Parallel Repetition: Simplifications and the No-Signaling Case

Thomas Holenstein*

Abstract

Consider a game where a referee chooses (x, y) according to a publicly known distribution P_{XY}, sends x to Alice, and y to Bob. Without communicating with each other, Alice responds with a value a and Bob responds with a value b. Alice and Bob jointly win if a publicly known predicate $Q(x, y, a, b)$ holds.

Let such a game be given and assume that the maximum probability that Alice and Bob can win is $v < 1$. Raz (SIAM J. Comput. 27, 1998) shows that if the game is repeated n times in parallel, then the probability that Alice and Bob win *all* games simultaneously is at most $\bar{v}^{\frac{n}{\log(s)}}$, where s is the maximal number of possible responses from Alice and Bob in the initial game, and $\bar{v} < 1$ is a constant depending only on v.

In this work, we simplify Raz's proof in various ways and thus shorten it significantly. Further we study the case where Alice and Bob are not restricted to local computations and can use any strategy which does not imply communication among them.

1 Introduction

The question how much parallel repetition of a game as in the abstract reduces the winning probability of the players was motivated by the study of two-prover interactive proofs, initiated by Ben-Or et al. [BOGKW88]. It was first conjectured that in a game which is repeated n times in parallel, the probability that Alice and Bob win all the games simultaneously is at most v^n (see [FRS94]). However, later a counterexample to this conjecture was given [For89].

Related Work Various papers give upper bounds on the winning probability of a game which is repeated n times in parallel [CCL92, Fei91, LS95, Raz98, Ver94]. However, the upper bound given by Raz [Raz98] is the only explicit bound for arbitrary distributions P_{XY} (it is also quantitatively the strongest). Parnafes, Raz, and Wigderson [PRW97] modify Raz's proof to show that the term $\log(s)$ can be replaced by a parameter which is much smaller for some games.

*Microsoft Research, Silicon Valley; thomahol@microsoft.com. This work was done while the author was at ETH Zurich.

Games for which the n-fold parallel repetition decreases the winning probability less than from v to v^n were also constructed: Fortnow [For89] gives[1] a game for which the maximal winning probability in two repetitions is larger than v^2 (see also [FL92]), Feige [Fei91] constructs a game where the winning probability in two parallel repetitions does not decrease at all, and Feige and Verbitsky [FV02] give, for infinitely many s, a game where $\Theta(\frac{\log(s)}{\log\log(s)})$ repetitions decrease the winning probability from at most $\frac{3}{4}$ to at least $\frac{1}{8}$, where s is the number of possible answers Alice and Bob can give. This last result shows that in general Raz's bound is close to optimal.

No-signaling strategies No-signaling strategies are all those strategies which do not imply communication. Popescu and Rohrlich [PR94] give an example of such a strategy: Alice receives a bit x, Bob receives a bit y, and they respond with uniform random bits a and b such that $a \oplus b = x \wedge y$. Note that even though we cannot implement this strategy with shared randomness and without communication, Alice and Bob cannot communicate if they only have black-box access to such functionality.

The study of no-signaling strategies is motivated by the idea that if Alice and Bob share some entangled quantum state, the set of possible strategies they might use increases, but stays a subset of the no-signaling strategies (this subset is strict: for example the above strategy which achieves $a \oplus b = x \wedge y$ from (x, y) cannot be simulated perfectly using quantum mechanics [NC00, Problem 2.3], [Cir80] — the corresponding game is called the CHSH-game [CHSH69]).

We remark that there are games which can be won with probability 1 given a shared quantum state (and thus with a no-signaling strategy), but not using local strategies. Those are called "pseudo-telepathy games" (see [BBT05] and the references therein).

A parallel repetition theorem for the case where Alice and Bob share a quantum state and the decision of the referee only depends on the XOR of the binary answers of Alice and Bob was recently given by Cleve et al. [CSUU06].

Contributions of this paper In this paper we simplify Raz's proof. Most importantly, we replace a large part (essentially Section 6) of Raz's paper with the simpler Lemma 8. This also allows us to give an explicit bound on the maximal winning probability of a game repeated n times in parallel (Raz does not explicitly describe the dependence of $\bar{v}$ on v).

The use of Lemma 8 also makes the rest of the argument simpler. We shortly explain why: The main part of the proof consists of showing that the information the players get in the n-fold repetition does not help them to win the subgame in some coordinate j, even conditioned on the event that certain other subgames are won. This is done in three steps. In two of these steps the information does not help the players because they can generate this information themselves with local computation only. Lemma 8 shows that this also holds for the third step. This allows us to merge some of the steps, which simplifies the overal structure.

[1]For readers not familiar with such counter-examples, a variation of Fortnow's game is reproduced in Appendix A.

We also study how much the term $\log(s)$ in the exponent in the parallel repetition theorem can be reduced. In [PRW97] it is shown that the logarithm of the partition number of the accepance predicate can be used instead of $\log(s)$. Based on the ideas from there, Theorem 17 gives a bound which might be stronger for some games.

Finally, we prove a parallel repetition theorem in case Alice and Bob are restricted to no-signaling strategies (in both the given game and the parallel repetition of it).

2 Notation and Basic Facts

2.1 Probability Distributions

We use calligraphic letters to denote sets. We denote random variables using capital letters, and values with lower case letters. We use superscripts to denote tuples, e.g., $X^n := (X_1, \ldots, X_n)$ and $x^n := (x_1, \ldots, x_n)$.

If a distribution P_{XY} over $\mathcal{X} \times \mathcal{Y}$ is given, we write P_X or P_Y to denote the marginal distribution, e.g., $\mathsf{P}_X(x) := \sum_{y \in \mathcal{Y}} \mathsf{P}_{XY}(x, y)$. The conditional distribution $\mathsf{P}_{Y|X=x}$ is $\mathsf{P}_{Y|X=x}(y) := \mathsf{P}_{XY}(x, y)/\mathsf{P}_X(x)$.

Let P_{X_0} be a distribution over $\mathcal{X}$ and $\mathsf{P}_{Y_1|X_1=x}$ be a conditional distribution over $\mathcal{Y}$. We define the distribution $\mathsf{P}_{X_0}\mathsf{P}_{Y_1|X_1}$ over $\mathcal{X} \times \mathcal{Y}$ as

$$(\mathsf{P}_{X_0}\mathsf{P}_{Y_1|X_1})(x, y) := \mathsf{P}_{X_0}(x) \cdot \mathsf{P}_{Y_1|X_1=x}(y). \tag{1}$$

For this, it is necessary that $\mathsf{P}_{Y_1|X_1=x}$ is defined for every $x \in \mathcal{X}$. We also use this notation when $\mathsf{P}_{Y_1|X_1=x}$ is defined as marginal of a given distribution $\mathsf{P}_{X_1Y_1}$. In this case, we define $\mathsf{P}_{Y_1|X_1=x}$ in an arbitrary way if $\mathsf{P}_{X_1}(x) = 0$. This notation is used for example in Corollary 9 in the form $\mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X}$, where it is understood as $(\mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X})(x, y, s) := \mathsf{P}_{X_0Y_0}(x, y)\mathsf{P}_{S|X=x}(s)$. Note that the conditional distribution $\mathsf{P}_{S|X=x}$ is defined there by the marginal distribution P_{SX} of the given distribution P_{SXY}. Our notation is not explicit since it does not specify which random variables are associated with each other. However, this will always be clear from the context.

For two probability distributions P_{X_0} and P_{X_1} over the same set $\mathcal{X}$ we define the statistical distance

$$\|\mathsf{P}_{X_0} - \mathsf{P}_{X_1}\| := \frac{1}{2} \sum_{x \in \mathcal{X}} \big|\mathsf{P}_{X_0}(x) - \mathsf{P}_{X_1}(x)\big|. \tag{2}$$

2.2 Games

Definition 1. *A* game $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$ *is a distribution* P_{XY} *over* $\mathcal{X} \times \mathcal{Y}$ *and a predicate* Q *over* $\mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$*. The* value $v(\mathfrak{G})$ *of a game is*

$$v(\mathfrak{G}) := \max_{h_a, h_b} \Pr_{XY}[Q(X, Y, h_a(X), h_b(Y))],$$

where the maximization is over functions $h_a : \mathcal{X} \to \mathcal{A}$ *and* $h_b : \mathcal{Y} \to \mathcal{B}$*. A* strategy (h_a, h_b) *for a game is a pair of such functions.*

Sometimes also randomized strategies for Alice and Bob are considered, where h_a and h_b also depend on (the same) shared randomness r chosen according to some distribution P_R. However, there always exists an $r \in \mathcal{R}$ such that

$$\begin{aligned}\Pr_{RXY}[Q(X,Y,h_a(X,R),h_b(Y,R))] &= \mathop{\mathrm{E}}_{R}\big[\Pr_{XY}[Q(X,Y,h_a(X,R),h_b(Y,R))]\big] \\ &\leq \Pr_{XY}[Q(X,Y,h_a(X,r),h_b(Y,r))], \end{aligned} \tag{3}$$

and we see that the definition of the value is robust against such a change. Individual (local) randomness can be obtained from shared randomness and is thus a special case of the above.

Definition 2. *The n-fold* parallel repetition $\mathfrak{G}^n$ *of a game* $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ *over* $\mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$ *is the game over* $\mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{A}^n \times \mathcal{B}^n$ *which is given by* $\mathfrak{G}^n := (\mathsf{P}_{X^nY^n}, Q^{\wedge n})$ *where*

$$\begin{aligned}\mathsf{P}_{X^nY^n}(x^n,y^n) &:= \prod_{i=1}^{n} \mathsf{P}_{XY}(x_i,y_i), \quad \textit{and} \\ Q^{\wedge n}(x^n,y^n,a^n,b^n) &:= \bigwedge_{i=1}^{n} Q(x_i,y_i,a_i,b_i).\end{aligned}$$

If a strategy is given, the distribution $\mathsf{P}_{X^nY^nA^nB^n}$ of queries and answers is defined in the obvious way. We further define, for all i, the event W_i which occurs if the ith subgame is won.

Definition 3. *For a game $\mathfrak{G}^n$ and a strategy (h_a, h_b) the distribution $\mathsf{P}_{X^nY^nA^nB^n}$ over $\mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{A}^n \times \mathcal{B}^n$ is given by*

$$\mathsf{P}_{X^nY^nA^nB^n}(x^n,y^n,a^n,b^n) := \begin{cases} \mathsf{P}_{X^nY^n}(x^n,y^n) & \textit{if } h_a(x^n) = a^n \textit{ and } h_b(y^n) = b^n \\ 0 & \textit{otherwise.} \end{cases}$$

Further, W^n is the tuple of events $(W_1, \ldots, W_n)$ where $W_i :\Longleftrightarrow Q(X_i, Y_i, A_i, B_i)$.

We prove the following version of the parallel repetition theorem.

Theorem 4 (Parallel Repetition Theorem)**.** *For any game $\mathfrak{G}$ with value $v := v(\mathfrak{G})$ and any integer n:*

$$v(\mathfrak{G}^n) \leq \Big(1 - \frac{(1-v)^3}{6000}\Big)^{\frac{n}{\log(|\mathcal{A}||\mathcal{B}|)}}.$$

The constant 6000 could be improved by a more carful analysis (we will not optimize constants which would improve it during the proof). However, we do not know whether the 3 in the exponent can be reduced.

