SIAM J. COMPUT.
Vol. 30, No. 1, pp. 324–346

TWO-PROVER PROTOCOLS—LOW ERROR AT AFFORDABLE RATES[*]

URIEL FEIGE[†] AND JOE KILIAN[‡]

Abstract. We introduce the *miss-match* form for two-prover one-round proof systems. Any two-prover one-round proof system can be easily modified so as to be in miss-match form. Proof systems in miss-match form have the "projection" property that is important for deriving hardness of approximation results for NP-hard combinatorial optimization problems.

Our main result is an upper bound on the number of parallel repetitions that suffice in order to reduce the error of miss-match proof systems from p to ϵ. This upper bound depends only on p and on ϵ (polynomial in $1/(1-p)$ and in $1/\epsilon$). Based on previous work, it follows that for any $\epsilon > 0$, NP has two-prover one-round proof systems with logarithmic-sized questions, constant-sized answers, and error at most ϵ.

As part of our proof we prove upper bounds on the influence of random variables on multivariate functions, which may be of independent interest.

Key words. interactive proof systems, complexity theory

AMS subject classification. 68Q15

PII. S0097539797325375

1. Introduction.

A two-prover one-round proof system [6] is a protocol by which two provers jointly try to convince a computationally limited probabilistic verifier that a common input belongs to a prespecified language. The verifier selects a pair of questions at random. Each prover sees only one of the two questions and sends back an answer. The verifier evaluates a predicate on the common input and the two questions and answers, and accepts or rejects according to the output of the predicate. For inputs in the language, the provers have a strategy (where a strategy for a prover is a function from incoming messages to outgoing messages) that always causes the verifier to accept. For inputs not in the language, regardless of the strategy used by the provers, the verifier accepts with probability at most ϵ. The smaller the value of ϵ, known as the *error*, the greater the verifier's confidence in the result of the proof system.

The class MIP(2,1) denotes those languages L accepted by a two-prover one-round proof system with a probabilistic polynomial-time verifier. More generally, an MIP(2,1) *game* is one in which a verifier engages in a single round of communication with two cooperating but not communicating players (the provers); the verifier determines based on the messages and its private coin tosses whether the players win. Similarly, in an MIP(2,1) *proof system*, "yes" instances (when $x \in L$) give rise to trivial games, in which the provers can always win (we only consider proof systems with perfect completeness) and "no" instances give rise to nontrivial games in which the provers can win with probability at most $\epsilon < 1$. We call ϵ the *error* of the proof

[*] Received by the editors August 4, 1997; accepted for publication (in revised form) November 23, 1999; published electronically May 15, 2000. A preliminary version of this manuscript appeared in Proceedings of the 26th Annual ACM Symposium on the Theory of Computing, Las Vegas, NV, 1994, pp. 172–183.

http://www.siam.org/journals/sicomp/30-1/32537.html

[†] Department of Computer Science and Applied Mathematics, The Weizmann Institute, Rehovot, Israel (feige@wisdom.weizmann.ac.il). This author is the incumbent of the Joseph and Celia Reskin Career Development Chair. This research was done in part while this author was visiting the NEC Research Institute.

[‡] NEC Research Institute, 4 Independence Way, Princeton, NJ (joe@research.nj.nec.com).

system. Reducing the error in MIP(2,1) proof systems (while preserving triviality for "yes" instances) is a subtle issue. A natural approach is to repeat an MIP(2,1) protocol n times and accept only if all executions are accepting. Ideally, one would hope that this method would reduce the error to ϵ^n. This is indeed true if each execution is performed with a fresh pair of provers, requiring n pairs of provers, or if the executions are performed sequentially (each prover must answer each question online before seeing the question for the next execution), requiring n rounds of communication. However, parallel repetition—in which there are only two provers and one round, and each prover sends out its answers only after receiving all its questions—is not guaranteed to reduce the error to ϵ^n [15]. Much work was invested in trying to analyze the rate at which parallel repetition reduces the error in MIP(2,1) proof systems (see, e.g., [15, 22, 7, 13, 14]).

