Parallel Repetition in Projection Games and a Concentration Bound

Anup Rao*
University of Washington
anuprao@cs.washington.edu

September 17, 2010

Abstract

A two player game is played by cooperating players who are not allowed to communicate. A referee asks the players questions sampled from some known distribution and decides whether they win or not based on a known predicate of the questions and the players' answers. The parallel repetition of the game is the game in which the referee samples n independent pairs of questions and sends the corresponding questions to the players simultaneously. If the players cannot win the original game with probability better than $(1-\epsilon)$, what's the best they can do in the repeated game?

We improve earlier results of [Raz98] and [Hol07], who showed that the players cannot win all copies in the repeated game with probability better than $(1-\epsilon/2)^{\Omega(n\epsilon^2/c)}$ (here c is the length of the answers in the game), in the following ways:

- We show that the probability of winning all copies is $(1-\epsilon/2)^{\Omega(\epsilon n)}$ as long as the game is a "projection game", the type of game most commonly used in hardness of approximation results.
- We prove a concentration bound for parallel repetition (of general games) showing that for any constant $0 < \delta < \epsilon$, the probability that the players win a $(1-\epsilon+\delta)$ fraction of the games in the parallel repetition is at most $\exp\left(-\Omega_\epsilon(\delta^3 n/c)\right)$ (here the constant may depend on ϵ). Our result has applications to testing Bell Inequalities, since it implies that the parallel repetition of the CHSH game can be used to get an experiment that has a very large classical versus quantum gap.

Our first bound is independent of the answer length and has a better dependence on ϵ. By the recent work of Raz [Raz08], this bound is tight. Our bound gives a generic way to improve the soundness of a Probabilistically Checkable Proof (PCP), in a way that is independent of the answer length of the PCP. Using it, for every k, one can convert any q query PCP with answer length c, size sc and soundness $(1-\epsilon)$ into a 2 query PCP with answer length ck, size $O(ck(2s)^k)$ and soundness $(1-\epsilon/2q)^{\Omega(\epsilon k/q)}$.

Another consequence of our bound is that the Unique Games Conjecture of Khot [Kho02] can now be shown to be equivalent to the following a priori weaker conjecture:

Unique Games Conjecture There is an unbounded increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that for every $\epsilon > 0$, there exists an alphabet size $M(\epsilon)$ for which it is NP-hard to distinguish a Unique Game with alphabet size M in which a $1-\epsilon^2$ fraction of the constraints can be satisfied from one in which a $1-\epsilon f(1/\epsilon)$ fraction of the constraints can be satisfied.

*Supported by the National Science Foundation under agreement No. CCR-0324906.

1 Introduction

We study two player games ($\mathcal{G}$) defined by a publicly known distribution on questions (X, Y) and a publicly known predicate V. A referee administers the game by sampling two questions (x, y) and revealing a question to each of the players. The players respond with answers $(a(x), b(y))$. The game is won if and only if the predicate $V(x, y, a, b)$ is satisfied. The players are not allowed to communicate during the game. The *value* of the game, usually denoted by $(1 - \epsilon)$ in this paper, is the maximum probability of success that the players can achieve.

Any one round two prover interactive proof (as introduced by [BGKW88]) can be viewed as such a two player game between the verifier and the two players. This proof system turns out to be powerful enough to capture all of non-deterministic exponential time (NEXP), with an exponentially small error. Such games also arise in cryptographic applications [BGKW88, BGKW89, DFK$^+$92, LS95], hardness of approximation results [FGL$^+$91, ALM$^+$98, FL92, LY93], and have been used to prove direct product theorems for communication complexity [PRW97].

Given any game $\mathcal{G}$, the n-fold parallel repetition of the game $\mathcal{G}^n$ is the game in which the referee samples n independent questions $(X_1, Y_1), \ldots, (X_n, Y_n)$, each distributed according to the distribution of $\mathcal{G}$, and sends all the x questions to the first player and all the y questions to the second player. The players then each respond with n answers that are functions of *all* the questions that each receives, and the referee decides that they win if and only if they win in each of the n coordinates. For each player, the i'th answer may depend on the question asked in some other coordinate. Can we bound the value of $\mathcal{G}^n$ in terms of the value of $\mathcal{G}$?

1.1 Earlier Work

The first bound on the value of $\mathcal{G}^n$ obtained by Verbitsky [Ver94], who showed that it must tend to 0 as n tends to infinity. This was followed by a much stronger bound due to Raz, involving the *answer length*, of the game. The answer length is c if the set of possible answers that the players may give is bounded by 2^c.

Theorem 1 (Raz [Raz98])**.** *There is a universal constant $\alpha > 0$ such that for every game $\mathcal{G}$ with value $1 - \epsilon$, the value of $\mathcal{G}^n$ is at most $(1 - \epsilon/2)^{\alpha\epsilon^{32}n/c}$.*

The dependence on the answer length c was shown to be almost optimal by Feige and Verbitsky [FV02]. Raz's proof was subsequently strengthened and simplified by Holenstein:

Theorem 2 (Holenstein [Hol07])**.** *There is a universal constant $\alpha > 0$ such that for every game $\mathcal{G}$ with value $1 - \epsilon$, the value of $\mathcal{G}^n$ is at most $(1 - \epsilon/2)^{\alpha\epsilon^2 n/c}$.*

We note that $(1 - \epsilon/2)$ in the above expression can actually be replaced by $(1 - \epsilon\beta)$ for any constant $\beta < 1$. In addition to our own results, this paper contains a proof of Theorem 2.

1.2 The Connection to Probabilistically Checkable Proofs and Hardness of Approximation

A *projection game* is a game in which the predicate V has a special kind of structure — every pair (x, y) defines a function f_{xy} and the predicate V is satisfied only when $f_{xy}(b) = a$. We note that the definition for projection games we use in this paper is slightly weaker than the one typically used, since we allow $V_{xy}(a, b)$ to be false even when $f_{xy}(b) = a$. This weaker form is useful for the

discussion below. If the game is such that f_{xy} is a permutation for every xy, then the game is called a *unique game*.

Both projection and unique games have played an important role in the study of approximation algorithms. Most hardness of approximation results are proved by showing that if one can given an algorithm for the hard problem under study, than this algorithm can be used to solve the *Label Cover* problem. An instance of this problem is a bipartite graph and a predicate V_{xy} associated with every edge xy in the graph. The problem is to estimate

$$\max_{a,b} \Pr_{\text{edge } (x,y)}[V(x,y,a(x),b(y))],$$

where here the maximum is taken over all functions a, b mapping the vertices to elements of the answer set, and the probability is over a uniformly random edge from the graph.

Every such instance L is associated with a two player game $\mathcal{G}_L$ in the natural way: the distribution on questions is the one induced by picking a uniformly random edge, and the predicate is the same as in the instance. The problem of finding the optimal a, b for L is the same as the problem of finding the best strategy for the game $\mathcal{G}_L$. Similarly, the parallel repetition of the game $\mathcal{G}_L^n$ is associated with another instance of Label Cover L^n, where again the value of $\mathcal{G}_L^n$ is the maximum fraction of edges of L^n that can be satisfied by any labeling. Further, the property of being a unique or projection game is preserved under parallel repetition.

The above games have played a key role in the development of Probabilistically Checkable Proofs (PCPs). The PCP theorem of Arora et al. [AS98, ALM+98] shows that there exists a constant $\epsilon_0 > 0$ and a constant alphabet size 2^c, for which we can construct a 2-query PCP with completeness 1 and soundness $1 - \epsilon_0$. Given any q query PCP, define the Label Cover instance where every left vertex corresponds to a particular q-tuple of questions that can be asked by the verifier. A valid label for the vertex is a q-tuple of answers for the corresponding questions. Every right vertex corresponds to a single query to the PCP, with valid labels being single answers. xy is an edge in the graph if and only if y corresponds to a single question from the q-tuple of x. The corresponding constraint is satisfied if and only if the q-tuple of answers would satisfy the verifier, and the label of y is consistent with the corresponding label of x. If the proof is valid, then every constraint can be satisfied. Any assignment to the Label Cover problem gives a PCP proof just by taking the answers to the vertices on the right, and any PCP proof gives an assignment to the Label Cover instance in the natural way. It turns out that if the best assignment to the Label Cover instance satisfies a $(1 - \epsilon)$ fraction of the edges, and the best proof in the PCP has a probability of success of $1 - \gamma$, then $q\epsilon \geq \gamma \geq \epsilon$. This shows that any q-query PCP with size s and soundness $1 - \epsilon$ can be converted into a 2-query projection PCP (a projection PCP is a PCP where the verifier accepts only when a projection constraint is satisfied) of size $s + s^q$ with soundness $1 - \gamma/q$.

Thus the results of Arora et al. can be viewed as proving that there is a constant $\epsilon_0 > 0$ such that given any language in NP, the instances of this problem can be converted to instances of Label Cover in polynomial time, so that the Label Cover instance has value 1 if the input instance belongs to the language, else it has value at most $1 - \epsilon_0$. Thus an algorithm that can distinguish Label Cover instances with value 1 from instances with value $1 - \epsilon_0$ can be used to solve all problems in NP.

Parallel repetition theorems for projection games can then be used to improve the above result. We can take the instance L obtained from the discussion above and encode it as the parallel repetition L^n. If L has value 1, then L^n clearly still has value 1. On the other hand, if L has value at most $1 - \epsilon_0$, by the parallel repetition theorem, n can be chosen to be large enough so that L^n

has value at most ϵ. So any algorithm that can distinguish Label Cover instances with value 1 from instances with value ϵ can be used to distinguish instances with value 1 from instances with value $1-\epsilon_0$ and so can be used to solve every problem in NP.