In [PRW97] it is shown that in Raz's proof the term $\log(|\mathcal{A}||\mathcal{B}|)$ in the exponent can be reduced to the maximum of the logarithm of the partition number of $Q(x, y, \cdot, \cdot)$. As shown by Beame [Bea06], the argument can be adapted to work with the proof given here. We give a slightly different argument in Section 8 which shows how the term can be reduced to a quantity which is a lower bound on the logarithm of the partition number.

3 Proof Sketch

Fix an arbitrary game $\mathfrak{G}$, its n-fold parallel repetition $\mathfrak{G}^n$, and a strategy h_a, h_b for $\mathfrak{G}^n$. With the notation from Definition 3, the parallel repetition theorem is simply an upper bound on $\Pr[W_1 \wedge \cdots \wedge W_n]$. To get such an upper bound, we show that for arbitrary indices $i_1, \ldots, i_m$ there exists an index j such that

$$\Pr[W_j | W_{i_1} \wedge \cdots \wedge W_{i_m}] \leq v(\mathfrak{G}) + \varepsilon, \tag{4}$$

where ε depends on m, n, $\log(|\mathcal{A}||\mathcal{B}|)$, and $\Pr[W_{i_1} \wedge \cdots \wedge W_{i_m}]$ (this is Lemma 15). From (4) a simple induction gives the parallel repetition theorem, thus we now concentrate on the proof of (4).

Locally Computable Embeddings In order to prove (4) we define the distribution

$$\mathsf{P}_{\widetilde{X}^n \widetilde{Y}^n} := \mathsf{P}_{X^n Y^n | W_{i_1} \wedge \cdots \wedge W_{i_m}} \tag{5}$$

(i.e., the distribution of the message which the referee sends to Alice and Bob conditioned on the event that the games i_1 to i_m are won).

We show (Lemma 14) that for some j the following can be achieved by Alice and Bob without communication and using shared randomness only:

1. Alice, on input x, produces a tuple $\bar{x}^n$ with $\bar{x}_j = x$.
2. Bob, on input y, produces a tuple $\bar{y}^n$ with $\bar{y}_j = y$.
3. Let $\mathsf{P}_{\overline{X}^n \overline{Y}^n}$ be the resulting joint distribution of the tuples $(\bar{x}^n, \bar{y}^n)$, assuming that (x, y) is chosen according to P_{XY}. Then

$$\|\mathsf{P}_{\overline{X}^n \overline{Y}^n} - \mathsf{P}_{\widetilde{X}^n \widetilde{Y}^n}\| \leq \varepsilon.$$

We say that (X, Y) can be $1-\varepsilon$-embedded into $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X, Y)$ by local computation.

If such an embedding is given, we can consider the following strategy for the initial game $\mathfrak{G}$: Alice and Bob embed their inputs (X, Y) in $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X, Y)$, and answer with coordinate j of $h_a(\widetilde{X}^n)$ and $h_b(\widetilde{Y}^n)$. This strategy wins with probability at least $\Pr[W_j | W_{i_1} \wedge \cdots \wedge W_{i_m}] - \varepsilon$. Since no strategy for the initial game has higher winning probability than $v(\mathfrak{G})$ this implies (4).

We remark that a necessary condition for such an embedding to exist is that

$$\|\mathsf{P}_{XY} - \mathsf{P}_{\widetilde{X}_j \widetilde{Y}_j}\| \leq \varepsilon, \tag{6}$$

and indeed this follows from Lemma 5 for $U_j = (X_j, Y_j)$ (of course this condition is not a sufficient one).

Constructing an Embedding We now give a more detailled explanation how Alice and Bob can embed (X, Y) into $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X, Y)$. For this, given values (x, y) distributed according to P_{XY}, Alice and Bob proceed as follows:

1. Alice and Bob use shared randomness to produce queries and responses for all the won games, i.e., values $(x_{i_1}, y_{i_1}, a_{i_1}, b_{i_1})$ to $(x_{i_m}, y_{i_m}, a_{i_m}, b_{i_m})$. Here, Alice and Bob *both* produce *all* these values.
2. For every index $i \notin \{i_1, \ldots, i_m, j\}$, Alice and Bob examine a shared random bit d_i. If $d_i = 1$ both locally produce x_i, otherwise both locally produce y_i. Again, Alice and Bob both produce all these values.
3. Using individual randomness, Alice and Bob locally expand their information such that Alice gets x^n and Bob y^n.

In steps 1 and 2 we have to take care of two things: first, the values produced should be distributed according to the the respective marginal of the distribution $\mathsf{P}_{\widetilde{A}^n \widetilde{B}^n \widetilde{X}^n \widetilde{Y}^n | \widetilde{X}_j = x \wedge \widetilde{Y}_j = y}$ (where $\mathsf{P}_{\widetilde{A}^n \widetilde{B}^n \widetilde{X}^n \widetilde{Y}^n}$ is defined analogously to (5)). Second, Alice and Bob should produce *equal values* (otherwise the resulting random variables $(\overline{X}^n, \overline{Y}^n)$ will not have the correct overall distribution).

For step 1 achieving both is simple: it follows from Corollary 6 that Alice and Bob can choose the values $(x_{i_1}, y_{i_1}, a_{i_1}, b_{i_1}), \ldots, (x_{i_m}, y_{i_m}, a_{i_m}, b_{i_m})$ independently of (x, y) according to $\mathsf{P}_{\widetilde{X}_{i_1} \widetilde{Y}_{i_1} \widetilde{A}_{i_1} \widetilde{B}_{i_1} \cdots \widetilde{X}_{i_m} \widetilde{Y}_{i_m} \widetilde{A}_{i_m} \widetilde{B}_{i_m}}$. Using shared randomness this can be done such that both get the same tuple.

The second step is harder, as in this case the values cannot be chosen independently of (x_j, y_j) anymore.[2] However, let $\widetilde{S}$ be the random variables which Alice and Bob produce in this step. It will follow from Corollary 6 that $\|\mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{X}_j} - \mathsf{P}_{XY\widetilde{S}}\|$ and $\|\mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{Y}_j} - \mathsf{P}_{XY\widetilde{S}}\|$ are both small, and Lemma 8 implies that this is sufficient to generate $\widetilde{S}$ locally.

In fact, Corollary 6 and Lemma 8 are strong enough to do steps 1 and 2 at the same time, and thus these steps are done simultaneously in the proof of Lemma 14.

Step 3 will be simpler to implement. Because the players also computed a_{i_ℓ} and b_{i_ℓ} in step 1, they can expand their known values according to the given distributions and the resulting distribution will be correct (this follows from Lemma 10, and a detailed explanation is in the proof of Lemma 14).

4 Conditioned Distributions

The following lemma is essentially Claim 5.1 in Raz's paper [Raz98] (and we use the proof given there). It states that if random variables U_i are chosen independently,

[2] The values d_i can be chosen independently, but *not* the values of x_i respective y_i. We quickly explain why this is impossible in general. Assume that the random variables X and Y contain a shared bit B. The game $\mathfrak{G}^n$ and the strategy (h_a, h_b) may be such that Alice and Bob win subgame i_1 in case $B_1 \oplus \cdots \oplus B_n = 0$. Generating the values independently of (x, y) would now produce a distribution with statistical distance at least $\frac{1}{2}$ from the target distribution. Therefore, a bit which is contained in both x and y *must* be considered when generating the values of x_i and y_i.

then conditioning on an event does not change the individual distributions a lot on average.

Lemma 5. *Let* $\mathsf{P}_{U^k} := \mathsf{P}_{U_1} \dots \mathsf{P}_{U_k}$ *be a probability distribution over* $\mathcal{U}^k$*,* W *an event. Then,*

$$\Pr[W] \leq 2^{-\sum_{j=1}^{k} (\|\mathsf{P}_{U_j|W} - \mathsf{P}_{U_j}\|)^2}. \tag{7}$$

As an example, let U_i be uniform and independent bits and W be the event that at least $k(\frac{1}{2} + \varepsilon)$ of these bits are one. Then $\|P_{U_i|W} - \mathsf{P}_{U_i}\| \geq \varepsilon$ and the lemma states that $\Pr[W] \leq 2^{-k\varepsilon^2}$, which is a version of Chernoff's inequality (note that this implies that Lemma 5 is almost tight; see, for example, [HR06]).

Using $(\sum_{j=1}^{k} a_j)^2 \leq k \sum_{j=1}^{k} a_j^2$ one easily checks that (7) implies

$$\sum_{j=1}^{k} \|\mathsf{P}_{U_j|W} - \mathsf{P}_{U_j}\| \leq \sqrt{k \log\Bigl(\frac{1}{\Pr[W]}\Bigr)}\ , \tag{8}$$

which is the form we use later.

Proof. For two distributions P_S and P_T over the same set $\mathcal{S}$, the *relative entropy* $D(\mathsf{P}_S \| \mathsf{P}_T)$ is defined as

$$D(\mathsf{P}_S \| \mathsf{P}_T) := \sum_{s \in \mathcal{S}} \mathsf{P}_S(s) \log\Bigl(\frac{\mathsf{P}_S(s)}{\mathsf{P}_T(s)}\Bigr). \tag{9}$$

This quantity satisfies $D(\mathsf{P}_S \| \mathsf{P}_T) \geq \bigl(\|\mathsf{P}_S - \mathsf{P}_T\|\bigr)^2$ (see [CT91, Lemma 12.6.1]). Also, if $\mathsf{P}_{U^k} = \mathsf{P}_{U_1} \dots \mathsf{P}_{U_k}$ and P_{V^k} are distributions over the set $\mathcal{U}^k$, then $\sum_{j=1}^{k} D(\mathsf{P}_{V_j} \| \mathsf{P}_{U_j}) \leq D(\mathsf{P}_{V^k} \| \mathsf{P}_{U^k})$ (see Appendix B).

Using the above we get

$$\begin{aligned}
\sum_{j=1}^{k} \Bigl(\|\mathsf{P}_{U_j|W} - \mathsf{P}_{U_j}\|\Bigr)^2 &\leq \sum_{j=1}^{k} D(\mathsf{P}_{U_j|W} \| \mathsf{P}_{U_j}) \\
&\leq D(\mathsf{P}_{U^k|W} \| \mathsf{P}_{U^k}) \\
&= \sum_{u^k} \mathsf{P}_{U^k|W}(u^k) \log\Bigl(\frac{\mathsf{P}_{U^k|W}(u^k)}{\mathsf{P}_{U^k}(u^k)}\Bigr) \\
&= \sum_{u^k} \mathsf{P}_{U^k|W}(u^k) \log\Bigl(\frac{\Pr[W|U^k = u^k]}{\Pr[W]}\Bigr) \\
&= \log\Bigl(\frac{1}{\Pr[W]}\Bigr) + \sum_{u^k} \mathsf{P}_{U^k|W}(u^k) \log\bigl(\Pr[W|U^k = u^k]\bigr) \\
&\leq \log\Bigl(\frac{1}{\Pr[W]}\Bigr).
\end{aligned}$$

□

We now give a slight extension of this lemma (this makes it simpler to apply later). First, the U_j are independent given the value of an additional random variable T. Second, an arbitrary third random variable V with bounded alphabet size gives side information about U_j. Then, choosing U_j without considering the fact that an event W happened and ignoring V does not change the distribution of U_j too much on average. For the notation in the following corollary we refer to Section 2.1, equation (1) and the subsequent remarks.