We analyze a specific class of MIP(2,1) proof systems which we call *miss-match* proof systems. In the basic one-round proof system, the question to the first prover is composed of two "half questions" (α_1, α_2), and the first prover replies with two "half answers" (β_1, β_2). Based on $\alpha_1, \alpha_2, \beta_1$, and β_2, the verifier makes an initial decision on whether to reject or provisionally accept, pending the results of its interaction with the second prover. This acceptance predicate depends on the specific proof system and may differ from one miss-match proof system to another. The common feature of all miss-match proof systems is the way in which the second prover is used to confirm a decision to accept. Here the verifier has two options. The first, *miss* option, is to ask the second prover a null question λ, ignore the second prover's answer, and accept. The second, the *match*, is to ask the second prover one of the two half questions α_i sent to the first prover and to accept only if the second prover's answer β matches (i.e., is equal to) the half answer β_i given by the first prover. The verifier chooses its question to the second prover uniformly from $\{\alpha_1, \alpha_2, \lambda\}$. For formal definitions, see section 2.

Our main result is an upper bound, for proof systems in the miss-match form, on the number of parallel repetitions that suffice in order to reduce the error from an initial value of p to a desired value of ϵ. This upper bound is polynomial in $1/(1-p)$ and in $1/\epsilon$.

Though our upper bound applies only to proof systems in miss-match form, it can be used in order to reduce the error from one constant to another in any MIP(2,1) proof system. The reason is that any MIP(2,1) proof system can be easily transformed into miss-match form (see Proposition 3.1), with only constant overhead in communication and randomness, and insignificant loss in the error p. In [2] it was shown that any NP-language has an MIP(2,1) proof system with logarithmic-size questions and constant-size answers. The error p for these proof systems is a constant less than 1 but larger than $\frac{1}{2}$. Our main result implies that the error can be reduced to any constant $\epsilon > 0$, while increasing the question and answer sizes by only a multiplicative constant factor. We remark that these error-reducing transformations preserve triviality; on "yes" instances of the proof system the provers can always make the verifier accept.

A major application of MIP(2,1) proof systems for NP is to prove hardness of approximation results. Decreasing the error of an MIP(2,1) proof system often translates into stronger hardness of approximation results (cf. [12, 5, 4]).

Our analysis of parallel repetition of MIP(2,1) proof systems in miss-match form is based in part on an analysis of the influence of random variables on an arbitrary function. This part is presented in a self-contained way in section 4 and may be of independent interest. Essentially, we present an exact formulation of the intuition that for any function, knowing the value of an α-fraction of the input variables (chosen at random) is expected to give at most an α-fraction of the information regarding the output of the function. Here, we measure "information" in terms of the variance of the function. Theorem 4.3 is an essentially optimal version of a lemma that appeared in the conference version of this paper [12]. It was developed with the help of Leonid Gurvits, who gave the first proof, based on Fourier analysis. In section 4, we present a more elementary proof of this theorem.

1.1. Related work. A preliminary version of this paper appeared in [12]. In discussing related work, we distinguish between work done prior to the preliminary version and work done since. We also discuss the improvements of the current version over the preliminary one.

Prior work. MIP(2,1) proof systems were introduced by Ben-Or, Goldwasser, Kilian, and Wigderson [6]. A major result there was a two-prover perfect zero-knowledge proof system for NP. The authors remarked that parallel repetition of these proof systems preserves zero-knowledge properties, but the issue of its effect on the error was not touched upon. Initial beliefs that n parallel repetitions reduce the error from ϵ to ϵ^n were refuted by an explicit example in [15]. Since then, the problem of error reduction for MIP(2,1) proof systems has attracted much attention, because of both the intellectual challenge and the potential applications of these proof systems. Applications were initially made to cryptography [18], and later to the design of efficient probabilistically checkable proofs and to proving hardness results for approximating NP-hard optimization problems [2, 20, 3].

Initially, bounds on the rate by which parallel repetition reduces the error in MIP(2,1) proof systems were obtained only in special cases (see, e.g., [8, 18]). In this respect, the result most related to our current work was the method of [7, 10] that can reduce the error in any MIP(2,1) proof system from $p > 1/2$ to $1/2 + \epsilon$ with an overhead that depends only on p and ϵ. Our work on miss-match proof systems gives a result of similar flavor that extends to arbitrarily small errors.

The most successful approach for reducing the error was through algebraic techniques [9, 19, 13]. These techniques may dramatically change the structure of the original protocol, but thereafter the analysis of error reduction becomes relatively simple and the rate of error reduction becomes exponential. Algebraic techniques became the method of choice for reducing the error in MIP(2,1) proof systems.