There is a simple algorithm to check if a Unique Games instance has value 1 or not — cycle over all choices for an assignment to a single vertex v in the graph and check that the induced unique assignment to the rest of the connected component satisfies all constraints, and repeat this for all connected components[1]. Still, we may hope that the following conjecture (due to Khot [Kho02]) is true:

Conjecture 3 (Unique Games Conjecture)**.** *For every ϵ, there exists an answer length $c(\epsilon)$ for which it is NP-hard to distinguish instances of Unique Games with answer length c that have value at least $1-\epsilon$ from instances that have value at most ϵ.*

Several tight or almost tight hardness results have been proved assuming the Unique Games Conjecture, including for Max 2-Lin [Kho02], Vertex Cover [KR03], Max-Cut [Kho02, KKMO04, MOO05], Approximate Coloring [DMR06], Sparsest Cut [CKK$^+$06, KV05] and Max 2-Sat [Aus07]. Thus the question of whether or not the conjecture is true is of considerable interest. On the other hand, approximation algorithms [Tre05, CMM06, CMM06, GT06] have been designed to approximate the value of a Unique Game. For example, given a unique games instance with value $1-\epsilon$, an algorithm due to Charikar, Makarychev and Makarychev [CMM06] can find an assignment with value $1-O(\sqrt{\epsilon c})$. This implies that the answer length $c(\epsilon)$ in the Unique Games Conjecture *must* be larger than $\Omega(1/\epsilon)$ if the conjecture is to hold and P is different from NP.

We might have hoped that we could use the parallel repetition thereoms of Raz or Holenstein to reduce the task of proving the conjecture to the task of proving it for a much smaller gap, just as we did above for the case of Label Cover, and then try and prove the conjecture for that small gap. However, the bounds of Raz and Holenstein are problematic for this purpose. If $\epsilon > \delta$ and we try to apply Holenstein's theorem to increase the gap between the $1-\delta$ instances and the $1-\epsilon$ instances, n repetitions maps these instances to ones with value $(1-\delta)^n$ and $(1-\epsilon^3)^{\alpha n/c}$ respectively, which is a big gap only if $\delta \ll \alpha\epsilon^3/c$. A conjecture with this kind of gap cannot hold, since the algorithm of Charikar et al. shows that if δ, ϵ, c satisfy this constraint, we can distinguish $(1-\delta)$ instances from $(1-\epsilon)$ instances in polynomial time.

1.3 Parallel Repetition and Bell Inequalities

Two player games also show up in the context of testing so called *Bell Inequalities* [Bel64] to confirm the existence of quantum entanglement. The idea is to consider games where two players who have access to entangled qubits can achieve a much higher success probability than two classical players can. Perhaps the most famous example of a game where such a gap exists is the CHSH game [CHSH69]. Here the verifier sends the players random bits (x, y) and receives one bit answers $(a(x), b(y))$. The players win when $a \oplus b = x \wedge y$. Two classical players cannot win with probability better than 0.75, but it can be shown that two players sharing entangled qubits can win with probability close to 0.85.

For the purpose of testing Bell Inequalities, it is important to be able to come up with games that have a big gap between the success probability of classical players and the success probability

[1]We need to cycle over assignments since our definition of a Unique Game allows for a constraint to be unsatisfied even if the corresponding bijection is satisfied.

of entangled players, and this seems to be an issue that has warranted a significant amount of effort [BCH+02, Gil03].

This motivates proving a *concentration bound* for parallel repetition. If we could prove that two players cannot hope to win more than the expected fraction of games in the parallel repetition, we would get a simple way to construct games with a large quantum vs classical gap. We can take the CHSH game (any game with a small classical vs quantum gap would work) and consider its n-fold parallel repetition. We say that the players win the repeated game as long as they win in 0.8 fraction of the coordinates. The concentration bound would imply that if the players were classical, they can win this game with a very small probability. On the other hand, the Chernoff bound shows that the obvious quantum strategy (play each game in the repetition independently) is sure to win the game with all but exponentially small probability.

1.4 Our Results

- We prove an essentially tight bound on the value of the parallel repetition in the case that the original game is a projection game.

  Theorem 4 (Parallel Repetition in Projection Games)**.** *There is a universal constant $\alpha > 0$ such that if $\mathcal{G}$ is a projection game with value at most $1 - \epsilon$, the value of $\mathcal{G}^n$ is at most $(1 - \epsilon/2)^{\alpha\epsilon n}$.*

  This improves on the earlier bounds in two ways: the dependence on ϵ is better, and the bound is independent of the answer length. As we discussed in the introduction, the work of Raz [Raz08] shows that our bound is tight up to the constant α.

  Every unique game is also a projection game, so this theorem can be used to amplify the gap between unique games instances. We obtain the the Unique Games Conjecture is actually equivalent to the following (a priori weaker) conjecture:

  Conjecture 5 (Unique Games Conjecture Restated)**.** *There exists an unbounded increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that for every $\epsilon > 0$, there is an answer length $c(\epsilon)$ for which it is NP-hard to distinguish instances of unique games with answer length c that have value $1 - \epsilon^2$ from instances that have value $1 - \epsilon f(1/\epsilon)$.*

  We sketch the proof of this equivalence in Section 7.

  As discussed in the introduction, every q-query PCP P of size s with soundness $1 - \epsilon$ can be converted into a 2-query projection PCP whose soundness can then be amplified to get a new 2-query PCP P^n. Theorem 4 gives a bound for the soundness of the resulting 2-query PCP as described in the following corollary:

  Corollary 6 (Soundness Reduction for General PCPs)**.** *There is a universal constant $\alpha > 0$ such that if P is a q-query PCP of size sc with alphabet length c, completeness $1 - \beta$ and soundness $1 - \epsilon$, then P^n is a 2-query PCP of size $c(s^n + s^{qn})$ with alphabet length cqn, completeness $(1 - \beta)^n$ and soundness $(1 - \epsilon/2q)^{\alpha\epsilon n/q}$.*

- We prove a concentration bound for parallel repetition:

Theorem 7 (Concentration in General Games)**.** *There is a universal constant $\alpha > 0$ such that for every game $\mathcal{G}$ with value $1-\epsilon$ and answer set size 2^c and every $\delta > 0$, the probability that the players can win more than a $1-\epsilon+\delta$ fraction of the games in the n-fold parallel repetition is bounded by*

$$2\left(1-\frac{\delta/2}{1-\epsilon+3\delta/4}\right)^r,$$

where $r = \frac{\alpha\delta^2 n}{c-\log(1-\epsilon+\delta/4)}$.

This theorem shows that the parallel repetition of the CHSH game gives a game with a large classical vs quantum gap.

1.5 Techniques

Our proofs build on the work of Raz and Holenstein. In this section we shall be vague (and slightly inaccurate) in order to convey what is new about our work without revealing too many technical details.

Fix a strategy for $\mathcal{G}^n$. We use the notation $X^n = X_1, \ldots, X_n$ and $Y^n = Y_1, \ldots, Y_n$ to denote the questions that are asked to the players in $\mathcal{G}^n$. It turns out that the heart of all the earlier proofs (and our own) is a lemma of the following type:

Informal Lemma 8. *Let $S \subset [n]$ be any set of small size k and W_S denote the event that the players win the games corresponding to the coordinates in S. Then, if* $\Pr[W_S]$ *is* large enough *there exists an index $i \notin S$ such that the probability that the players win the i'th coordinate conditioned on W_S is at most $1-\epsilon/2$.*

Here we need $\Pr[W_S]$ to be larger than some function of ϵ, n, k and the answer length c. Once we have this kind of lemma, it is not too hard to show that the players cannot win $\mathcal{G}^n$ with a high probability, and we leave this to the formal parts of the paper.

The lemma is proved via a reduction — we can show that if the lemma is false, i.e. if there exists a small set S and a dense event W_S for which the lemma is false, we can find a strategy for $\mathcal{G}$ that wins with probability larger than $1-\epsilon$, which is a contradiction. Suppose there exists such a set S for which W_S is dense. Then the players decide on an index i ahead of time. When asked the questions (X, Y), the players place these questions in the i'th coordinate and use shared randomness to generate $n-1$ other pairs of questions (X_j, Y_j) such the joint distribution of the questions they end up with is $\epsilon/2$ close to $\left(X^nY^n \middle| W_S\right)$ in statistical distance. Since the lemma is assumed to be false, the players can then use the i'th coordinate answers dictated by the strategy for $\mathcal{G}^n$ to win $\mathcal{G}$ with probability more than $1-\epsilon$.

The questions are actually generated in two steps. In the first step, the players *simultaneously* sample two random variables R, A, i.e. they end up with the same sample for these random variables with high probability. The random variable A is just the answers of the first player in the coordinates in S. The random variable R contains at least one question from every pair (X_j, Y_j), and both questions from the pairs corresponding to the coordinates in S. These properties allow us to argue that for every r, a, $\left(X^nY^n \middle| (R, A) = (r, a) \wedge W_S\right)$ is a product distribution. This means that once the players have agreed on the sample for R, A, they can use independent randomness to sample the rest of the questions conditioned on the information they have, and end up with a distribution on questions that is close to $\left(X^nY^n \middle| W_S\right)$.