Corollary 6. *Let $\mathsf{P}_{TU^kV} := \mathsf{P}_T \mathsf{P}_{U_1|T} \mathsf{P}_{U_2|T} \ldots \mathsf{P}_{U_k|T} \mathsf{P}_{V|TU^k}$ be a probability distribution over $\mathcal{T} \times \mathcal{U}^k \times \mathcal{V}$, W be an event. Then,*

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W} \mathsf{P}_{U_j|T} \right\| \leq \sqrt{k} \sqrt{\log(|\mathcal{V}^*|) + \log\Big(\frac{1}{\Pr[W]}\Big)},$$

where $\mathcal{V}^* := \{v \in \mathcal{V} | \mathsf{P}_{V|W}(v) > 0\}$.

The proof is essentially an application of Jensen's inequality on Lemma 5.

Proof. Fix a pair $(t, v) \in \mathcal{T} \times \mathcal{V}$ and consider the distributions $\mathsf{P}_{U^k|T=t,V=v,W}$ and $\mathsf{P}_{U^k|T=t}$. We apply Lemma 5 (in the form given by (8)) on these distributions (with the event $(V{=}v) \wedge W$) and get

$$\begin{aligned}
\sum_{j=1}^{k} \left\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W} \mathsf{P}_{U_j|T} \right\| &= \sum_{t \in \mathcal{T}, v \in \mathcal{V}^*} \mathsf{P}_{TV|W}(t,v) \cdot \sum_{j=1}^{k} \left\| \mathsf{P}_{U_j|T=t,V=v,W} - \mathsf{P}_{U_j|T=t} \right\| \\
&\leq \sum_{t \in \mathcal{T}, v \in \mathcal{V}^*} \mathsf{P}_{TV|W}(t,v) \sqrt{k \log\Big(\frac{1}{\Pr[W \wedge V = v | T = t]}\Big)} \\
&\leq \sqrt{k \log\Big(\sum_{t \in \mathcal{T}, v \in \mathcal{V}^*} \mathsf{P}_{TV|W}(t,v) \frac{1}{\Pr[W \wedge V = v | T = t]}\Big)},
\end{aligned} \tag{10}$$

where the last inequality is Jensen's inequality applied on the function $\sqrt{\log(\cdot)}$ which is concave on $[1, \infty)$. We compute

$$\begin{aligned}
\sum_{t \in \mathcal{T}, v \in \mathcal{V}^*} \mathsf{P}_{TV|W}(t,v) \frac{1}{\Pr[W \wedge V = v | T = t]} &= \sum_{t \in \mathcal{T}, v \in \mathcal{V}^*} \frac{\Pr[T = t \wedge V = v | W]}{\Pr[W \wedge V = v | T = t]} \\
&= \sum_{t \in \mathcal{T}, v \in \mathcal{V}^*} \frac{\Pr[T = t \wedge V = v \wedge W] \Pr[T = t]}{\Pr[W] \Pr[V = v \wedge T = t \wedge W]} \\
&= \sum_{t \in \mathcal{T}, v \in \mathcal{V}^*} \frac{\Pr[T = t]}{\Pr[W]} = \frac{|\mathcal{V}^*|}{\Pr[W]}.
\end{aligned}$$

Inserting this into (10) completes the proof. ∎

5 Embedding by Local Computation

We next study under what conditions random variables can be embedded into other random variables by local computations.

Definition 7 (Embeddable). *For two distributions $\mathsf{P}_{X_0Y_0}$ and $\mathsf{P}_{X_1SY_1T}$ we say that (X_0, Y_0) is $1-\varepsilon$-embeddable in (X_1S, Y_1T) with $(X_1, Y_1) = (X_0, Y_0)$ if there exists a probability measure P_R over a set $\mathcal{R}$ and functions $f_A : \mathcal{X} \times \mathcal{R} \to \mathcal{S}$, $f_B : \mathcal{Y} \times \mathcal{R} \to \mathcal{T}$, such that*

$$\left\| \mathsf{P}_{X_0Y_0}\mathsf{P}_{F_AF_B|XY} - \mathsf{P}_{X_1Y_1ST} \right\| \leq \varepsilon,$$

where $\mathsf{P}_{F_AF_B|X=xY=y}$ is the distribution defined by the random variable $(f_A(x,R), f_B(y,R))$.

The following lemma gives a condition under which (X,Y) is embeddable in (XS, YS). It is one of the main contributions of this paper.

Lemma 8. *Let a distribution P_{SXY} be given. If*

$$\|\mathsf{P}_{SXY} - \mathsf{P}_{XY}\mathsf{P}_{S|X}\| \leq \varepsilon_1 \tag{11}$$

and

$$\|\mathsf{P}_{SXY} - \mathsf{P}_{XY}\mathsf{P}_{S|Y}\| \leq \varepsilon_2, \tag{12}$$

then (X,Y) is $1-2\varepsilon_1-2\varepsilon_2$-embeddable[3] *in (XS, YS).*

Even if $\varepsilon_1 = \varepsilon_2 = 0$, equations (11) and (12) do not imply that S is independent of X and Y. For example, if X and Y contain the same uniform random bit, then S can depend on this bit. However, if $\varepsilon_1 = \varepsilon_2 = 0$ the lemma is obviously true: Alice uses shared randomness to choose S according to $\mathsf{P}_{S|X=x}$ (more concretely: Alice chooses a uniform random real $\rho \in [0,1]$ and uses the smallest element s for which the cumulative distribution function $\sum_{s' \leq s} \mathsf{P}_{S|X=x}(s')$ is larger than ρ). Since Bob has the same distribution $\mathsf{P}_{S|Y=y}$ he will find the same value if he uses the same shared randomness.

In case $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, we have to overcome the following problem: $\mathsf{P}_{S|Y=y}$ is unknown to Alice (since y is unknown to Alice), and analogously $\mathsf{P}_{S|X=x}$ is unknown to Bob. The solution is to define the function $f_A : \mathcal{X} \times \mathcal{R} \to \mathcal{S}$ with the following process: Alice chooses, using shared randomness, a uniform random element s from $\mathcal{S}$ and a uniform random real number $\rho \in [0,1]$. If $\mathsf{P}_{S|X=x}(s) > \rho$ she outputs s, otherwise Alice repeats the above. The function $f_B : \mathcal{Y} \times \mathcal{R} \to \mathcal{S}$ is defined by the analogous process given y. It is easy to see that Alice outputs elements according to the distribution $\mathsf{P}_{S|X=x}$, Bob according to $\mathsf{P}_{S|Y=y}$. We further show that usually the output of f_A is equal to the output of f_B.

[3] It is understood that the embedding satisfies $(X,Y) = (X,Y)$, i.e., that the original random variables will result if from the resulting (XS, YS) the S-part is omitted.

Proof. Let $\mathcal{R} := (\mathcal{S} \times [0,1])^\infty$ be the set of infinite sequences over $\mathcal{S} \times [0,1]$. For a fixed x, y and a sequence $r := \{(s_i, \rho_i)\}_{i \geq 0}$, we define $f_A(x,r) := s_i$ if i is the smallest index for which $\mathsf{P}_{S|X=x}(s_i) > \rho_i$. Analogously, $f_B(y,r) := s_j$ if j is the smallest index with $\mathsf{P}_{S|Y=y}(s_j) > \rho_j$ and[4] $f_{AB}(x,y,r) := s_k$ if k is the smallest index with $\mathsf{P}_{S|X=xY=y}(s_k) > \rho_k$. If no such index exist the respective function is defined in an arbitrary way (this happens with probability 0).

Let $\mathsf{P}_{XYF_AF_BF_{AB}}$ be the joint distribution of $(x, y, f_A(x,r), f_B(y,r), f_{AB}(x,y,r))$ where (x,y) is chosen according to P_{XY} and r uniformly from $\mathcal{R}$. We have $\mathsf{P}_{F_{AB}|X=xY=y} = \mathsf{P}_{S|X=xY=y}$, $\mathsf{P}_{F_A|X=x} = \mathsf{P}_{S|X=x}$ and $\mathsf{P}_{F_B|Y=y} = \mathsf{P}_{S|Y=y}$, since these equalities hold conditioned on the event that the respective function accepts in round i, for any fixed i.

Further, we have $\Pr[F_A = F_{AB}|X = x, Y = y] \geq 1 - 2\|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|$: the two values F_A, F_{AB} are equal if $\rho_j < \min(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))$ for the smallest j for which $\rho_j < \max(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))$ is satisfied. This happens with probability

$$\begin{aligned}\frac{\sum_s \min(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))}{\sum_s \max(\mathsf{P}_{F_A|X=x}(s_j), \mathsf{P}_{F_{AB}|X=xY=y}(s_j))} &= \frac{1 - \|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|}{1 + \|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|} \\ &\geq 1 - 2\|\mathsf{P}_{F_A|X=x} - \mathsf{P}_{F_{AB}|X=xY=y}\|.\end{aligned}$$

This yields $\Pr[F_A = F_{AB}] \geq 1 - 2\varepsilon_1$, and analogously we get $\Pr[F_B = F_{AB}] \geq 1 - 2\varepsilon_2$, and thus $\Pr[F_A = F_B = F_{AB}] \geq 1 - 2\varepsilon_1 - 2\varepsilon_2$. This implies

$$\|\mathsf{P}_{XYSS} - \mathsf{P}_{XY}\mathsf{P}_{F_AF_B|XY}\| = \|\mathsf{P}_{XYF_{AB}F_{AB}} - \mathsf{P}_{XYF_AF_B}\| \geq 1 - 2\varepsilon_1 - 2\varepsilon_2. \qquad \square$$

In the following corollary, the input distribution is changed slightly. This makes it a bit easier to apply later.

Corollary 9. *Let distributions P_{SXY} and $\mathsf{P}_{X_0Y_0}$ be given. If*

$$\|\mathsf{P}_{SXY} - \mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X}\| \leq \varepsilon_1 \tag{13}$$

and

$$\|\mathsf{P}_{SXY} - \mathsf{P}_{X_0Y_0}\mathsf{P}_{S|Y}\| \leq \varepsilon_2, \tag{14}$$

then (X_0, Y_0) is $1 - 3\varepsilon_1 - 2\varepsilon_2$-embeddable[5] *in (XS, YS) with $(X,Y) = (X_0, Y_0)$.*

Proof. From (13) we get $\|\mathsf{P}_{XY} - \mathsf{P}_{X_0Y_0}\| \leq \varepsilon_1$. One can now find a joint distribution $\mathsf{P}_{XYX_0Y_0}$ with $\Pr[(X,Y) = (X_0,Y_0)] \geq 1 - \varepsilon_1$. The corollary now follows by applying f_A and f_B from Lemma 8. $\square$

Random variables S, T, U form a Markov chain, written $S \leftrightarrow T \leftrightarrow U$ if $\mathsf{P}_{STU} = \mathsf{P}_T\mathsf{P}_{S|T}\mathsf{P}_{U|T}$ (i.e., if given T the probability distribution of U does not depend on S). The following lemma is essentially Lemma 4.1 in Raz's paper.

[4]The use of f_{AB} in order to simplify the analysis was suggested by Anup Rao.

[5]The statement could be made symmetric (i.e., (X_0, Y_0) is $1 - 2\varepsilon_1 - 2\varepsilon_2 - \min(\varepsilon_1, \varepsilon_2)$-embeddable).

Lemma 10. *Let P_{XYST} be any distribution. If*

$$S \leftrightarrow X \leftrightarrow YT$$

and

$$XS \leftrightarrow Y \leftrightarrow T$$

then (X, Y) is 1-embeddable in (XS, YT).