Our preliminary version. The preliminary version of our paper [12] had two parts. The first part, on which the current version is based, analyzed the effect of parallel repetition on MIP(2,1) proof systems of a certain form, which we called *confuse-or-compare.* The second part showed how to preserve the zero-knowledge property when algebraic error reduction techniques are applied. We shall not discuss this second part further, nor shall we discuss zero-knowledge.

Confuse-or-compare proof systems give the verifier two options of what to do with the second prover, similar to the case of miss-match proof systems. The "compare" option is identical to the "match" option. The "confuse" option was our original version of the "miss" option. Rather than send the second prover the null question, the verifier sends him a random half question γ, unrelated to α_1 and α_2. As with the miss option, the verifier ignores the second prover's answer if the confuse option is employed. The results presented for the confuse-or-compare proof systems were qualitatively similar to the results presented here for miss-match proof systems, and the analysis was slightly simpler. The reason we switched from confuse-or-compare to miss-match will be explained shortly.

The main point of the results in [12] was that they provided a way of reducing the error from one constant to another with only constant overhead. In contrast, known algebraic error reduction techniques require more than a constant overhead in modifying the original proof system. Using the confuse-or-compare proof systems, previously known hardness of approximation results could be derived under weaker complexity assumptions (e.g., under the assumption that $P \neq NP$ instead of the assumption that NP does not have quasi-polynomial algorithms). Additional results in [12] included further strengthening of hardness of approximation results for *clique* and *chromatic number*, based on observations regarding which parameters of proof systems are really relevant to obtaining such results.

Results obtained since. Concurrently with our preliminary work in [12], Verbitsky proved that for any MIP(2,1) proof system, parallel repetition can reduce the error to be arbitrarily small. However, the method of analysis used by Verbitsky does not give useful bounds on the number of repetitions needed [22]. Subsequently, Raz proved the following parallel repetition theorem, which gives almost tight bounds on the rate of error reduction by parallel repetition.

Parallel repetition theorem (*see* [21]). For any MIP(2,1) proof system with error ϵ there exists some constant $\alpha > 0$ (that depends on ϵ and on the answer length of the proof system) such that the error obtained by n parallel repetitions is at most $\epsilon^{\alpha n}$.

Raz's theorem is strongest when the initial error and answer length are constant, as then α is just some other constant. This is indeed the case in the places where the use of error reduction by transformation to a confuse-or-compare proof system was suggested, and hence Raz's theorem could replace the results of [12] in all suggested applications. Moreover, as was pointed out in [5, 4], Raz's theorem has the further advantage of not introducing the confuse rounds. Certain efficient constructions in the theory of probabilistically checkable proofs [4] (leading to better inapproximability ratios for NP-hard optimization problems) require that the underlying MIP(2,1) proof system have the "projection" property: the answer of the second prover is a function of the answer of the first prover. This indeed holds for compare rounds, but does not hold for confuse rounds. Even though the verifier ignores the answer of the second prover on the confuse rounds, the second prover cannot tell which are the confuse rounds, and for this reason cannot be used by the construction of Bellare, Goldreich, and Sudan [4]. With Raz's theorem, the problem does not arise because the confuse rounds are not needed in order to reduce the error.

Despite the above, the confuse-or-compare structure does offer advantages that cannot be achieved by straight parallel repetition. There are families of MIP(2,1) proof systems for which the number of parallel repetitions required in order to reduce the error from 3/4 to 1/8 is not bounded by any constant [14]. However, as will be discussed in section 3, any MIP(2,1) proof system with error at most 3/4 can be converted to one in miss-match form, with error at most 11/12, and then amplified by parallel repetition to achieve any constant error. This method uses only a constant number of "black box" invocations of the original proof system. Thus, we have an alternative generic amplification method that in some cases works substantially better than parallel repetition.