It turns out that $\Pr[W_S \wedge A = a | R = r]$ needs to be *large enough* for typical fixings of $R = r$ for this argument to go through. Raz and Holenstein argue that this quantity is large, just by counting. They argue that if the answer length is c bits, $\Pr[W_S \wedge (R, A) = (r, a)] / \Pr[W_S \wedge R = r]$ is typically at least 2^{-ck}, since there are at most 2^{ck} possible ways to set the random variable A if the answer length is c. In our work, we get a stronger bound by observing that in the case of a projection game, the players cannot be using *all* of their answers equally often.

For simplicity, let us assume that the game is unique. Then note that for every fixing of $R = r$, the bijection between the answers of the players in the coordinates of S is determined, but the answers are now two independent random variables. It is not too hard to show that if two independent random variables satisfy some bijection with probability γ, there must exist a set of size $100/\gamma$ such that the probability that the bijection is satisfied and the first random variable does not land in this set is less than $\gamma/100$ (simply take the the set to be the elements of weight at least $\gamma/100$). The argument also works in the case that the constraints are projections instead of bijections.

So we can argue that for every fixing of $R = r$, there is a small set of *heavy answers* that the players must be using. This argument lets us get a lowerbound on $\Pr[W_S \wedge (R, A) = (r, a)]$ that is independent of the answer length.

To prove the concentration bound, we first observe that the lemma above can be generalized slightly in the following way:

Informal Lemma 9. *Let $S \subset [n]$ be any set of small size k and E be* any *event that is determined by what happens in the games of S. Then, if* $\Pr[E]$ *is* large enough*,* most *indices $i \notin S$ are such that the probability that the players win the i'th coordinate conditioned on E is at most* $1 - \epsilon/2$.

Once we have this lemma, we can show that if the referee samples a small fraction of the coordinates uniformly at random and checks that the players have won in those coordinates, his count of how many games the players have won in the random sample behaves like a supermartingale; conditioned on the result of his sampling so far, the outcome at the next random coordinate is biased towards losing. This allows us to bound the probability that the referee sees a larger fraction of wins than he should. On the other hand, the Chernoff bound gives that with high probability, the referee's experiment gives a good estimate for the fraction of games that the players won. These arguments allow us to bound the probability that the players win a large fraction of the games.

2 Preliminaries

2.1 Notation

We use calligraphic letters to denote sets, capital letters to denote random variables and small letters to denote instantiations of random variables/elements of sets. We shall use the same letter to denote objects of this type that are related to each other. For example, we shall use X to denote a random variable taking values in the set $\mathcal{X}$ and x to denote an instantiation of that random variable.

In this paper we shall often need to start with some probability space and modify it in certain ways. We explain our notation with the help of some examples. If A, B, C are random variables in some probability space taking values in $\mathcal{A}, \mathcal{B}, \mathcal{C}$, then:

- $A'B'C' \stackrel{def}{=} \{A\}\{B\}\{C\}$ defines a new probability space in which A', B', C' take values in $\mathcal{A}, \mathcal{B}, \mathcal{C}$ such that

$$\Pr[A' = a \wedge B' = b \wedge C' = c] \stackrel{def}{=} \Pr[A = a] \Pr[B = b] \Pr[C = c]$$

- $A'B'C' \stackrel{def}{=} \{AB\}\{C\}$ means

$$\Pr[A' = a \wedge B' = b \wedge C' = c] \stackrel{def}{=} \Pr[A = a \wedge B = b] \Pr[C = c]$$

- $A'B'C' \stackrel{def}{=} \{AB\}\left\{C\middle|B\right\}$ means

$$\Pr[A' = a \wedge B' = b \wedge C' = c] \stackrel{def}{=} \Pr[A = a \wedge B = b] \Pr[C = c|B = b]$$

- Let $\tilde{A}$ be a random variable taking values in $\mathsf{supp}(A)$. Then $A', B' \stackrel{def}{=} \left\{\tilde{A}\right\}\left\{B\middle|\tilde{A}\right\}$ means

$$\Pr[A' = a \wedge B' = b] \stackrel{def}{=} \Pr[\tilde{A} = a] \Pr[B = b|A = a]$$

2.2 Statistical Distance

Sometimes the distributions we get are not exactly the distributions we want, but they may be *close* enough. The measure of *closeness* we will use is this one:

Definition 10. *Let D and F be two random variables taking values in a set $\mathcal{S}$. Their* statistical distance *is*

$$|D - F| \stackrel{def}{=} \max_{\mathcal{T} \subseteq \mathcal{S}} (|\Pr[D \in \mathcal{T}] - \Pr[F \in \mathcal{T}]|) = \frac{1}{2} \sum_{s \in \mathcal{S}} |\Pr[D = s] - \Pr[F = s]|$$

*If $|D - F| \leq \epsilon$ we shall say that D is ϵ-*close *to F. We shall also use the notation $D \stackrel{\epsilon}{\approx} F$ to mean D is ϵ-close to F.*

Proposition 11 (Triangle Inequality)**.** *Let A, B, C be random variables over $\mathcal{S}$, with $A \stackrel{\epsilon_1}{\approx} B \stackrel{\epsilon_2}{\approx} C$. Then $A \stackrel{\epsilon_1+\epsilon_2}{\approx} C$.*

Proposition 12 (Conditioning Close Distributions)**.** *Let A, B, A', B' be random variables such that $|A - A'| = 0$. Then for every a, $\left|B|A = a - B'|A' = a\right| \leq \left|AB - A'B'\right| / \Pr[A = a]$.*

Proposition 13. *Let A, A' be two random variables over $\mathcal{A}$ in the same probability space such that $\Pr[A \neq A'] \leq \epsilon$. Then $|A - A'| \leq \epsilon$.*

Proof. Let $\mathcal{S} \subset \mathcal{A}$ be any set. Then by the union bound we get $\Pr[A \in \mathcal{S}] \leq \Pr[A' \in \mathcal{S}] + \Pr[A \neq A']$, which clearly implies the proposition. □

Proposition 14 (Maintaining Independence)**.** *Let X, Y be independent random variables and E be an event that depends only on Y. Then $XY|E$ are independent random variables.*

2.3 Games

In this paper, a game is defined by a distribution (X, Y) on a set of questions, $\mathcal{X} \times \mathcal{Y}$, a set of possible answers $\mathcal{A} \times \mathcal{B}$ and a predicate $V : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$. A strategy for the game is a pair of functions $a : \mathcal{X} \to \mathcal{A}$ and $b : \mathcal{Y} \to \mathcal{B}$. The value of the game is the maximum of $\Pr_{X,Y}[V(X, Y, a(X), b(Y)]$, over all choices of strategies $a(\cdot), b(\cdot)$.

We call a game a *projection* game if there exists a family of functions $f_{x,y}$ indexed by $\mathcal{X} \times \mathcal{Y}$ such that $V(x, y, a, b)$ is equivalent to $f_{x,y}(b) = a$.

A game is called *unique* if it is a projection game with the additional property that all functions $f_{x,y}$ are bijections.

The *answer length* of a game is the quantity $\log |\mathcal{A}| + \log |\mathcal{B}|$.

Given a game $\mathcal{G}^n$ the parallel repetition of the game is the game with distribution on questions obtained by taking n independent samples $(X_1, Y_1) \cdots (X_n, Y_n)$. A strategy for the new game is specified by two functions $a : \mathcal{X}^n \to \mathcal{A}^n$ and $b : \mathcal{Y}^n \to \mathcal{B}^n$. When the game is played, the referee samples n independent pairs of questions as above and sends one question from each pair to each of the players. The players respond with n answers each. The referee then checks that the players win by checking the AND of the predicate in the original game in each of the n copies. Thus the value of the game is the maximum of $\Pr[V(X_1, Y_1, a_1(X_1), b_1(Y_1)) \wedge \cdots \wedge V(X_n, Y_n, a_n(X_n), b_n(Y_n))]$, over all choices of strategies $a(\cdot), b(\cdot)$.

3 Main Theorems

In this section, we prove our main theorems, assuming two lemmas which we prove in a later section. Fix an optimal strategy for the repeated game. For any set $S \subset [n]$, we let W_S denote the event that the players win all the games in coordinates included in the set S.

Lemma 15 (Main Lemma for General Games)**.** *Let $\mathcal{G}$ be a game with value at most $1 - \epsilon$, such that one player gives answers from a set of size 2^c. Let $S \subset [n]$ be a set of coordinates and $\gamma > 0$ be such that $\Pr[W_S] \geq 2^{-\gamma^2(n-|S|)+|S|c}$. Then for i chosen uniformly from outside S, we have that $\mathbb{E}_{i \notin S}\left[\Pr[W_{\{i\}}|W_S]\right] \leq 1 - \epsilon + 25\gamma$.*

The lemma says that as long as the probability of winning in S is not too small already, then then there is an index $i \notin S$ such that the probability of winning in $S \cup \{i\}$ is even smaller. Let $S_1 = \{1\}$, and let $S_j = S_{j-1} \cup i_j$ be chosen by picking the element $i_j \notin S_{j-1}$ that minimizes W_{S_j}. We shall use the lemma to show that the probability of winning in the set S_k (for large k) must be small. The idea is that either the probability in winning S_{k-1} is already extremely small, or we can use Lemma 17 to show that the probability of winning in S_k is significantly smaller than the probability of winning in S_{k-1} and proceed inductively.