Proof. Using individual (non-shared) randomness, Alice computes S according to $\mathsf{P}_{S|X=x}$ and Bob computes T according to $\mathsf{P}_{T|Y=y}$. Since

$$\mathsf{P}_{STXY} = \mathsf{P}_{XY}\mathsf{P}_{S|XY}\mathsf{P}_{T|SXY} = \mathsf{P}_{XY}\mathsf{P}_{S|X}\mathsf{P}_{T|Y} \tag{15}$$

this gives the correct (global) distribution. □

6 Embeddings for Games

Given a game $\mathfrak{G}$ and its n-fold parallel repetition, we now show that (X, Y) can be embedded into $(\widetilde{X}^n, \widetilde{Y}^n)$, where $\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{X^nY^n|W_{k+1}\wedge\cdots\wedge W_n}$.

We need the following simple fact on statistical distance.

Fact 11. *Let P_{Z_0} and P_{Z_1} be distributions over $\mathcal{Z}$. Let $\mathcal{S} \subseteq \mathcal{Z}$ be such that $\Pr[Z_0 \in \mathcal{S}] = \Pr[Z_1 \in \mathcal{S}] = \frac{1}{2}$. Then,*

$$\|\mathsf{P}_{Z_0|Z_0\in\mathcal{S}} - \mathsf{P}_{Z_1|Z_1\in\mathcal{S}}\| \leq 2\|\mathsf{P}_{Z_0} - \mathsf{P}_{Z_1}\|\,.$$

Also, we need the following statements about Markov chains.

Claim 12. *Let $\mathsf{P}_{X_0Y_0}\mathsf{P}_{X_1Y_1}$ be a distribution over $\mathcal{X}_0 \times \mathcal{Y}_0 \times \mathcal{X}_1 \times \mathcal{Y}_1$, $f : \mathcal{X}_0 \times \mathcal{X}_1 \to \mathcal{U}$ and $g : \mathcal{Y}_0 \times \mathcal{Y}_1 \to \mathcal{V}$ be arbitrary. Then,*

$$X_0X_1 \leftrightarrow X_0 f(X_0, X_1) Y_1 g(Y_0, Y_1) \leftrightarrow Y_0Y_1. \tag{16}$$

Proof. It is sufficient to show this for all possible values $x_0 \in \mathcal{X}_0$ and $y_1 \in \mathcal{Y}_1$. Let $\mathsf{P}_{\widetilde{Y}_0\widetilde{X}_1} := \mathsf{P}_{Y_0X_1|X_0=x_0Y_1=y_1} = \mathsf{P}_{Y_0|X_0=x_0}\mathsf{P}_{X_1|Y_1=y_1}$. In this case, (16) reduces to

$$\widetilde{X}_1 \leftrightarrow f(x_0, \widetilde{X}_1) g(\widetilde{Y}_0, y_1) \leftrightarrow \widetilde{Y}_0.$$

Since $\widetilde{X}_1$ and $\widetilde{Y}_0$ are independent this is obvious. □

Claim 13. *Let P_{TUV} be a distribution over $\mathcal{T} \times \mathcal{U} \times \mathcal{V}$ and W an event with*

$$T \leftrightarrow U \leftrightarrow V,$$
$$W \leftrightarrow U \leftrightarrow TV.$$

Then, for $\mathsf{P}_{\widetilde{T}\widetilde{U}\widetilde{V}} := \mathsf{P}_{TUV|W}$ we have

$$\widetilde{T} \leftrightarrow \widetilde{U} \leftrightarrow \widetilde{V}.$$

Proof.

$$\begin{aligned}
\mathsf{P}_{\widetilde{T}\widetilde{U}\widetilde{V}}(t,u,v) &= \mathsf{P}_{TUV|W}(t,u,v) \\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{TV|U=u,W}(t,v) \\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{TV|U=u}(t,v) \\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{T|U=u}(t)\mathsf{P}_{V|U=u}(v) \\
&= \mathsf{P}_{U|W}(u)\mathsf{P}_{T|U=u,W}(t)\mathsf{P}_{V|U=u,W}(v).
\end{aligned}$$

□

Lemma 14. *Let a game $\mathfrak{G}^n = (Q^n, (\mathsf{P}_{XY})^n)$, a strategy (h_a, h_b), and $k \le n$ be given. Let*

$$\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{X^nY^n|W_{k+1}\wedge\cdots\wedge W_n}$$

Then, for $1 \le j \le k$, there exists $\varepsilon_j \ge 0$ such that (X,Y) is $1-\varepsilon_j$-embeddable in $(\widetilde{X}^n, \widetilde{Y}^n)$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X,Y)$ and

$$\sum_{j=1}^{k} \varepsilon_j \le 15\sqrt{k}\sqrt{(n-k)\log(|\mathcal{A}|\,|\mathcal{B}|) + \log\Big(\frac{1}{\Pr[W_{k+1}\wedge\cdots\wedge W_n]}\Big)}. \tag{17}$$

Proof. As described in Definition 3 we consider the distribution $\mathsf{P}_{X^nY^nA^nB^nW^n}$ and the corresponding random variables. Additionally, we let $D_1, \ldots, D_k$ be uniform and independent bits. For $1 \le j \le k$ we define

$$U_j := \begin{cases} X_j & \text{if } D_j = 0 \\ Y_j & \text{otherwise} \end{cases}$$

and

$$\overline{U}_j := \begin{cases} Y_j & \text{if } D_j = 0 \\ X_j & \text{otherwise.} \end{cases}$$

Also, we set

$$T := (X_{k+1}, \ldots, X_n, Y_{k+1}, \ldots, Y_n, D^k, \overline{U}^k), \tag{18}$$

$$V := (A_{k+1}, \ldots, A_n, B_{k+1}, \ldots, B_n), \tag{19}$$

and define the event $W := W_{k+1} \wedge \cdots \wedge W_n$.

From Corollary 6 we get

$$\sum_{j=1}^{k} \Big\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W}\mathsf{P}_{U_j|T} \Big\| \le \varepsilon_{\mathrm{Tot}}\,, \tag{20}$$

where we set

$$\varepsilon_{\mathrm{Tot}} := \sqrt{k}\sqrt{(n-k)\log(|\mathcal{A}||\mathcal{B}|) + \log\Big(\frac{1}{\Pr[W]}\Big)} \tag{21}$$

(we applied Corollary 6 using $|\mathcal{V}^*| \leq |\mathcal{V}|$).

In (20), we condition on both sides on the event $D_j = 0$, which is, on both sides, a restriction on a subset which has probability $\frac{1}{2}$. Fact 11 implies

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{TU_jV|W\wedge(D_j=0)} - \mathsf{P}_{TV|W\wedge(D_j=0)}\mathsf{P}_{U_j|T} \right\| \leq 2\varepsilon_{\mathrm{Tot}}\,, \tag{22}$$

where we do not need to condition on $D_j = 0$ in $\mathsf{P}_{U_j|T}$ since this is included in the given t anyhow; in fact we can now write $\mathsf{P}_{X_j|Y_j}$ instead of $\mathsf{P}_{U_j|T}$.

For a fixed j, define the random variable

$$\begin{aligned} T^{(\setminus j)} := (&X_{k+1}, \ldots, X_n, Y_{k+1}, \ldots, Y_n, \\ &D_1, \ldots, D_{j-1}, D_{j+1}, \ldots, D_k, \\ &\overline{U}_1, \ldots, \overline{U}_{j-1}, \overline{U}_{j+1}, \ldots, \overline{U}_k). \end{aligned} \tag{23}$$

With this notation (22) is equivalent to

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{T^{(\setminus j)}X_jY_jV|W\wedge(D_j=0)} - \mathsf{P}_{T^{(\setminus j)}Y_jV|W\wedge(D_j=0)}\mathsf{P}_{X_j|Y_j} \right\| \leq 2\varepsilon_{\mathrm{Tot}}\,. \tag{24}$$

But now nothing depends on $D_j = 0$ anymore, so this also means

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{T^{(\setminus j)}X_jY_jV|W} - \mathsf{P}_{T^{(\setminus j)}Y_jV|W}\mathsf{P}_{X_j|Y_j} \right\| \leq 2\varepsilon_{\mathrm{Tot}}\,. \tag{25}$$

We set $S := (T^{(\setminus j)}, V)$ and define the probability distribution

$$\mathsf{P}_{\widetilde{S}\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{SX^nY^n|W}. \tag{26}$$

With this, (25) becomes

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{\widetilde{S}\widetilde{Y}_j}\mathsf{P}_{X_j|Y_j} \right\| \leq 2\varepsilon_{\mathrm{Tot}}\,, \tag{27}$$

or, equivalently

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{\widetilde{Y}_j}\mathsf{P}_{\widetilde{S}|\widetilde{Y}_j}\mathsf{P}_{X|Y} \right\| \leq 2\varepsilon_{\mathrm{Tot}}\,. \tag{28}$$

Lemma 5 implies

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{Y}_j} - \mathsf{P}_{Y} \right\| \leq \varepsilon_{\mathrm{Tot}}, \tag{29}$$

and thus

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{Y}_j} \right\| \leq 3\varepsilon_{\mathrm{Tot}}\,. \tag{30}$$

Symmetric reasoning yields

$$\sum_{j=1}^{k} \left\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{X}_j} \right\| \leq 3\varepsilon_{\mathrm{Tot}}\,. \tag{31}$$

From (30) and (31), Corollary 9 implies that (X, Y) is $1-\varepsilon_j$-embeddable in $(\widetilde{X}_j\widetilde{S}, \widetilde{Y}_j\widetilde{S})$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X, Y)$ and such that $\sum_{j=1}^{k} \varepsilon_j \leq 15\varepsilon_{\mathrm{Tot}}$.

We next show that

$$X^k \leftrightarrow TV \leftrightarrow Y^k. \tag{32}$$

If the bits D^k and the values $X_{k+1}, \ldots, X_n$, $Y_{k+1}, \ldots, Y_n$ are fixed, this follows immediately from Claim 12. Since it holds for all these values it must also hold overall.

From (32) we easily get

$$X^n \leftrightarrow X_jS \leftrightarrow Y^nY_jS$$
$$X^nX_jS \leftrightarrow Y_jS \leftrightarrow Y^n.$$

Claim 13 yields

$$\widetilde{X}^n \leftrightarrow \widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n\widetilde{Y}_j\widetilde{S} \tag{33}$$
$$\widetilde{X}^n\widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n. \tag{34}$$

Above we have seen that (X, Y) is embeddable in $(\widetilde{X}_j\widetilde{S}, \widetilde{Y}_j\widetilde{S})$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X, Y)$. Lemma 10 together with (33) and (34) now implies that we can 1-locally embed this in $(\widetilde{X}^n\widetilde{X}_j\widetilde{S}, \widetilde{Y}^n\widetilde{Y}_j\widetilde{S})$. Since Alice and Bob can then ignore part of the constructed information this completes the proof. □

Lemma 15. *Let a game $\mathfrak{G} = (Q, \mathsf{P}_{XY})$, its n-fold repetition $\mathfrak{G}^n$, and a strategy (h_a, h_b) for $\mathfrak{G}^n$ be given. Let indices $i_1, \ldots, i_m$ be given. Then, there exists an index i_{m+1} such that*

$$\begin{aligned}\Pr[&W_{i_{m+1}}|W_{i_1} \wedge \cdots \wedge W_{i_m}] \\ &\leq v(\mathfrak{G}) + 15\sqrt{\frac{1}{n-m}}\sqrt{m\log(|\mathcal{A}||\mathcal{B}|) + \log\Bigl(\frac{1}{\Pr[W_{i_1} \wedge \cdots \wedge W_{i_m}]}\Bigr)}.\end{aligned} \tag{35}$$

Proof. First, we can assume that the given indices i_ℓ, $1 \leq \ell \leq m$, are pairwise different (otherwise we get a stronger statement). Given this we can even assume that $i_\ell = n - \ell + 1$ by appropriately redefining the functions (h_a, h_b).

