\mathcal{P}''$, the players will share the same initial entangled state as in $\mathcal{P}'$, and also $2c$ uniformly random classical bits. If player j communicates in the i-th round, let $r_i = r_i^1 \ldots r_i^{j-1} r_i^{j+1} \ldots r_i^l$ denote the portion of the shared randomness that corresponds to the bits in the i-th round of communication in $\mathcal{P}'$, with r_i^k corresponding to the message to the k-th player. On inputs $x^1 \ldots x^l$, the players do the following in $\mathcal{P}''$:

- For each round i, if player j is the one communicating in that round, player j assumes $r_{i-l+1}^j \ldots r_{i-1}^j$ are the classical messages she has received from the other $l-1$ players between the $(i-l)$-th and the i-th round. They do a measurement on their part of the entangled state as she does in the i-th round of $\mathcal{P}'$, depending on x^j, their previous measurement outcomes, and messages from the other players. If r_i is not compatible with her input and these measurement outcomes and previous messages, then player j outputs $\perp$.

- At the end, if a player has not output $\perp$ yet, they output according to $\mathcal{P}'$.

Once the outputs of the measurements are fixed, the protocol is deterministic. So a transcript that is separately compatible for all the players, is compatible for all of them, and there is exactly one such transcript. $\{r_i\}_i$ is equal to this transcript with probability 2^{-2c}, and hence no player outputs $\perp$ with probability 2^{-2c}. When they do not output $\perp$, the trancript is correct for input $x^1 \dots x^l$, and hence $\mathcal{P}''$ is correct with probability at least $1 - \varepsilon$ over the distribution p on $x^1 \dots x^l$, due to the correctness of $\mathcal{P}'$.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il