Formally, let $t(\gamma, n, c, \epsilon)$ be such that

$$2^{-\gamma^2(n-t)+tc} = (1 - \epsilon + 25\gamma)^t \tag{1}$$

It is a computation to show that the unique t satisfying this equation is

$$t = \frac{\gamma^2 n}{c + \gamma^2 - \log(1 - \epsilon + 25\gamma)}$$

We can then use Lemma 15 to prove:

Lemma 16. *For every $\gamma > 0$, and $k \leq \lceil t \rceil$, $\Pr[W_{S_k}] \leq \max\{2^{-\gamma^2(n-t)+tc}, (1-\epsilon+25\gamma)^k\}$.*

Theorem 2 follows from Lemma 16 by setting $\gamma = \epsilon/50$ and setting $k = \lceil t \rceil$. For this setting of parameters, Lemma 16 and Equation 1 give that

$$\Pr[W_{[n]}] \leq \Pr[W_{S_k}] \leq (1-\epsilon/2)^t = (1-\epsilon/2)^{\frac{\gamma^2 n}{c+\gamma^2-\log(1-\epsilon/2)}}$$

Since $-\log(1-\epsilon/2) \leq 1$, this quantity is at most $(1-\epsilon/2)^{\alpha\epsilon^2 n/c}$ for some constant α.

Proof of Lemma 16. We prove the lemma by induction on k. For $k = 1$, $\Pr[W_{S_1}] \leq (1-\epsilon) \leq (1-\epsilon+25\gamma)$, so the bound holds. For general k, we have that $\Pr[W_{S_k}] = \Pr[W_{i_k}|W_{S_{k-1}}]\Pr[W_{S_{k-1}}]$. If $\Pr[W_{S_{k-1}}] \leq 2^{-\gamma^2(n-t)+tc}$, we are done since $\Pr[W_{S_k}] \leq \Pr[W_{S_{k-1}}]$. Otherwise by induction it must be the case that $2^{-\gamma^2(n-t)+tc} < \Pr[W_{S_{k-1}}] \leq (1-\epsilon+25\gamma)^{k-1}$.

Since $\Pr[W_{S_{k-1}}] \geq 2^{-\gamma^2(n-t)+tc} \geq 2^{-\gamma^2(n-(k-1))+(k-1)c}$, we apply Lemma 15 to show that there must exist $i \notin S_{k-1}$ for which $\Pr[W_{\{i\}}|W_{S_{k-1}}] \leq (1-\epsilon+25\gamma)$. Thus

$$\Pr[W_{S_k}] \leq \Pr[W_{\{i\}}|W_{S_{k-1}}]\Pr[W_{S_{k-1}}] \leq (1-\epsilon+25\gamma)^k.$$

□

For projection games, we shall prove a lemma that is independent of the alphabet length c:

Lemma 17 (Main Lemma for Projection Games)**.** *Let $S \subset [n]$ be a set, $\beta > 0$ be a number and $\mathcal{G}$ be a projection game such that $n - |S| \geq \frac{\log(1/5\beta^2)}{2\beta^2}$ and $\Pr[W_S] \geq 2^{-\beta^2(n-|S|)}$. Then $\mathbb{E}_{i\notin S}\left[\Pr[W_{\{i\}}|W_S]\right] \leq 1-\epsilon+75\beta$.*

The lemma can then be used to prove Theorem 4. Set $t(\gamma, n, \epsilon)$ so that

$$2^{-\beta^2(n-t)} = (1-\epsilon+75\beta)^t \tag{2}$$

It is a computation to show that

$$t = \frac{\beta^2 n}{\beta^2 - \log(1-\epsilon+75\beta)}$$

Lemma 18. *In a projection game, for every $\beta > 0$, and $k \leq \lceil t \rceil$, if $n \geq t + \frac{\log(1/5\beta^2)}{2\beta^2}$, then $\Pr[W_{S_k}] \leq \max\{2^{-\beta^2(n-t)+tc}, (1-\epsilon+75\beta)^k\}$.*

Theorem 4 follows from Lemma 18 by setting $\beta = \epsilon/150$ and setting $k = \lceil t \rceil$. For this setting of parameters, if $n \geq t + \frac{\log(1/5\beta^2)}{2\beta^2}$, Lemma 18 and Equation 2 give that

$$\Pr[W_{[n]}] \leq \Pr[W_{S_k}] \leq (1-\epsilon/2)^t = (1-\epsilon/2)^{\frac{\beta^2 n}{\beta^2-\log(1-\epsilon/2)}}$$

Since $-\log(1-\epsilon/2) = O(\epsilon)$, this quantity is at most $(1-\epsilon/2)^{\alpha\epsilon n}$ for some constant α. Further, if Theorem 4 is true for large n, then it must hold for every n:[2] if the value of $\mathcal{G}^n$ is larger than $(1-\epsilon/2)^{\alpha\epsilon n}$, then for every r, the players can win the game $\mathcal{G}^{nr}$ with probability at least $(1-\epsilon/2)^{\alpha\epsilon nr}$, just by playing the same strategy independently on each block of n games. Thus, we get a contradiction for large enough r.

[2]This argument was suggested by an anonymous referee.

Proof of Lemma 18. The proof is almost identical to the proof of Lemma 16. We prove the lemma by induction on k. For $k = 1$, $\Pr[W_{S_1}] \leq (1 - \epsilon) \leq (1 - \epsilon + 75\beta)$, so the bound holds. For general k, we have that $\Pr[W_{S_k}] = \Pr[W_{i_k}|W_{S_{k-1}}] \Pr[W_{S_{k-1}}]$. If $\Pr[W_{S_{k-1}}] \leq 2^{-\beta^2(n-t)}$, we are done since $\Pr[W_{S_k}] \leq \Pr[W_{S_{k-1}}]$. Otherwise by induction it must be the case that $2^{-\beta^2(n-t)} < \Pr[W_{S_{k-1}}] \leq (1 - \epsilon + 75\beta)^{k-1}$.

Since $\Pr[W_{S_{k-1}}] \geq 2^{-\beta^2(n-t)} \geq 2^{-\beta^2(n-(k-1))}$, and $n - (k-1) \geq n - t \geq \frac{\log(1/5\beta^2)}{2\beta^2}$ we can apply Lemma 17 to show that there must exist $i \notin S_{k-1}$ for which $\Pr[W_{\{i\}}|W_{S_{k-1}}] \leq (1 - \epsilon + 75\beta)$. Thus

$$\Pr[W_{S_k}] \leq \Pr[W_{\{i\}}|W_{S_{k-1}}] \Pr[W_{S_{k-1}}] \leq (1 - \epsilon + 75\beta)^k.$$

□

4 Sampling From Close Distributions

A variant of the following lemma was proved by Holenstein [Hol07]. The proof we sketch here is due to Boaz Barak.

Lemma 19 (Sampling similar distributions [Hol07])**.** *There exists a protocol for l non-communicating players such that given distributions $A_1, \ldots, A_l$ taking values in $\mathcal{A}$ such that $|A_l - A_i| \leq \epsilon_i$ for every $i \in [l]$, the players can use shared randomness to sample $B_1, \ldots, B_l$ with the property that:*

- *For every i, B_i has the same distribution as A_i.*
- *For every $i \leq l - 1$, $\Pr[B_l \neq B_i] \leq 2\epsilon_i$.*
- $\Pr[$*all samples are the same*$] \geq 1 - 2\sum_{i=1}^{l-1} \epsilon_i$

Proof Sketch: First note that the last guarantee follows from the second guarantee and the union bound.

To prove the first two guarantees, let us first consider the case that the A_i's are promised to be uniform over (possibly different) subsets of $\mathcal{A}$. In this case the protocol for the players is simple: the shared randomness is interpreted as a permutation of the universe $\mathcal{A}$. Each player then samples the first element of the permutation that lies in the support of her distribution. The lemma is then easily seen to be true.

To handle the general case, identify each distribution A_i with the uniform distribution on the set $\cup_{a \in \mathcal{A}} \{a\} \times [0, \Pr[A_i' = a]]$, which is a subset of $\mathcal{A} \times [0, 1]$. Then by tiling the set $\mathcal{A} \times [0, 1]$ with a fine enough grid, we can interpret the shared randomness as a permutation of the parts of this grid to get a protocol that is arbitrarily close to getting the bounds promised above. □

5 Conditioning Product Distributions

Fact 20. *Let A, B be random variables in some probability space. Let A' be another random variable such that $|A - A'| \leq \epsilon$. Then $\left|\{AB\} - \{A'\}\left\{B\middle|A'\right\}\right| \leq \epsilon$.*

We need a basic definition:

Definition 21 (Informational Divergence)**.** *Given two random variables U, V taking values in the same set $\mathcal{U}$, we define the informational divergence*

$$\mathsf{D}\left(U\middle||V\right) \stackrel{def}{=} \sum_{u\in\mathcal{U}} \Pr[U=u]\log\left(\frac{\Pr[U=u]}{\Pr[V=u]}\right)$$

where we adopt the convention that $0\log 0 = 0\log\frac{0}{0} = 0$. If there exists a $u \in \mathcal{U}$ for which $\Pr[V=u]=0$ but $\Pr[U=u]\neq 0$, we say that that $\mathsf{D}\left(U\middle||V\right)=\infty$.