Define the distribution $\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{X^nY^n|W_{n-m+1}\wedge\cdots\wedge W_n}$. Lemma 14 implies that there exists an index j such that (X,Y) is $1-\varepsilon$-embeddable in $(\widetilde{X}^n,\widetilde{Y}^n)$ with $(\widetilde{X}_j,\widetilde{Y}_j) = (X,Y)$ and

$$\varepsilon := 15\sqrt{\frac{1}{n-m}}\sqrt{m\log(|\mathcal{A}||\mathcal{B}|) + \log\Big(\frac{1}{\Pr[W_{n-m+1}\wedge\cdots\wedge W_n]}\Big)}.$$

Consider the following strategy for $\mathfrak{G}$. On input (X,Y) Alice and Bob $1-\varepsilon$-embed this into $(\widetilde{X}^n,\widetilde{Y}^n)$ with $(\widetilde{X}_j,\widetilde{Y}_j) = (X,Y)$. Since the resulting distribution has statistical distance at most ε from $\mathsf{P}_{\widetilde{X}^n\widetilde{Y}^n}$, if they output coordinate j of $h_a(\widetilde{X}^n)$ and $h_b(\widetilde{Y}^n)$ they have probability at least $\Pr[W_j|W_{n-m+1}\wedge\cdots\wedge W_n]-\varepsilon$ to win the initial game. The shared randomness can be eliminated (see the remark after Definition 1), and thus

$$v(\mathfrak{G}) \geq \Pr[W_j|W_{n-m+1}\wedge\cdots\wedge W_n] - \varepsilon.$$

□

7 Parallel Repetition Theorem

Proof (of Theorem 4). Fix a strategy (h_a,h_b) for $\mathfrak{G}^n$. Then, repeatedly choose the index i_{m+1} for which $\Pr[W_{i_{m+1}}|W_{i_1}\wedge\cdots\wedge W_{i_m}]$ is minimized. We set $p_0 := 1$ and $p_m := \Pr[W_{i_1}\wedge\cdots\wedge W_{i_m}]$. Lemma 15 implies

$$p_{m+1} \leq p_m\cdot\left(v + 15\sqrt{\frac{1}{n-m}}\sqrt{m\log(|\mathcal{A}||\mathcal{B}|)+\log\Big(\frac{1}{p_m}\Big)}\right). \tag{36}$$

We show per induction that

$$p_m \leq \Big(\frac{1+v}{2}\Big)^m,$$

as long as $m \leq \frac{(1-v)^2(n-m)}{2700\log(|\mathcal{A}||\mathcal{B}|)}$. The statement holds for $m=0$ and we now make a step from m to $m+1$. First, we can assume that $p_m \geq \big(\frac{1+v}{2}\big)^{m+1} > \frac{1}{2}^{m+1}$, as otherwise the induction step is trivial. In this case, (36) yields

$$\begin{aligned} p_{m+1} &\leq p_m\cdot\left(v + 15\sqrt{\frac{1}{n-m}}\sqrt{m\log(|\mathcal{A}||\mathcal{B}|)+(m+1)}\right)\\ &\leq p_m\cdot\left(v + \sqrt{\frac{1}{n-m}}\sqrt{675m\log(|\mathcal{A}||\mathcal{B}|)}\right) \end{aligned} \tag{37}$$

Since we assume $m \leq \frac{(1-v)^2}{2700\log(|\mathcal{A}||\mathcal{B}|)}(n-m)$ this proves the induction step.

In total we get for $m = \frac{n(1-v)^2}{3000\log(|\mathcal{A}||\mathcal{B}|)}$

$$p_m \leq \Big(\frac{1+v}{2}\Big)^{\frac{n(1-v)^2}{3000\log(|\mathcal{A}||\mathcal{B}|)}}. \tag{38}$$

We have

$$\Big(\frac{1+v}{2}\Big)^{\frac{(1-v)^2}{3000}} = \Big(1-\frac{1-v}{2}\Big)^{\frac{(1-v)^2}{3000}} \leq 1-\frac{(1-v)^3}{6000}, \tag{39}$$

where the last inequality follows from $(1-b)^a \leq 1-ab$ which holds for all $a \in [0,1]$, $b \leq 1$. Since $\Pr[W_1 \wedge \cdots \wedge W_n] \leq p_m$, (38) and (39) imply the theorem.[6] □

8 Improving the Rate

Theorem 4 shows that the n-fold parallel repetition reduces the winning probability from $v(\mathfrak{G})$ to $(1-\Theta(1-v(\mathfrak{G}))^3)^{\Omega(\frac{n}{\log(|\mathcal{A}||\mathcal{B}|)})}$. As shown in [PRW97], the term $|\mathcal{A}| \cdot |\mathcal{B}|$ in the exponent can be reduced to the the maximum (over x, y) number of (fractional) rectangles needed to cover the 1-entries in $Q(x,y,\cdot,\cdot)$. Here, we show that it can be reduced to a quantity which is possibly smaller in some cases.

Definition 16 (Exact Fractional Product Cover)**.** *Let $Q : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be an arbitrary predicate. Two functions $f : \mathcal{A} \times \{1,\ldots,\alpha\} \to [0,1]$ and $g : \mathcal{B} \times \{1,\ldots,\alpha\} \to [0,1]$ form an* exact fractional product cover of size α *for Q if for all a, b:*

$$Q(a,b) = \sum_{i=1}^{\alpha} f(a,i)g(b,i).$$

Clearly, any partition by rectangles gives an exact fractional product cover (by defining $f(a,i)$ and $g(b,i)$ as appropriate predicates). We will prove the following strengthening of Theorem 4.

Theorem 17. *Let $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ be a game. Let α be such that for all (x,y) there exists an exact fractional product cover of size α for $Q_{x,y}(a,b) := Q(x,y,a,b)$. If $\alpha > 1$ then*

$$v(\mathfrak{G}^n) \leq \Big(1-\frac{(1-v)^3}{6000}\Big)^{\frac{n}{\log(\alpha)}}, \tag{40}$$

[6] The minimal value of the sequence defined by $p_0 := 1$ and $p_{m+1} := p_m\big(v + \sqrt{\frac{225}{n-m}}\sqrt{m\ell + \log(1/p_m)}\big)$ is indeed $\Big(1-\Theta((1-v)^3)\Big)^{\frac{n}{\ell}}$. The argument in the proof above shows that the minimal value can only be lower. On the other hand, the sequence given by $p'_0 := 1$, $p'_{m+1} := p'_m\Big(v + \sqrt{\frac{m\ell}{n}}\Big)$ is strictly smaller than the sequence $\{p_j\}_{j\geq 0}$. This sequence does not decrease anymore if $m > m' := n(1-v)^2/\ell$, and

$$\begin{aligned} p'_{m'} &= \prod_{i=0}^{m'-1}\Big(v+\sqrt{\frac{i\ell}{n}}\Big) = \exp\left(\sum_{i=0}^{m'-1} \ln\Big(v+\sqrt{\frac{i\ell}{n}}\Big)\right) \\ &\approx \exp\left(\int_0^{m'} \ln\Big(v+\sqrt{\frac{i\ell}{n}}\Big)\right) = \exp\Big((4v-1+2v^2\ln(v)-3v^2)\cdot\frac{n}{2\ell}\Big) \approx \Big(1-\frac{(1-v)^3}{2}\Big)^{\frac{3n}{4\ell}}. \end{aligned}$$

and if $\alpha = 1$ *then*

$$v(\mathfrak{G}^n) \leq \left(1 - \frac{(1-v)^2}{6000}\right)^n. \tag{41}$$

To prove Theorem 17 we first need a characterization of fractional product covers by Markov chains.

Lemma 18. *Let a distribution* $\mathsf{P}_{ABZ} = \mathsf{P}_A \mathsf{P}_B \mathsf{P}_{Z|AB}$ *be given for which there exists functions* $f(a,z) : \mathcal{A} \times \mathcal{Z} \to [0,1]$ *and* $g(b,z) : \mathcal{B} \times \mathcal{Z} \to [0,1]$ *which satisfy*

$$\mathsf{P}_{Z|A=aB=b}(z) = f(a,z) \cdot g(b,z). \tag{42}$$

Then, $A \leftrightarrow Z \leftrightarrow B$*.*

Lemma 18 could be strengthened as follows: if $\mathsf{P}_{Z|AB}$ is such that $A \leftrightarrow Z \leftrightarrow B$ for all distributions $\mathsf{P}_A \mathsf{P}_B$, then $\mathsf{P}_{Z|AB}$ is of the form (42) for some functions f and g. For completeness, we prove this in Appendix C.

Lemma 18 implies the following: if $Q : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ has a fractional product cover of size α, then there exists a random variable Z over some set $\mathcal{Z}$ given by a conditional distribution $\mathsf{P}_{Z|AB}$ with the following properties:

- For any product distribution $\mathsf{P}_{AB} = \mathsf{P}_A \mathsf{P}_B$ we have $A \leftrightarrow Z \leftrightarrow B$
- $|\{z \in \mathcal{Z} | \exists a, b : Q(a,b) = 1 \wedge \mathsf{P}_{Z|A=aB=b}(z) > 0\}| \leq \alpha$
- $Q(a,b)$ can be inferred from z.

(Note that we do not restrict the alphabet size of Z in case $Q(a,b) = 0$, which means that in this case z can be, for example, (a,b).)

Proof (of Lemma 18). We get

$$\begin{aligned}
\mathsf{P}_{A|B=bZ=z}(a) &= \frac{\mathsf{P}_{ABZ}(a,b,z)}{\mathsf{P}_{BZ}(b,z)} \\
&= \frac{\mathsf{P}_A(a)\mathsf{P}_B(b)f(a,z)g(b,z)}{\sum_{a'} \mathsf{P}_B(b)\mathsf{P}_A(a')f(a',z)g(b,z)} \\
&= \frac{\mathsf{P}_A(a)f(a,z)}{\sum_{a'} \mathsf{P}_A(a')f(a',z)} \\
&= \frac{\sum_{b'} \mathsf{P}_A(a)\mathsf{P}_B(b')f(a,z)g(b',z)}{\sum_{a',b'} \mathsf{P}_A(a')\mathsf{P}_B(b')f(a',z)g(b',z)} \\
&= \frac{\mathsf{P}_{AZ}(a,z)}{\mathsf{P}_Z(z)} \\
&= \mathsf{P}_{A|Z=z}(a),
\end{aligned}$$

and thus $\mathsf{P}_{ABZ}(a,b,z) = \mathsf{P}_Z(z)\mathsf{P}_{B|Z=z}(b)\mathsf{P}_{A|B=bZ=z}(a) = \mathsf{P}_Z(z)\mathsf{P}_{B|Z=z}(b)\mathsf{P}_{A|Z=z}(a)$, which means that $A \leftrightarrow Z \leftrightarrow B$. □

Given the characterization from Lemma 18 we can now prove Theorem 17.

Proof (of Theorem 17). We first show that Lemma 14 still holds if we replace (17) by

$$\sum_{j=1}^{k} \varepsilon_j \leq 15\sqrt{k}\sqrt{(n-k)\log(\alpha) + \log\Big(\frac{1}{\Pr[W_{k+1} \wedge \cdots \wedge W_n]}\Big)}. \tag{43}$$

For this, we define the random variables D^k, U^k, $\overline{U}^k$, and T exactly as in the proof of Lemma 14. Instead of (19) we now define

$$V := (Z_{k+1}, \ldots, Z_n), \tag{44}$$

where Z_i is obtained from (A_i, B_i, X_i, Y_i) by a channel that has alphabet size at most α in case W_i, which ensures $A_i \leftrightarrow X_i Y_i Z_i \leftrightarrow B_i$ in case A_i and B_i are independent, and for which W_i can be inferred from (X_i, Y_i, Z_i). The existence of such a random variable is ensured by Lemma 18 and the fact that for every (x, y) there exists a exact fractional product cover of size α for $Q(x, y, \cdot, \cdot)$ (the alphabet size of Z in case $Q(x, y, a, b) = 0$ is irrelevant and Z can be defined, for example, as (A, B) in this case).

From Corollary 6 we now get

$$\sum_{j=1}^{k} \Big\| \mathsf{P}_{TU_jV|W} - \mathsf{P}_{TV|W}\mathsf{P}_{U_j|T} \Big\| \leq \varepsilon_{\mathrm{Tot}}\,, \tag{45}$$

where we set

$$\varepsilon_{\mathrm{Tot}} := \sqrt{k}\sqrt{(n-k)\log(\alpha) + \log\Big(\frac{1}{\Pr[W]}\Big)}. \tag{46}$$

For a fixed j we define $T^{(\setminus j)}$ as in the proof of Lemma 14 and obtain in exactly the same way for $S := (T^{(\setminus j)}, V)$ and

$$\mathsf{P}_{\widetilde{S}\widetilde{X}^n\widetilde{Y}^n} := \mathsf{P}_{SX^nY^n|W} \tag{47}$$

the equations

$$\sum_{j=1}^{k} \Big\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{Y}_j} \Big\| \leq 3\varepsilon_{\mathrm{Tot}} \tag{48}$$

and

$$\sum_{j=1}^{k} \Big\| \mathsf{P}_{\widetilde{S}\widetilde{X}_j\widetilde{Y}_j} - \mathsf{P}_{XY}\mathsf{P}_{\widetilde{S}|\widetilde{X}_j} \Big\| \leq 3\varepsilon_{\mathrm{Tot}}\,. \tag{49}$$

Again, Corollary 9 implies that (X, Y) is $1 - \varepsilon_j$-embeddable in $(\widetilde{X}_j\widetilde{S}, \widetilde{Y}_j\widetilde{S})$ with $(\widetilde{X}_j, \widetilde{Y}_j) = (X, Y)$ and such that $\sum_{j=1}^{k} \varepsilon_j \leq 15\varepsilon_{\mathrm{Tot}}$.

Again we get

$$X^k \leftrightarrow TV \leftrightarrow Y^k, \tag{50}$$

now using the properties of the Z_i. (This is done as follows: clearly, $X^k \leftrightarrow T \leftrightarrow Y^k$, i.e. for a fixed values t for T the X^k and Y^k are independent. Now, inductively adding Z_i will not change this in any step.) Claim 13 now yields

$$\widetilde{X}^n \leftrightarrow \widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n\widetilde{Y}_j\widetilde{S} \tag{51}$$

$$\widetilde{X}^n\widetilde{X}_j\widetilde{S} \leftrightarrow \widetilde{Y}_j\widetilde{S} \leftrightarrow \widetilde{Y}^n, \tag{52}$$

and Lemma 10 completes the proof that (43) can replace (17) in Lemma 14.

From Lemma 14 where (17) is replaced by (43) we obtain (40) exactly as in the proof of Theorem 4. To get (41) we note first that in this case (43) reduces to

$$\sum_{j=1}^{k} \varepsilon_j \le 15\sqrt{k}\sqrt{\log\Big(\frac{1}{\Pr[W_{k+1} \wedge \cdots \wedge W_n]}\Big)}. \tag{53}$$

Using an analogous definition for p_m as previously, we get

$$p_{m+1} \le p_m \cdot \Big(v + 15\sqrt{\frac{1}{n-m}\log\Big(\frac{1}{p_m}\Big)}\Big). \tag{54}$$

Here, we show per induction that $p_m \le (\frac{1+v}{2})^m$ as long as $m+1 \le \frac{(n-m)(1-v)}{400}$. To make a step from m to $m+1$ we can assume $p_m \ge (\frac{1+v}{2})^{m+1}$, which implies $p_m \ge 2^{-(1-v)(m+1)}$ (since $(1-\frac{1}{2})^{1-v} \le 1 - \frac{1-v}{2}$, see inequality below), which means that

$$p_{m+1} \le p_m \cdot \Big(v + \sqrt{\frac{225(m+1)(1-v)}{n-m}}\Big), \tag{55}$$

for relevant values of m. If $m+1 \le \frac{(n-m)(1-v)}{900}$ this implies the hypothesis. We thus get for $m = \frac{n(1-v)}{1800}$

$$p_m \le \Big(\frac{1+v}{2}\Big)^{\frac{n(1-v)}{1800}}.$$

Finally, $(\frac{1+v}{2})^{(1-v)/1800} = (1 - \frac{(1-v)}{2})^{(1-v)/1800} \le 1 - \frac{(1-v)^2}{3600}$, again using $(1-b)^a \le 1 - ab$ for $a \in [0,1]$, $b \le 1$.[7] □

[7]Note that in case $p_m \le \Big(1 - \Theta((1-v)^2)\Big)^n$ equation (54) only implies $p_{m+1} \le p_m(v + 15\sqrt{-\log(1-\Theta((1-v)^2))}) = p_m(v + 15\sqrt{\Theta((1-v)^2)}) = p_m(v+1-v) = p_m$, and thus (54) cannot be used to get a significantly stronger version of the theorem.

9 No-signaling Strategies

No-signaling strategies are those where the only restriction on the response of Alice and Bob is that they do not *imply* communication.

Definition 19 (No-signaling). *A pair* (h_a, h_b) *of functions is* no-signaling *if* $h_a : \mathcal{X} \times \mathcal{Y} \times \mathcal{R} \to \mathcal{A}$ *and* $h_b : \mathcal{X} \times \mathcal{Y} \times \mathcal{R} \to \mathcal{B}$ *satisfy*

$$\Pr_R[h_a(x,y,R)] = \Pr_R[h_a(x,y',R)]$$
$$\Pr_R[h_b(x,y,R)] = \Pr_R[h_b(x',y,R)],$$

for all x, x', y, y'.

Definition 20 (No-signaling value). *The no-signaling value* $v_{ns}(\mathfrak{G})$ *of a* game $\mathfrak{G} = (\mathsf{P}_{XY}, Q)$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$ *is*

$$v_{ns}(\mathfrak{G}) := \max \Pr_{XYR}[Q(X, Y, h_a(X,Y,R), h_b(X,Y,R))],$$

where the maximum is over all no-signaling functions (h_a, h_b).

Clearly, $v(\mathfrak{G}) \leq v_{ns}(\mathfrak{G})$, since any local strategy is a no-signaling strategy. We further note that for no-signaling strategies $v_{ns}(\mathfrak{G}^2) > (v_{ns}(\mathfrak{G}))^2$ is also possible, similar to the local case (see Appendix A).

We will prove the following Theorem:

Theorem 21. *For any game* $\mathfrak{G}$ *with no-signaling value* $v_{ns} := v_{ns}(\mathfrak{G})$ *and any integer* n*:*

$$v_{ns}(\mathfrak{G}^n) \leq \Big(1 - \frac{(1-v_{ns})^2}{6400}\Big)^n. \tag{56}$$

We remark that the proof of this theorem will be much simpler than the proof of Theorem 4.

We first show that if P_{XYST} a distribution which is close to no-signalling (i.e., $\|\mathsf{P}_{XYS} - \mathsf{P}_{XY}\mathsf{P}_{S|X}\|$ and $\mathsf{P}_{XYT} - \mathsf{P}_{XY}\mathsf{P}_{S|Y}$) then there exists a no-signalling strategy which produces value which are statistically close to S and T from X and Y.

Lemma 22. *Let* P_{ST}*,* $\mathsf{P}_{S'}$ *be arbitrary distributions over* $\mathcal{S} \times \mathcal{T}$ *and* $\mathcal{S}$*,* $\mathcal{S}$ *and* $\mathcal{T}$ *finite. Then, there exists a distribution* $\mathsf{P}_{\overline{S}\,\overline{T}}$ *such that*

$$\|\mathsf{P}_{\overline{S}\,\overline{T}} - \mathsf{P}_{ST}\| \leq \|\mathsf{P}_{S'} - \mathsf{P}_S\| \tag{57}$$
$$\|\mathsf{P}_{\overline{S}} - \mathsf{P}_{S'}\| = 0 \tag{58}$$
$$\|\mathsf{P}_{\overline{T}} - \mathsf{P}_T\| = 0. \tag{59}$$

Proof. We change P_{ST} gradually to $\mathsf{P}_{\overline{S}\,\overline{T}}$ such that in the end (58) and (59) hold.

For this, fix values s_0 and s_1 with

$$\mathsf{P}_S(s_0) < \mathsf{P}_{S'}(s_0) \text{ and } \mathsf{P}_S(s_1) > \mathsf{P}_{S'}(s_1). \tag{60}$$

Then, as long as (60) holds find a value t for which $\mathsf{P}_{ST}(s_1,t) > 0$. Decrease $\mathsf{P}_{ST}(s_1,t)$ by ε and increase $\mathsf{P}_{ST}(s_0,t)$ by ε, such that afterwards $\mathsf{P}_{ST}(s_1,t) = 0$ or (60) does not hold anymore for s_0, s_1. After a finite number of repetitions (60) is not true anymore, and we start the process over again with new values for s_0, s_1. However, this can also only happen a finite number of times, thus the process terminates.

If (60) cannot be satisfied then clearly (58) holds. We never change $\mathsf{P}_T(t)$ for any t which implies (59). Finally, (57) is ensured by the fact that we only decrease $\|\mathsf{P}_{S'} - \mathsf{P}_S\|$ and do not change P_{ST} more than P_S. $\square$

Lemma 23. *Let $\mathsf{P}_{X_0Y_0}$ and P_{XYST} be arbitrary distributions. If*

$$\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{S|X} - \mathsf{P}_{XYS}\| \le \varepsilon_1, \tag{61}$$

$$\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{T|Y} - \mathsf{P}_{XYT}\| \le \varepsilon_2, \tag{62}$$

then there exists a conditional distribution $\mathsf{P}_{S'T'|X'=xY'=y}$ with $\mathsf{P}_{S'|X'=xY'=y} = \mathsf{P}_{S'|X'=x}$ and $\mathsf{P}_{T'|X'=xY'=y} = \mathsf{P}_{T'|Y'=y}$ such that

$$\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{S'T'|XY} - \mathsf{P}_{XYST}\| \le 3\varepsilon_1 + 2\varepsilon_2. \tag{63}$$

Proof. For fixed x, y we define $\mathsf{P}_{S_0T_0|X=xY=y}$ using Lemma 22 with the following properties:

$$\begin{aligned}
\|\mathsf{P}_{S_0T_0|X=xY=y} - \mathsf{P}_{ST|X=xY=y}\| &\le \|\mathsf{P}_{S|X=x} - \mathsf{P}_{S|X=xY=y}\| \\
\|\mathsf{P}_{S_0|X=xY=y} - \mathsf{P}_{S|X=x}\| &= 0 \\
\|\mathsf{P}_{T_0|X=xY=y} - \mathsf{P}_{T|X=xY=y}\| &= 0.
\end{aligned}$$

Then, again using Lemma 22 we define $\mathsf{P}_{S'T'|X=xY=y}$ such that

$$\begin{aligned}
\|\mathsf{P}_{S'T'|X=xY=y} - \mathsf{P}_{S_0T_0|X=xY=y}\| &\le \|\mathsf{P}_{T_0|Y=y} - \mathsf{P}_{T_0|X=xY=y}\| \\
\|\mathsf{P}_{T'|X=xY=y} - \mathsf{P}_{T_0|Y=y}\| &= 0 \\
\|\mathsf{P}_{S'|X=xY=y} - \mathsf{P}_{S_0|X=xY=y}\| &= 0.
\end{aligned}$$

We see that for all pairs x, y we have $\mathsf{P}_{S'|X=xY=y} = \mathsf{P}_{S'|X=x}$ and $\mathsf{P}_{T'|X=xY=y} = \mathsf{P}_{T'|Y=y}$.