The following are standard facts about informational divergence:

Fact 22. $\mathsf{D}\left(V\middle||U\right) \geq |U-V|^2$

Fact 23. *If V is a random variable, E is any event and $\tilde{V} \stackrel{def}{=} V|E$, in the same space with $\Pr[E]=2^{-d}$, then $\mathsf{D}\left(\tilde{V}\middle||V\right)\leq d$.*

Proof. We have $\mathsf{D}\left(\tilde{V}\middle||V\right) = \sum_{v\in\mathcal{V}}\Pr[V=v|E]\log\left(\frac{\Pr[V=v|E]}{\Pr[V=v]}\right)$. However, for every v,

$$\frac{\Pr[V=v|E]}{\Pr[V=v]} \leq \frac{1}{\Pr[E]} \leq 2^d$$

Thus $\mathsf{D}\left(\tilde{V}\middle||V\right) \leq \sum_{v\in\mathcal{V}}\Pr[V=v|E]d \leq d$. □

Fact 24. *If $U_1,\ldots,U_n$ are independent random variables and $V_1,\ldots,V_n$ are other random variables,*

$$\sum_{i=1}^{n}\mathsf{D}\left(V_i\middle||U_i\right) \leq \mathsf{D}\left(V_1\ldots V_n\middle||U_1\ldots U_n\right)$$

A key part of the proof will be showing that if we condition a product distribution on an event whose probability is not too low, there must be some coordinate which remains distributed how it was before the conditioning.

Lemma 25 ([Raz98])**.** *Let $U_1,U_2,\ldots,U_n$ be independent random variables. Suppose E is any event in the same probability space such that $\Pr[E]=2^{-d}$, then*

$$\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\{U_i\}-\left\{U_i\middle|E\right\}\right|\right] \leq \sqrt{\frac{d}{n}}$$

Proof.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\{U_i\}-\left\{U_i\middle|E\right\}\right|\right]^2 \\
&\leq \mathop{\mathbb{E}}_{i\in[n]}\left[\left|\{U_i\}-\left\{U_i\middle|E\right\}\right|^2\right] && \text{by convexity of the square function} \\
&\leq \mathop{\mathbb{E}}_{i\in[n]}\left[\mathsf{D}\left(\left\{U_i\middle|E\right\}\middle|\middle|\{U_i\}\right)\right] && \text{by Fact 22} \\
&\leq \frac{1}{n}\mathsf{D}\left(\left(U_1U_2\ldots U_n\middle|E\right)\middle||U_1\ldots U_n\right) && \text{by Fact 24} \\
&\leq \frac{d}{n} && \text{by Fact 23}
\end{aligned}$$

□

Next, we show that all of the above still holds if in addition to dense event, we condition on the value of some random variable with small support, and the variables are only independent in convex combination:

Corollary 26. *Let $R, U_1, U_2, \ldots, U_n, A$ be random variables and E be an event with $\Pr[E] = 2^{-d}$ such that*

- *For every r, $U_1, \ldots, U_n$ are independent conditioned on the event $R = r$.*
- *For every r, $|\mathsf{supp}\left(A \middle| E \wedge (R = r)\right)| \leq 2^h$*

Then,

$$\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\left\{RA\middle|E\right\}\left\{U_i\middle|R\right\} - \left\{RAU_i\middle|E\right\}\right|\right] \leq \sqrt{\frac{d+h}{n}}$$

Proof.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\left\{RA\middle|E\right\}\left\{U_i\middle|R\right\} - \left\{RAU_i\middle|E\right\}\right|\right]^2 \\
&= \mathop{\mathbb{E}}_{\substack{i\in[n]\\(a,r)\leftarrow(AR|E)}} \left[\left|\left\{U_i\middle|R=r\right\} - \left\{U_i\middle|E\wedge(A,R)=(a,r)\right\}\right|\right]^2 \\
&\leq \mathop{\mathbb{E}}_{(a,r)\leftarrow(AR|E)}\left[\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\left\{U_i\middle|R=r\right\} - \left\{U_i\middle|E\wedge(A,R)=(a,r)\right\}\right|\right]^2\right] && \text{by convexity} \\
&\leq \mathop{\mathbb{E}}_{(a,r)\leftarrow(AR|E)}\left[\frac{\log(1/\Pr[E\wedge A=a|R=r])}{n}\right] && \text{by Lemma 25} \\
&\leq (1/n)\log\left(\mathop{\mathbb{E}}_{r\leftarrow(R|E)}\left[\sum_{a\in\mathsf{supp}(A|E\wedge R=r)}\frac{\Pr[A=a|E\wedge(R=r)]}{\Pr[E\wedge A=a|R=r]}\right]\right) && \text{by concavity of log} \\
&= (1/n)\log\left(\mathop{\mathbb{E}}_{r\leftarrow(R|E)}\left[\sum_{a\in\mathsf{supp}(A|E\wedge R=r)}\frac{\Pr[R=r]}{\Pr[E\wedge R=r]}\right]\right) \\
&\leq (1/n)\log\left(2^h\mathop{\mathbb{E}}_{r\leftarrow(R|E)}\left[\frac{\Pr[R=r]}{\Pr[E\wedge R=r]}\right]\right) \\
&= (1/n)\log\left(2^h\mathop{\mathbb{E}}_{r\leftarrow(R|E)}\left[\frac{\Pr[R=r]}{\Pr[E]\Pr[R=r|E]}\right]\right) \\
&= (1/n)\log\left(2^h\sum_{r\in\mathcal{R}}\frac{\Pr[R=r]\Pr[R=r|E]}{\Pr[E]\Pr[R=r|E]}\right) \\
&= (1/n)\log\left(2^h/\Pr[E]\right) \\
&= \frac{d+h}{n}
\end{aligned}$$

□

Finally, we need the following Corollary to carry out the proof for the case of projection games:

Corollary 27. *Let $R, U_1, U_2, \ldots, U_n, A$ be random variables and E be an event. Suppose that for every r, $U_1, \ldots, U_n$ are independent conditioned on the event $R = r$. Let $H \subset \mathsf{supp}(R) \times \mathsf{supp}(A)$ be a set such that for every r, $|\{a|(r,a) \in H\}| \leq 2^h$, and $\Pr[H \wedge E] \geq 2^{-d}$. Then,*

$$\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\left\{RA\middle|E\right\}\left\{U_i\middle|R\right\} - \left\{RAU_i\middle|E\right\}\right|\right] \leq \sqrt{\Pr[H|E]\frac{d+h}{n} + (1-\Pr[H|E])}$$

Proof. We can separate the expectation into the part where H occurs and the part where it does not.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[n]}\left[\left|\left\{RA\middle|E\right\}\left\{U_i\middle|R\right\} - \left\{RAU_i\middle|E\right\}\right|\right]^2 \\
&= \mathop{\mathbb{E}}_{\substack{i\in[n]\\(a,r)\leftarrow(AR|E)}}\left[\left|\left\{U_i\middle|R=r\right\} - \left\{U_i\middle|E\wedge(A,R)=(a,r)\right\}\right|\right]^2 \\
&\leq \mathop{\mathbb{E}}_{i\in[n]}\left[\Pr[H|E]\mathop{\mathbb{E}}_{(a,r)\leftarrow(AR|E\wedge H)}\left[\left|\left\{U_i\middle|R=r\right\} - \left\{U_i\middle|E\wedge(A,R)=(a,r)\right\}\right|\right] + (1-\Pr[H|E])\right]^2 \\
&\leq \Pr[H|E]\mathop{\mathbb{E}}_{i\in[n]}\left[\mathop{\mathbb{E}}_{(a,r)\leftarrow(AR|E\wedge H)}\left[\left|\left\{U_i\middle|R=r\right\} - \left\{U_i\middle|E\wedge(A,R)=(a,r)\right\}\right|\right]\right]^2 + (1-\Pr[H|E])
\end{aligned}$$

where the last inequality is by convexity. By Corollary 26, we get that first term in this expectation is bounded by $\frac{d+h}{n}$. This gives us the final bound.

□

6 Proof of Main Lemmas

In this section, we shall prove Lemma 15 and Lemma 17.

These lemmas say that as long the probability of winning in the k coordinates in S is not too small, then on average, the players must be doing pretty badly on the remaining coordinates even conditioned on winning in W_S.

Without loss of generality, we assume that $S = \{n-k+1, n-k, \ldots, n\}$. We shall prove these lemmas by contradiction. Fix a strategy for $\mathcal{G}^n$. Suppose for the sake of contradiction that $\Pr[W_S]$ is high and $\mathbb{E}_{i\notin S}\left[\Pr[W_{\{i\}}|W_S]\right] > 1 - \epsilon/2$. Then we shall use the players strategy for $\mathcal{G}^n$ to get an extremely good strategy for $\mathcal{G}$, one that wins with probability more than $1-\epsilon$, and thus contradicting the bound on the value of $\mathcal{G}$.

6.1 Intuition for the proof

We first outline a natural way to use a strategy for $\mathcal{G}^n$ to get a strategy for $\mathcal{G}$, which we shall ultimately refine to complete the proof: the players decide on an index i such that given questions (X, Y) in $\mathcal{G}$, they can use shared randomness to generate $n-1$ pairs of questions such that when the questions (X, Y) are placed in the i'th coordinate, and the rest of the questions are placed in the appropriate coordinates, the resulting distribution is statistically close to the distribution

$(X_1, Y_1) \dots (X_n, Y_n)|W_S$. If the players can find such an index i, then they could just use the strategy of the players for $\mathcal{G}^n$ to win $\mathcal{G}$ in the i'th coordinate with probability more than $1 - \epsilon$.