We further get

$$\begin{aligned}
&\|\mathsf{P}_{X_0Y_0}\mathsf{P}_{S'T'|XY} - \mathsf{P}_{XYST}\| \\
&\quad\le \varepsilon_1 + \|\mathsf{P}_{XY}\mathsf{P}_{S'T'|XY} - \mathsf{P}_{XYST}\| \\
&\quad= \varepsilon_1 + \sum_{x,y,s,t} \left|\mathsf{P}_{XY}(x,y)\mathsf{P}_{S'T'|X=xY=y}(s,t) - \mathsf{P}_{XY}(x,y)\mathsf{P}_{ST|X=xY=y}(s,t)\right| \\
&\quad\le \varepsilon_1 + \sum_{x,y} \mathsf{P}_{XY}(x,y)\Big(\|\mathsf{P}_{S|X=x} - \mathsf{P}_{S|X=xY=y}\| + \|\mathsf{P}_{T|Y=y} - \mathsf{P}_{T|X=xY=y}\|\Big) \\
&\quad\le \varepsilon_1 + \|\mathsf{P}_{XY}\mathsf{P}_{S|X} - \mathsf{P}_{SXY}\| + \|\mathsf{P}_{XY}\mathsf{P}_{T|Y} - \mathsf{P}_{TXY}\| \\
&\quad\le 3\varepsilon_1 + 2\varepsilon_2.
\end{aligned}$$

$\square$

We can now prove a non-signaling analogue of Lemma 15.[8]

Lemma 24. *Let a game $\mathfrak{G} = (Q, \mathsf{P}_{XY})$, its n-fold repetition $\mathfrak{G}^n$, and a no-signaling strategy (h_a, h_b) for $\mathfrak{G}^n$ be given. Let indices $i_1, \ldots, i_m$ be given. Then, there exists an index i_{m+1} such that*

$$\begin{aligned}
&\Pr[W_{i_{m+1}} | W_{i_1} \wedge \cdots \wedge W_{i_m}] \\
&\qquad \leq v_{ns}(\mathfrak{G}) + 10\sqrt{\frac{1}{n-m}}\sqrt{\log\Bigl(\frac{1}{\Pr[W_{i_1} \wedge \cdots \wedge W_{i_m}]}\Bigr)}.
\end{aligned} \tag{64}$$

Proof. As in the proof of Lemma 15 we assume that $i_\ell = n - \ell + 1$ and we define $W := W_{n-m+1} \wedge \cdots \wedge W_n$. The no-signaling property of (h_a, h_b) implies $\mathsf{P}_{X^n Y^n A^n} = \mathsf{P}_{X^n} \mathsf{P}_{Y^n|X^n} \mathsf{P}_{A^n|X^n} = \mathsf{P}_{A^n X^n} \mathsf{P}_{Y^n|X^n}$. Thus, when we apply Corollary 6 on this distribution (with the event W and the random variables $T = (X^n, A^n)$ and $U_j = Y_j$) we get

$$\begin{aligned}
\sum_{j=1}^{n-m} \Bigl\| \mathsf{P}_{X^n A^n Y_j | W} - \mathsf{P}_{X^n A^n | W} \mathsf{P}_{Y_j | X_j} \Bigr\| &= \sum_{j=1}^{n-m} \Bigl\| \mathsf{P}_{T Y_j | W} - \mathsf{P}_{T|W} \mathsf{P}_{Y_j|T} \Bigr\| \\
&\leq \sqrt{(n-m)\log\Bigl(\frac{1}{\Pr[W]}\Bigr)}.
\end{aligned}$$

Taking appropriate marginals this gives

$$\sum_{j=1}^{n-m} \Bigl\| \mathsf{P}_{X_j Y_j A_j | W} - \mathsf{P}_{X_j A_j | W} \mathsf{P}_{Y_j | X_j} \Bigr\| \leq \sqrt{(n-m)\log\Bigl(\frac{1}{\Pr[W]}\Bigr)}.$$

Applying Lemma 5 once more and rearranging we get

$$\sum_{j=1}^{n-m} \Bigl\| \mathsf{P}_{X_j Y_j A_j | W} - \mathsf{P}_{XY} \mathsf{P}_{A_j | X_j W} \Bigr\| \leq 2\sqrt{(n-m)\log\Bigl(\frac{1}{\Pr[W]}\Bigr)}. \tag{65}$$

Symmetrically, we obtain

$$\sum_{j=1}^{n-m} \Bigl\| \mathsf{P}_{X_j Y_j B_j | W} - \mathsf{P}_{XY} \mathsf{P}_{B_j | Y_j W} \Bigr\| \leq 2\sqrt{(n-m)\log\Bigl(\frac{1}{\Pr[W]}\Bigr)}. \tag{66}$$

From (65), (66), and Lemma 23 we get that there exists a distribution $\mathsf{P}_{A'_j B'_j | XY}$ which can be implemented by no-signaling functions and for which

$$\sum_{j=1}^{n-m} \Bigl\| \mathsf{P}_{XY} \mathsf{P}_{A'_j B'_j | XY} - \mathsf{P}_{X_j Y_j A_j B_j | W} \Bigr\| \leq 10\sqrt{(n-m)\Bigl(\log\Bigl(\frac{1}{\Pr[W]}\Bigr)\Bigr)}.$$

[8] A previous version of the proof of this lemma contained an error, which was first noticed by Oded Regev and Ricky Rosen.

Thus, if Alice and Bob use the strategy implied by $\mathsf{P}_{A'_jB'_j|XY}$ (which is no-signaling) they can win the initial game with probability $\Pr[W_j|W]-10\sqrt{1/(n-m)}\sqrt{\log(1/\Pr[W])}$ for some j, which implies the lemma. □

Proof (of Theorem 21). Fix a no-signaling strategy (h_a, h_b) for $\mathfrak{G}$. As in the proof of Theorem 4 we repeatedly select indices i_{m+1} such that $\Pr[W_{i_{m+1}}|W_{i_1} \wedge \cdots \wedge W_{i_m}]$ is minimized. Let $p_m := \Pr[W_{i_1} \wedge \cdots \wedge W_{i_m}]$. Lemma 24 implies

$$p_{m+1} \leq p_m \cdot \left(v + 10\sqrt{\frac{1}{n-m}\log\left(\frac{1}{p_m}\right)}\right). \tag{67}$$

From (67) we obtain (56) in the same way as we obtained (41) from (54) in the proof of Theorem 17. □

Acknowledgments

I would like to thank Paul Beame, Georges Baatz, Ryan O'Donnell, Johan Håstad, Bartosz Przydatek, Anup Rao, Ran Raz, Renato Renner, Ricky Rosen, Stefano Tessaro, Stefan Wolf, and Jürg Wullschleger for helpful discussions. Paul Beame pointed out [PRW97] to me and explained how the strengthening given there can be adapted to the proof given here; he also simplified an argument in a previous version of this paper. Anup Rao simplified an argument in the proof of Lemma 8. Oded Regev and Ricky Rosen found an error in a previous version of the proof of Lemma 24. I was supported by the Swiss National Science Foundation, project no. 200020-103847/1.

References

[BBT05] Gilles Brassard, Anne Broadbent, and Alain Tapp. Quantum pseudo-telepathy. *Foundations of Physics*, 35(11):1877–1907, 2005. (quant-ph/0407221).

[Bea06] Paul Beame. Personal communication, 2006.

[BOGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 113–132, 1988.

[CCL92] Jin-yi Cai, Anne Condon, and Richard J. Lipton. On games of incomplete information. *Theoretical Computer Science*, 103(1):25–38, 1992.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969.

[Cir80] Boris S. Cirel'son. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[CSUU06] Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Strong parallel repetition theorem for quantum XOR proof systems, 2006. (quant-ph/0608146).

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., first edition, 1991. ISBN 0-471-06259-6.

[Fei91] Uriel Feige. On the success probability of the two provers in one-round proof systems. In *Proceedings of the sixth Annual Structure in Complexity Theory Conference*, pages 116–123, 1991.

[FL92] Uriel Feige and László Lóvasz. Two-prover one-round proof systems: Their power and their problems. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, pages 733–744, 1992.

[For89] Lance Fortnow. *Complexity-Theoretic Aspects of Interactive Proof Systems.* PhD thesis, Massachusetts Institute of Technology, 1989.

[FRS94] Lance Fortnow, John Rompel, and Michael Sipser. On the power of multi-prover interactive proof systems. *Theoretical Computer Science*, 134(21):545–557, 1994.

[FV02] Uriel Feige and Oleg Verbitsky. Error reduction by parallel repetition – a negative result. *Combinatorica*, 22(4):461–478, 2002.

[HR06] Thomas Holenstein and Renato Renner. On the randomness of independent experiments, 2006. (cs.IT/0608007).

[LS95] Dror Lapidot and Adi Shamir. A one-round, two-prover, zero-knowledge protocol for NP. *Combinatorica*, 15(2):204–214, 1995.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, first edition, 2000. ISBN 0-521-63503-9.

[PR94] Sandu Popescu and Daniel Rohrlich. Nonlocality as an axiom for quantum theory. *Foundations of Physics*, 24(3):379–385, March 1994. (quant-ph/9508009).

[PRW97] Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 363–372, 1997.

[Raz98] Ran Raz. A parallel reptition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[Ver94] Oleg Verbitsky. Towards the parallel repetition conjecture. In *Structure in Complexity Theory Conference*, pages 304–307, 1994.

[Wat02] John Watrous. A note on parallel repetition of two-prover quantum-interactive proofs, 2002. Manuscript.

A Non-triviality

Local case We quickly reproduce a slight modification[9] of Fortnow's example [For89] which shows that in general $v(\mathfrak{G}^2) > (v(\mathfrak{G}))^2$. The same variation was also considered by Feige and Lovász [FL92].

The game we describe is over bits (i.e., all the queries and all the responses are bits). We set $\mathsf{P}_{XY}(0,0) := \mathsf{P}_{XY}(0,1) := \mathsf{P}_{XY}(1,0) := \frac{1}{3}$, and define

$$Q(x,y,a,b) := \big((x \vee a) \neq (y \vee b)\big). \tag{68}$$

This can be described in words: Alice and Bob receive a bit, and at least one of these bits is 0. If both players receive 0, exactly one player must respond with 1. If one of the players receives 1, the other must respond with 0.

We first show that for this game $v = \frac{2}{3}$. Clearly, $v \geq \frac{2}{3}$ (e.g., both players always answer 0). To show $v \leq \frac{2}{3}$ we check all deterministic strategies. If both players reply 0 on query 0, this fails in case $x = y = 0$ (and thus with probability $\frac{1}{3}$). If one player, w.l.o.g. Alice, answers 0 with 1 the players fail in case $x = 0$ and $y = 1$.

[9] Fortnow also lets the referee choose $x = y = 1$ with some probability, in which case the players cannot win the game.

If this game is repeated twice in parallel, setting $(a_1, a_2) := (x_2, x_1)$, $(b_1, b_2) := (y_2, y_1)$ also wins with probability $\frac{2}{3}$. One can check this as follows: for every fixed query (x_1, y_1) answering with (x_2, y_2) wins the first subgame with probability $\frac{2}{3}$. Moreover, with this strategy $Q(x_1, y_1, a_1, b_1) \equiv Q(x_2, y_2, a_2, b_2)$ which implies the claim.

No-signaling case We now show that for the above game

$$v(\mathfrak{G}) = v(\mathfrak{G}^2) = v_{ns}(\mathfrak{G}) = v_{ns}(\mathfrak{G}^2). \tag{69}$$

Previously, it was known that quantum strategies do not help Alice and Bob to win this game [Wat02] (in both the single instance case and where two parallel instances are used).

To show (69) it is sufficient to show that that $v(\mathfrak{G}) = v_{ns}(\mathfrak{G})$ (since $v_{ns}(\mathfrak{G}^2) \leq v_{ns}(\mathfrak{G})$ and $v_{ns}(\mathfrak{G}^2) \geq v(\mathfrak{G}^2) = v(\mathfrak{G})$ are already known). There are two ways to see that $v(\mathfrak{G}) = v_{ns}(\mathfrak{G})$. First, one can notice that the joint probability of Alice's and Bob's reply only matters if $x = y = 0$; i.e., only for one query. In such a case one can always get a local strategy which is as good as a given no-signaling strategy. Alternatively, let p be the probability that Alice replies 0 on query 0 and q be the probability that Bob replies 0 on query 0. In this case, the players win with probability at most p on query $(x, y) = (0, 1)$, with probability at most q on query $(1, 0)$, and with probability at most $(1 - p) + (1 - q)$ on query $(0, 0)$, which gives an overall winnig probability of at most $\frac{2}{3}$.

B A Lemma on Relative Entropy

This following lemma is well known, but we do not know of a standard reference containing a proof of it.

Lemma 25. *Let $\mathsf{P}_{U^k} = \mathsf{P}_{U_1} \dots \mathsf{P}_{U_k}$ and P_{V^k} be distributions over the same set. Then,*

$$\sum_{j=1}^{k} D(\mathsf{P}_{V_j} \| \mathsf{P}_{U_j}) \leq D(\mathsf{P}_{V^k} \| \mathsf{P}_{U^k}).$$

Proof. We prove the bipartite case; the general case follows by induction.

$$\begin{aligned}
&D(\mathsf{P}_{V_1 V_2} \| \mathsf{P}_{U_1} \mathsf{P}_{U_2}) \\
&\quad = \sum_{(u_1, u_2)} \mathsf{P}_{V_1 V_2}(u_1, u_2) \log\Big(\frac{\mathsf{P}_{V_1 V_2}(u_1, u_2)}{\mathsf{P}_{U_1}(u_1) \mathsf{P}_{U_2}(u_2)}\Big) \\
&\quad = \sum_{(u_1, u_2)} \mathsf{P}_{V_1 V_2}(u_1, u_2) \log\Big(\frac{\mathsf{P}_{V_1}(u_1)}{\mathsf{P}_{U_1}(u_1)}\Big) + \\
&\qquad\qquad + \sum_{(u_1, u_2)} \mathsf{P}_{V_1 V_2}(u_1, u_2) \log\Big(\frac{\mathsf{P}_{V_2 | V_1 = u_1}(u_2)}{\mathsf{P}_{U_2}(u_2)}\Big)
\end{aligned}$$

$$
\begin{aligned}
&= D(\mathsf{P}_{V_1}\|\mathsf{P}_{U_1}) + \sum_{(u_1,u_2)} \mathsf{P}_{V_1V_2}(u_1,u_2)\log\Big(\frac{\mathsf{P}_{V_2}(u_2)}{\mathsf{P}_{U_2}(u_2)}\cdot\frac{\mathsf{P}_{V_1V_2}(u_1,u_2)}{\mathsf{P}_{V_1}(u_1)\mathsf{P}_{V_2}(u_2)}\Big)\\
&= D(\mathsf{P}_{V_1}\|\mathsf{P}_{U_1}) + D(\mathsf{P}_{V_2}\|\mathsf{P}_{U_2}) + \sum_{(u_1,u_2)} \mathsf{P}_{V_1V_2}(u_1,u_2)\log\Big(\frac{\mathsf{P}_{V_1V_2}(u_1,u_2)}{\mathsf{P}_{V_1}(u_1)\mathsf{P}_{V_2}(u_2)}\Big)\\
&\geq D(\mathsf{P}_{V_1}\|\mathsf{P}_{U_1}) + D(\mathsf{P}_{V_2}\|\mathsf{P}_{U_2}),
\end{aligned}
$$

where the last inequality follows from the log-sum inequality (see [CT91, Theorem 2.7.1]). □

C Converse of Lemma 18

In this appendix we show that Lemma 18 can be strengthened to get an "if and only if" condition.

Lemma 26. *Let a conditional distribution $\mathsf{P}_{Z|AB}$ be given. If, for all product distributions $\mathsf{P}_{AB} = \mathsf{P}_A\mathsf{P}_B$ the Markov condition $A \leftrightarrow Z \leftrightarrow B$ is satisfied, then there exists functions $f(a,z) : \mathcal{A}\times\mathcal{Z} \to [0,1]$ and $g(b,z) : \mathcal{B}\times\mathcal{Z}\to[0,1]$ such that*

$$
\mathsf{P}_{Z|A=aB=b} = f(a,z)\cdot g(b,z). \tag{70}
$$

Proof. Fix an arbitrary z throughout the proof, and consider arbitrary elements $a, a' \in \mathcal{A}$ and $b, b' \in \mathcal{B}$. We set $\mathsf{P}_A(a) = \mathsf{P}_A(a') = \frac{1}{2}$ and $\mathsf{P}_B(b) = \mathsf{P}_B(b') = \frac{1}{2}$. The Markov condition implies

$$
\mathsf{P}_{A|Z=zB=b}(a) = \mathsf{P}_{A|Z=zB=b'}(a)
$$

which is equivalent to

$$
\frac{\mathsf{P}_{ABZ}(a,b,z)}{\mathsf{P}_{ABZ}(a,b,z)+\mathsf{P}_{ABZ}(a',b,z)} = \frac{\mathsf{P}_{ABZ}(a,b',z)}{\mathsf{P}_{ABZ}(a,b',z)+\mathsf{P}_{ABZ}(a',b',z)}
$$

or (because of our choice of P_{AB})

$$
\frac{\mathsf{P}_{Z|A=a,B=b}(z)}{\mathsf{P}_{Z|A=a,B=b}(z)+\mathsf{P}_{Z|A=a',B=b}(z)} = \frac{\mathsf{P}_{Z|A=a,B=b'}(z)}{\mathsf{P}_{Z|A=a,B=b'}(z)+\mathsf{P}_{Z|A=a',B=b'}(z)}.
$$

Analogously one gets (by swapping the roles of a and a')

$$
\frac{\mathsf{P}_{Z|A=a',B=b}(z)}{\mathsf{P}_{Z|A=a,B=b}(z)+\mathsf{P}_{Z|A=a',B=b}(z)} = \frac{\mathsf{P}_{Z|A=a',B=b'}(z)}{\mathsf{P}_{Z|A=a,B=b'}(z)+\mathsf{P}_{Z|A=a',B=b'}(z)}.
$$

Together, this implies

$$
\mathsf{P}_{Z|A=a,B=b}(z)\mathsf{P}_{Z|A=a',B=b'}(z) = \mathsf{P}_{Z|A=a,B=b'}(z)\mathsf{P}_{Z|A=a',B=b}(z). \tag{71}
$$

Fix now z and let $f(\cdot,z)$ and $g(\cdot,z)$ be functions onto $[0,1]$ which satisfy

$$f(a,z)g(b,z) \geq \mathsf{P}_{Z|A=a,B=b}(z) \tag{72}$$

for all (a,b) and for which the number of pairs (a,b) for which $f(a,z)g(b,z) > \mathsf{P}_{Z|A=aB=b}(z)$ is minimal (such functions exist since $f(a,z) = g(b,z) = 1$ satisfy (72)). We assume this number is non-zero and obtain a contradiction. For this, let (a_1,b_1) be a pair for which $f(a_1,z)g(b_1,z) > 0$ and for which the quotient $\mathsf{P}_{Z|A=a_1,B=b_1}(z)/f(a_1,z)g(b_1,z) < 1$ is minimal.

We define

$$f'(a,z) := \begin{cases} f(a,z) & \text{if } a \neq a_1, \\ \frac{\mathsf{P}_{Z|A=a,B=b_1}(z)}{g(b_1,z)} & \text{if } a = a_1 \end{cases}$$

and

$$g'(b,z) := \begin{cases} g(b,z) & \text{if } b \neq b_1, \\ \frac{\mathsf{P}_{Z|A=a_1,B=b}(z)}{f(a_1,z)} & \text{if } b = b_1. \end{cases}$$

We note that f' and g' cannot take values larger than 1. For example, $f'(a_1,z) > 1$ implies $\mathsf{P}_{Z|A=a_1,B=b_1}(z) > g(b_1,z) \geq f(a_1,z)g(b_1,z)$, which contradicts (72). We further claim that either the pair (f',g) or (f,g') still satisfies (72). Otherwise, there are values a_2 and b_2 such that

$$f'(a_1,z)g(b_2,z) = \frac{\mathsf{P}_{Z|A=a_1,B=b_1}(z)}{g(b_1,z)} g(b_2,z) > \mathsf{P}_{Z|A=a_1B=b_2}(z)$$

and

$$f(a_2,z)g'(b_1,z) = f(a_2,z)\frac{\mathsf{P}_{Z|A=a_1,B=b_1}(z)}{f(a_1,z)} > \mathsf{P}_{Z|A=a_2B=b_1}(z),$$

which implies

$$\frac{\mathsf{P}_{Z|A=a_1,B=b_1}(z)}{g(b_1,z)}\frac{\mathsf{P}_{Z|A=a_1,B=b_1}(z)}{f(a_1,z)} g(b_2,z)f(a_2,z) > \mathsf{P}_{Z|A=a_2B=b_1}(z)\mathsf{P}_{Z|A=a_1B=b_2}(z),$$

and using (71)

$$\frac{\mathsf{P}_{Z|A=a_1,B=b_1}(z)}{f(a_1,z)g(b_1,z)} > \frac{\mathsf{P}_{Z|A=a_2,B=b_2}(z)}{g(b_2,z)f(a_2,z)}$$

contradicting the way we chose (a_1,b_1). Thus, either (f',g) or (f,g') still satisfies (72) and since the respective version of (72) is satisfied with equality for at least one more pair (a,b) (namely for (a_1,b_1)) than for which (f,g) satisfies it, we get a contradiction. □