There are a couple of obstacles to getting this approach to work. The most immediately apparent obstacle is that it must be true that there exists an index i for which $(X_i, Y_i)|W_S$ is statistically close to (X, Y). This obstacle can easily be circumvented via Lemma 25. A more subtle issue is that the players have to generate the rest of the questions without communicating. Dealing with this issue will take up most of our effort in the proof. To understand under what circumstances it is possible for two players to generate questions that satisfy the above properties, let us first look at some simple cases. Below, let X^n denote $(X_1, \dots, X_n)$ and Y^n denote $(Y_1, \dots, Y_n)$.

Independent Distributions. Suppose every $(x, y) \in (\mathcal{X}, \mathcal{Y})$ was such that $\left(X^n Y^n \middle| W_S \wedge (X_i = x)\right)$ and $\left(X^n Y^n \middle| W_S \wedge (Y_i = y)\right)$ are both product distributions. Then given the questions (x, y), the first player can sample $\left(X^n \middle| W_S \wedge (X_i = x)\right)$ and the second player can independently sample $\left(Y^n \middle| W_S \wedge (Y_i = y)\right)$. It is then easy to see that if X, Y was statistically close to $(X_i, Y_i)|W_S$ (which we can guarantee using Lemma 25), the players do sample questions which are statistically close to $X^n Y^n | W_S$. Of course the assumption that we have such independence is unreasonable. In general, the first player's questions are not independent of the second players questions. Even if the game was such that (X, Y) is a product distribution, conditioning on W_S could potentially introduce complicated dependencies between the questions of the players.

Completely correlated distributions. Next suppose we could somehow prove that there exists some random variable R and functions f, g such that for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$:

- Learning R would allow both players to generate the random variables they want using their inputs
$$(f(R, x), g(R, y)) \approx \left(X^n Y^n \middle| W_S \wedge (X_i = x) \wedge (Y_i = y)\right)$$
- Although R may depend on (X_i, Y_i), all the information needed to generate R is contained in any one of these variables:
$$\left(R \middle| W_S \wedge (X_i = x)\right) \approx \left(R \middle| W_S \wedge (Y_i = y)\right) \approx \left(R \middle| W_S \wedge (X_i = x) \wedge (Y_i = y)\right)$$

Given these conditions, it is easy to design a protocol for the players — each player computes the distribution for R based on his question and they then use Lemma 19 to agree on a sample for $\left(R \middle| W_S \wedge (X_i = x) \wedge (Y_i = y)\right)$. The lemma and the second condition above guarantee that the distribution they end up with will be statistically close to the right one. Once they have generated the sample for R, they simply apply the functions f, g to generate their corresponding questions.

The solution for the general case will be a mixture of the solutions in the above two cases. We shall identify an index i and a random variable R such that:

- Fixing $R = r$ will determine at least one question of (X_i, Y_i) for every coordinate i. If A denotes the answers of one of the players in the coordinates $n - k + 1, \dots, n$, this condition guarantees that for every r, a, x, y, X^n, Y^n are independent conditioned on the event $(R, A, X_i) = (r, a, x) \wedge W_S$. Similarly X^n, Y^n are independent conditioned on the event $(R, A, Y_i) = (r, a, y) \wedge W_S$.

- Conditioned on W_S, all the information needed to generate RA given (X_i, Y_i) is contained in any one of these variables:

$$\left\{X_iY_i\middle|W_S\right\}\left\{RA\middle|X_i \wedge W_S\right\} \approx \left\{X_iY_iRA\middle|W_S\right\} \approx \left\{X_iY_i\middle|W_S\right\}\left\{RA\middle|Y_i \wedge W_S\right\}$$

Once we are able to determine such R, A and prove the above properties, we shall be done. On receiving the questions x, y, the players will use the protocols from Lemma 19 to generate $\left(RA\middle|(X_i, Y_i) = (x, y) \wedge W_S\right)$. Once they have sampled this random variable, they can generate the rest of their questions independently. This would prove that $\Pr[W_{\{i\}}|W_S]$ must be small.

The stronger results that apply to projection games in this paper come about by proving that in these kinds of games, the only way the player can win is by using a strategy that restricts itself to using a few possible answers. We can define a new sub-event of W_S that ensures that not only do the players win in the coordinates of S, they do so by using answers that have a relatively high probability. We can show that this event has an extremely high density in W_S, so that conditioning on W_S is essentially the same as conditioning on this event. This allows us to carry out the proof as if the effective answer size of the provers is much smaller than it actually is.

6.2 The proof

Let $A = A_{n-k+1} \dots A_n$ and $B = B_{n-k+1} \dots B_n$ denote the answers of the players in the last k games. Let $V = V_1, V_2, \dots, V_{n-k}$ denote uniformly random bits.

For $i = 1, 2, \dots, n-k$, let T_i denote a random question in every coordinate:

$$T_i \stackrel{def}{=} \begin{cases} X_i & \text{if } V_i = 1, \\ Y_i & \text{if } V_i = 0. \end{cases}$$

Let U_i's denote the opposite questions:

$$U_i \stackrel{def}{=} \begin{cases} X_i & \text{if } V_i = 0, \\ Y_i & \text{if } V_i = 1. \end{cases}$$

Set $Q \stackrel{def}{=} X_{n-k+1}X_{n-k+2} \dots X_nY_{n-k+1}Y_{n-k+2} \dots Y_n$ — the questions in the last k games.

Set $R \stackrel{def}{=} VQT_1T_2 \dots T_{n-k}$ — the "won" questions and a random question from each of the remaining question pairs.

Set $R^{-j} \stackrel{def}{=} V_1V_2 \dots V_{j-1}V_{j+1} \dots V_{n-k}QT_1T_2 \dots T_{j-1}T_{j+1} \dots T_{n-k}$ — removing the j'th coordinate from R.

The most technical part of the proof is the following lemma:

Lemma 28. *Let E be any event that is determined by ABR and let $H \subset \mathsf{supp}(R) \times \mathsf{supp}(A)$ be a set determining the event $(R, A) \in H$. Let $h, \gamma > 0$ be such that*

- *For every r, $|\{a|H\}| \leq 2^h$*
- $\Pr[E \wedge H] \geq 2^{-\gamma^2(n-k)+h}$.
- $\Pr[H|E] \geq 1 - \gamma^2$.

Then $\mathbb{E}_{i \notin S}\left[\Pr[W_{\{i\}}|E]\right] \leq 1 - \epsilon + 25\gamma$.

Before proving this lemma, let us first see how we can use it to prove Lemma 15 and Lemma 17.

Proof of Lemma 15. Set $E = W_S$ and $H = \mathsf{supp}(R) \times \mathsf{supp}(A)$ to be all points. It is clear that E is determined by QAB which is contained in ABR. If the game is such that each answer comes from a set of size 2^c, $|\mathsf{supp}(A|R = r)|$ is trivially bounded by 2^{kc}. Set $h = kc$ and apply Lemma 28 to get Lemma 15. □

Next we give the proof for the case of projection games.

Proof of Lemma 17. Recall that in a projection game, we have the condition that there is some function f_Q, determined by the questions Q such that W_S implies that $f_Q(B) = A$.

For any tuple $(a, r) \in (\mathcal{A}, \mathcal{R})$, say that (a, r) is heavy if $\Pr[A = a|R = r] \geq 2^{-h}$, where h is a parameter that we shall fix later. The intuition behind this definition is that conditioned on $R = r$, the answers $A, B|R = r$ are independent. Thus the players should be able to win the projection game with a decent probability only when they pick one of the heavy elements, and there cannot be too many of those. For instance, imagine that f was the identity function. Then it is easy to check that if A, B are independent and $\Pr[A = B]$ is θ, there must be a set of size $O(1/\theta)$ (namely the elements with weight at least $\theta/100$) which A lands in with high probability.

Let H denote the event that (A, R) is heavy. The first condition in Lemma 28 is trivially satisfied. We shall argue that when the players win, they usually win inside the event H. Note that for every r, $A, B|R = r$ is a product distribution.

$$\Pr[W_S \wedge H^c] \leq \sum_{(b,r) \text{ s.t. } (f_q(b),r) \text{ is not heavy}} \Pr[R = r, B = b] \Pr[A = f_q(b)|R = r] \leq 2^{-h}$$

Set $h = 3\beta^2(n - k)$. Then the above equation implies that:

$$\begin{aligned}
\Pr[H \wedge W_S] &= \Pr[H|W_S] \Pr[W_S] \geq \Pr[W_S] - 2^{-h} \\
&\geq 2^{-\beta^2(n-k)} - 2^{-3\beta^2(n-k)} \\
&= 2^{-2\beta^2(n-k)}(2^{\beta^2(n-k)} - 2^{-\beta^2(n-k)}) \\
&\geq 2^{-2\beta^2(n-k)} \\
&= 2^{-5\beta^2(n-k)+h}
\end{aligned}$$

Set $E = W_S$. Note that E is determined by ABR. Observe that for every i, $\Pr[W_{\{i\}}|W_S] \leq \Pr[W_{\{i\}}|H \wedge W_S] \Pr[H|W_S] + 2^{-h}/\Pr[W_S]$.

Next observe that

$$\Pr[H|W_S] = \frac{\Pr[H \wedge W_S]}{\Pr[W_S]} = \frac{\Pr[W_S] - \Pr[W_S \wedge H^c]}{\Pr[W_S]} \geq 1 - 2^{-h}/\Pr[W_S]$$

This last quantity is bounded by $1 - 2^{-3\beta^2(n-k)+\beta^2(n-k)} \geq 1 - 5\beta^2$ by the assumption on $n - k$. Applying Lemma 28 with $\gamma = \sqrt{5}\beta \leq 3\beta$, we get that

$$\mathop{\mathbb{E}}_{i \notin S}\left[\Pr[W_{\{i\}}|W_S]\right] \leq 1 - \epsilon + 75\beta$$

□

Finally, we prove Lemma 28.

Proof of Lemma 28. We shall first show that in expectation over a random choice of the index i, if the players use the protocol from Lemma 19 to generate $AR^{-i}|E$ assuming that their questions came from the distribution $X_iY_i|E$, then with high probability they sample the same value for this variable which implies that the distribution they sample is close to

$$\{X_iY_i\}\{AR^{-i}|EX_iY_i\} \approx \{X_iY_i|E\}\{AR^{-i}|EX_iY_i\}$$

Then we shall argue that if the players complete the rest of the questions they need independently, the joint distribution of questions they get is close to $X^nY^n|E$.

We shall use the following shorthand to simplify notation: an expression like $F_i \overset{\gamma}{\approx}_{\mathbb{E}_{i\notin S}} G_i$ stands for the statement $\mathbb{E}_{i\notin S}[|F_i - G_i|] \leq \gamma$.

Claim 29. $\{X_iY_i|E\} \overset{\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{X_iY_i\} =_{\mathbb{E}_{i\notin S}} \{XY\}$

Proof. The claim follows by Lemma 25 applied to the event E and the product distribution of the questions. Note that $\Pr[E] \geq \Pr[E \wedge H] \geq 2^{-\gamma^2(n-k)+h} \geq 2^{-\gamma^2(n-k)}$.

$$\mathop{\mathbb{E}}_{i\notin S}\left[\left|\{X_iY_i|E\} - \{X_iY_i\}\right|\right] \leq \sqrt{\frac{\gamma^2(n-k)}{n-k}} = \gamma$$

□

We apply Corollary 27 to get that:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\notin S}\left[\left|\{AR|E\}\{U_i|R\} - \{ARU_i|E\}\right|\right] \\
&\leq \sqrt{\Pr[H|E]\frac{\gamma^2(n-k)-h+h}{n-k} + 1 - \Pr[H|E]} \\
&\leq \sqrt{2\gamma^2} \leq 2\gamma
\end{aligned}$$

Note that for every i, $\{AR|E\}\{U_i|R\} = \{AR|E\}\{U_i|T_iV_i\}$, since U_i is independent of all the other random variables in R. Also, we have that for every i, V_i is a uniformly random bit in both distributions. Thus, we can conditioning on $V_i = 0$ and apply Proposition 12 to get

$$\{ARU_i|E\} \overset{2\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{AR|E\}\{U_i|T_iV_i\} \Rightarrow \{AR^{-i}Y_iX_i|E\} \overset{4\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{AR^{-i}Y_i|E\}\{X_i|Y_i\} \tag{3}$$

We can then argue that

$$\begin{aligned}
&\{X_iY_i\}\{AR^{-i}|Y_iX_iE\} && \\
&\overset{\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{X_iY_i|E\}\{AR^{-i}|Y_iX_iE\} && \text{by Claim 29} \\
&=_{\mathbb{E}_{i\notin S}} \{AR^{-i}X_iY_i|E\} && \text{rearranging} \\
&\overset{4\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{AR^{-i}Y_i|E\}\{X_i|Y_i\} && \text{by Equation 3} \\
&=_{\mathbb{E}_{i\notin S}} \{Y_i|E\}\{X_i|Y_i\}\{AR^{-i}|Y_iE\} && \text{rearranging} \\
&\overset{\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{Y_i\}\{X_i|Y_i\}\{AR^{-i}|Y_iE\} && \text{by Claim 29} \\
&=_{\mathbb{E}_{i\notin S}} \{X_iY_i\}\{AR^{-i}|Y_iE\} && \text{rearranging}
\end{aligned}$$

Repeating the argument but conditioning on $V_i = 1$, we get

Claim 30. $\{X_iY_i\}\left\{AR^{-i}\middle|X_iE\right\} \overset{6\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{X_iY_i\}\left\{AR^{-i}\middle|X_iY_iE\right\} \overset{6\gamma}{\approx}_{\mathbb{E}_{i\notin S}} \{X_iY_i\}\left\{AR^{-i}\middle|Y_iE\right\}$

At this point we have made a lot of progress. We have shown that each player has roughly the same information about the random variable AR^{-i}, even in the event E. We imagine that we run the protocol promised by Lemma 19 using the two players in our game, plus an additional player who gets access to both questions (x, y). All players generate AR^{-i} conditioned on E and whatever questions they have. Then by Lemma 19 we get a protocol which has the effect that

- player 1's variables have the distribution $\{X_i\}\left\{AR^{-i}\middle|X_iE\right\}$
- player 2's variables have the distribution $\{Y_i\}\left\{AR^{-i}\middle|Y_iE\right\}$
- player 3's variables have the distribution $\{X_iY_i\}\left\{AR^{-i}\middle|X_iY_iE\right\}$
- $\mathbb{E}_{i\notin S}\left[\Pr[\text{the players have inconsistent variables when they use the index } i]\right] \leq 2(6\gamma)+2(6\gamma) = 24\gamma$

This means that the joint distribution that the first two players get is 24γ-close to the distribution of the third player. But this third player samples from a distribution that is close to the one we want:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\notin S}\left[\left|\{X_iY_i\}\left\{AR^{-i}\middle|X_iY_iE\right\} - \left\{X_iY_iAR^{-i}\middle|E\right\}\right|\right] \\
&\leq \mathop{\mathbb{E}}_{i\notin S}\left[\left|\{X_iY_i\} - \left\{X_iY_i\middle|E\right\}\right|\right] \\
&\leq \gamma \qquad \text{by Claim 29}
\end{aligned}$$

For an average i, the first two players sample from a distribution that is 25γ close to the correct distribution.

Each of the players then sample the rest of her questions conditioned on the questions and answers that she knows. To end the proof, observe that for every i, x, y, a, r^{-i},

$$\left|\left(X^n\middle|(X_i, R^{-i}, A) = (x, r^{-i}, a) \wedge E\right) - \left(X^n\middle|(X_i, Y_i, R^{-i}, A) = (x, y, r^{-i}, a) \wedge E\right)\right| = 0$$

and

$$\left|\left(Y^n\middle|(Y_i, R^{-i}, A) = (y, r^{-i}, a) \wedge E\right) - \left(Y^n\middle|(X_i, Y_i, R^{-i}, A) = (x, y, r^{-i}, a) \wedge E\right)\right| = 0$$

This is because after fixing i, x, a, r^{-i}, X^n is independent of Y^n. After this fixing, E depends only on Y^n. Thus, by Proposition 14 X^n, Y^n remain independent even after conditioning on the event E. In particular, X^n is independent of Y_i, proving the first statement above. The second equation follows similarly: after fixing i, y, a, r^{-i}, X^n is independent of Y^n and they remain independent even after conditioning on the event E. Thus Y^n is independent of X_i in this space.

We have argued that the players can generate a distribution that is 25γ close to $X^nY^n|E$. This gives us our final bound:

$$1 - \epsilon \geq \mathop{\mathbb{E}}_{i\notin S}\left[\Pr[W_{\{i\}}|E]\right] - 25\gamma$$

□

7 Consequences for Unique Games

It is clear that Conjecture 3 implies Conjecture 5. To prove the converse, we show that if Conjecture 3 is false, then so is Conjecture 5. Suppose that there is some γ such that for every constant c, there is a polynomial time algorithm that can distinguish Unique Games instances with value $(1-\gamma)$ from those with value γ. Without loss of generality, we may assume that $f(1/\epsilon) < \epsilon^{-0.9}$ by decreasing the value of f at every point at which it violates this condition.

Given a unique games instance $\mathcal{G}$, we can run the algorithm on $\mathcal{G}^{1/f(1/\epsilon)\epsilon^2}$. We show how to set ϵ to be small enough so that if the original game had value $(1-\epsilon^2)$, the new game must have a value of at least $1-\gamma$, and if the original game had a value of $(1-\epsilon f(1/\epsilon))$, the new game must a value less than γ. In the first case, the value of $\mathcal{G}^{1/f(1/\epsilon)\epsilon^2}$ is at least $(1-\epsilon^2)^{1/f(1/\epsilon)\epsilon^2}$. Since $\lim_{\epsilon\to 0}(1-\epsilon^2)^{1/f(1/\epsilon)\epsilon^2} = 1$, for small enough ϵ, this is at least $(1-\gamma)$. In the second case, by Theorem 4, the value of $\mathcal{G}^{1/f(1/\epsilon)\epsilon^2}$ is at most $(1-\epsilon f(1/\epsilon)/2)^{\alpha/\epsilon} = \left((1-\epsilon f(1/\epsilon)/2)^{2/\epsilon f(1/\epsilon)}\right)^{\alpha f(1/\epsilon)/2}$.

Since $\lim_{x\to 0}(1-x)^{1/x} = 1/e$, and $\epsilon f(1/\epsilon) \leq \epsilon^{0.1}$, we have that

$$\lim_{\epsilon\to 0}\left((1-\epsilon f(1/\epsilon)/2)^{2/\epsilon f(1/\epsilon)}\right)^{\alpha f(1/\epsilon)/2} = 0,$$

and we can set ϵ to be small enough so that this quantity is bounded by γ. Thus, the algorithm contradicting Conjecture 3 can be used to get a new algorithm that distinguishes games with value $(1-\epsilon^2)$ from games with value $(1-\epsilon f(1/\epsilon))$, contradicting Conjecture 5.

8 The Concentration Bound

In this section we prove Theorem 7.

Here we made no effort to optimize the constants appearing in the bound. It is conceivable that these can be improved significantly. In analogy with the the results from the earlier section, the bounds can easily be improved in the case of projection games to remove the dependence on c. Here we present an alternate, simpler proof of our original result suggested to us by an anonymous referee.

Set $t(\gamma, n, c, \epsilon)$ such that

$$2^{-\gamma^2(n-t)+tc} = (1-\epsilon+25\gamma)^t \tag{4}$$

It is a computation to show that the unique t satisfying the constraint is

$$t = \frac{\gamma^2 n}{c+\gamma^2-\log(1-\epsilon+25\gamma)}$$

Fix a strategy for the two players. Define a distribution on sets as following. Let $i_1, i_2, \ldots, i_k$ denote a sequence of random distinct elements of $[n]$. For each $j = 1, \ldots, \lceil t\rceil$, let S_j denote the union of the first j elements. Recall that W_S denotes the event that the players win in the coordinates included in the set S.

We start by proving the following lemma:

Lemma 31. *For every $\gamma > 0$, $\Pr[W_{S_{\lceil t\rceil}}] \leq 2(1-\epsilon+25\gamma)^t$.*

Proof. For each $j = 1, \ldots, \lceil t \rceil$, let L_j denote the event determined by S_j that $\Pr[W_{S_j}] \leq (1 - \epsilon + 25\gamma)^t$, and H_j denote the complement event. Then $\Pr[W_{S_{\lceil t \rceil}}] = \Pr[W_{S_{\lceil t \rceil}} \wedge L_{\lceil t \rceil}] + \Pr[W_{S_{\lceil t \rceil}} \wedge H_{\lceil t \rceil}]$. The first term is bounded by $\Pr[W_{S_{\lceil t \rceil}} | L_{\lceil t \rceil}] \leq (1 - \epsilon + 25\gamma)^t$ by definition. To bound the second term,

$$\Pr[W_{S_{\lceil t \rceil}} \wedge H_{\lceil t \rceil}] = \prod_{j=1}^{\lceil t \rceil} \Pr[W_{S_j} \wedge H_j | W_{S_{j-1}} \wedge H_{j-1}]$$

By Lemma 15, the j'th term in this product is bounded by $\Pr[W_{S_j} | W_{S_{j-1}} \wedge H_{j-1}] \leq (1 - \epsilon + 25\gamma)$. Thus, $\Pr[W_{S_{\lceil t \rceil}}] \leq 2(1 - \epsilon + 25\gamma)^t$. □

We set $\gamma = \delta/100$. For this setting of parameters, we have that $\Pr[W_{S_{\lceil t \rceil}}] \leq 2(1 - \epsilon + \delta/4)^{\lceil t \rceil}$.

Now we use an argument implicit in the work of Schmidt et al. [SSS95]:

Let Z be the random variable that denotes the number of games that are won. Then whenever $Z \geq (1 - \epsilon + \delta)n$, let us pick a random subset S of size $\lceil t \rceil$ from the set of coordinates where the players won, and blame this set for this bad event. Then the probability that we blame any fixed set S is at most $\Pr[W_S]\binom{\lceil (1-\epsilon+\delta)n \rceil}{\lceil t \rceil}^{-1}$. Thus, by the union bound,

$$\Pr[Z \geq (1 - \epsilon + \delta)n] \leq \sum_S \Pr[W_S] \binom{\lceil (1 - \epsilon + \delta)n \rceil}{\lceil t \rceil}^{-1}$$

This quantity is at most

$$\binom{n}{\lceil t \rceil} (1 - \epsilon + \delta/4)^{\lceil t \rceil} \binom{\lceil (1 - \epsilon + \delta)n \rceil}{\lceil t \rceil}^{-1}$$

by Lemma 31. Simplifying, we get that

$$\begin{aligned}
\Pr[Z \geq (1 - \epsilon + \delta)n] &\leq (1 - \epsilon + \delta/4)^{\lceil t \rceil} \left(\frac{n}{(1 - \epsilon + \delta)n - \lceil t \rceil + 1} \right)^{\lceil t \rceil} \\
&\leq 2 \left(\frac{n(1 - \epsilon + \delta/4)}{(1 - \epsilon + \delta)n - \lceil t \rceil + 1} \right)^{\lceil t \rceil} \\
&\leq 2 \left(\frac{n(1 - \epsilon + \delta/4)}{(1 - \epsilon + \delta)n - t} \right)^{\lceil t \rceil} \\
&\leq 2 \left(\frac{1 - \epsilon + \delta/4}{1 - \epsilon + 3\delta/4} \right)^{\lceil t \rceil}
\end{aligned}$$

Where for the last inequality we used the fact that $t/n \leq (\delta/100)^2 \leq \delta/4$. Thus,

$$\Pr[Z \geq (1 - \epsilon + \delta)n] \leq 2 \left(\frac{1 - \epsilon + \delta/4}{1 - \epsilon + 3\delta/4} \right)^{\lceil t \rceil} = \left(1 - \frac{\delta/2}{1 - \epsilon + 3\delta/4} \right)^{\lceil t \rceil}$$

Finally, we observe that $\lceil t \rceil \geq r$ (as in the statement of the theorem), for a small enough constant α.

9 Acknowledgments

Thanks to Boaz Barak, Guy Kindler, Venkatesan Guruswami, Scott Aaronson, Avi Wigderson, David Zuckerman and Parikshit Gopalan for useful discussions. Boaz Barak gave the intuition for the proof of Holenstein's lemma sketched in Lemma 19. Venkatesan Guruswami pointed out the application of our work to the unique games conjecture. Scott Aaronson told us the motivation for getting a concentration bound. We would also like to thank anonymous referees who suggested how to remove a required lowerbound on n from our main theorem in an earlier version of this work, found a mistake in an earlier version of this paper, and gave us a reference that simplified the proof of the concentration bound.

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45, 1998.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998.

[Aus07] Per Austrin. Balanced max 2-sat might not be the hardest. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 189–197. ACM, 2007.

[BCH+02] Jonathan Barrett, Daniel Collins, Lucien Hardy, Adrian Kent, and Sandu Popescu. Quantum nonlocality, bell inequalities and the memory loophole. *Physical Review A*, 66:042111, 2002.

[Bel64] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–290, 1964.

[BGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pages 113–131, 1988.

[BGKW89] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Efficient identification schemes using two prover interactive proofs. In *Advances in Cryptology — CRYPTO '92, 12th Annual International Cryptology Conference, Proceedings*, 1989.

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[CKK+06] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. *Proceedings of the 21th Annual IEEE Conference on Computational Complexity*, 15, 2006.

[CMM06] Eden Chlamtac, Konstantin Makarychev, and Yury Makarychev. How to play unique games using embeddings. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006.

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

[DMR06] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*. ACM, 2006.

[DFK+92] Cynthia Dwork, Uriel Feige, Joe Kilian, Moni Naor, and Shmuel Safra. Low communication 2-prover zero-knowledge proofs for NP. In *Advances in Cryptology — CRYPTO '92, 12th Annual International Cryptology Conference, Proceedings*, 1992.

[FGL+91] Uriel Feige, Shafi Goldwasser, Laszlo Lovasz, Shmuel Safra, and Mario Szegedy. Approximating clique is almost NP-complete (preliminary version). In *Proceedings of the 32nd Annual IEEE Symposium on Foundations of Computer Science*, 1991.

[FL92] Uriel Feige and Laszlo Lovasz. Two-prover one-round proof systems: their power and their problems. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 733–744, 1992.

[FV02] Uriel Feige and Oleg Verbitsky. Error reduction by parallel repetition–A negative result. *Combinatorica*, 22, 2002.

[Gil03] Richard D. Gill. Accardi contra bell (cum mundi): The impossible coupling. *IMS LECTURE NOTES-MONOGRAPH SERIES*, 42, 2003.

[GT06] Anupam Gupta and Kunal Talwar. Approximating unique games. In *SODA*, pages 99–106. ACM Press, 2006.

[Hol07] Thomas Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 2007.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[KKMO04] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable CSPs? In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, 2004.

[KR03] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2-ϵ. In *IEEE Conference on Computational Complexity*, page 379. IEEE Computer Society, 2003.

[KV05] Subhash Khot and Nisheeth Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, 2005.

[LS95] Dror Lapidot and Adi Shamir. A one-round, two-prover, zero-knowledge protocol for NP. *Combinatorica*, 15, 1995.

[LY93] Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 286–293, 1993.

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low in.uences invariance and optimality. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 21–30. IEEE Computer Society, 2005.

[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 29th Annual ACM Symposium on Theory of Computing*, pages 363–372, 1997.

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

[Raz08] Ran Raz. A counterexample to strong parallel repetition. Technical Report TR08-018, ECCC: Electronic Colloquium on Computational Complexity, 2008.

[SSS95] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff–Hoeffding bounds for applications with limited independence. *SIAM Journal on Discrete Mathematics*, 8(2):223–250, May 1995.

[Tre05] Luca Trevisan. Approximation algorithms for unique games. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, 2005.

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