The current version. The main disadvantage of the confuse-or-compare structure, not having the projection property, is circumvented in the current version of the paper. Here the confuse-or-compare structure is replaced by the miss-match structure, which does have the projection property (as explained in section 2). This makes our results applicable for the construction of [4]. We analyze error reduction by parallel repetition for miss-match proof systems and show that the number of repetitions that is required in order to reduce the error from p to ϵ is some constant that depends only on p and ϵ. Our analysis follows closely our preliminary analysis in [12], with a slight twist toward the end. (The inspiration for this slight twist came from [16].) We also improve our analysis of the influence of random variables on a function. This stronger analysis was developed with the help of Leonid Gurvits and gives essentially tight bounds. (Our previous bounds gave away unnecessary logarithmic terms.) In the current version of the paper, we omit discussion of applications of our results to the construction of efficient probabilistically checkable proofs and to obtaining hardness of approximation results. The reader is referred to [4] for this purpose.

For further discussion on error reduction by parallel repetition, see [11]. For a discussion on the influence of variables on Boolean functions, see [17].

2. Definitions.

Two-prover one-round proof systems are often modeled as a game between a verifier and two cooperating but not communicating provers. It is not an adversarial game, as the strategy of the verifier is fixed in advance. Rather, it is a cooperative game with two players (provers) who coordinate a joint strategy that gives them the highest probability of winning. Modeling MIP(2,1) proof systems as games suppresses the language recognition aspects of these proof systems, but preserves the concept of error in a proof system and the issue of how error can be reduced. One may best think of a game as an instantiation of an MIP(2,1) proof system on a single input that is not in the given language. We use the following notation.

$G = G(X, Y, Q, \pi, A, B, V)$ - a two-prover one-round game;
X - set of questions to prover P_1;
Y - set of questions to prover P_2;
A - set of answers available to P_1;
B - set of answers available to P_2;
π - probability distribution on $X \times Y$;
Q - support of π. $Q \subseteq X \times Y$;
V - acceptance predicate on (X, Y, A, B).

Game G proceeds as follows. The verifier selects at random a question pair $(x, y) \in Q$, according to probability distribution π. Question x is sent to P_1, who replies with an answer $P_1(x) \in A$. Question y is sent to P_2, who replies with an answer $P_2(y) \in B$. (We identify the name of a prover and the strategy that it employs.) The verifier then evaluates a predicate $V(x, y, P_1(x), P_2(y))$ and accepts if the predicate is satisfied. The goal of the provers is to select a strategy (namely, two functions, one for each prover, specifying an answer to each possible question) that maximizes the probability that the verifier accepts. The probability that the verifier accepts under the optimal strategy of the provers is denoted by $\omega(G)$. If $\omega(G) = 1$ the provers are said to have a *perfect strategy* for G, and the game G is *trivial.* For nontrivial games, $\omega(G)$ is also called the *error* of the game. We shall be interested only in nontrivial games.

An *n-fold parallel repetition* of game G is a new game, denoted by G^n, played on n coordinates, corresponding to playing n versions of game G in parallel. The questions and answers are now n-vectors, the ith element corresponding to the ith game. The verifier treats each coordinate of G^n as an independent copy of the original game G and accepts in G^n only if it would have accepted all the n copies of G. The support set of G^n is $Q^n \subseteq X^n \times Y^n$, the answer sets are A^n and B^n. The verifier selects n question pairs $(x_i, y_i) \in Q$ independently, each according to the probability distribution π. We denote $\vec{x} = x_1 x_2 \cdots x_n$ and $\vec{y} = y_1 y_2 \cdots y_n$. (Hence $\vec{x} \in X^n$ and $\vec{y} \in Y^n$.) A strategy for the provers is $2n$ functions, $P_1^i : X^n \longrightarrow A$ and $P_2^i : Y^n \longrightarrow B$, where $1 \leq i \leq n$. The acceptance predicate is $V^n = \bigwedge_{i=1}^n V(x_i, y_i, P_1^i(\vec{x}), P_2^i(\vec{y}))$. That is, the verifier accepts if all n copies of the original game G are accepting.

Observe that even though the verifier treats each coordinate of G^n independently, the provers may not. In particular, the answer a prover gives in coordinate i may depend on the questions that the prover receives in other coordinates. For this reason, it is not true in general that $\omega(G^n) = (\omega(G))^n$.

We shall often refer to the following special classes of games. A game is *uniform* if π is uniform on Q. A game is *free* if it is uniform and has full support, i.e., $Q = X \times Y$. (The term *free* was introduced in [8].) A game has the *projection* property if the question to prover P_2 is a projection of the question to prover P_1, and the verifier accepts only if the answer of prover P_2 is a projection of the answer to prover P_1. More formally, the question sets are of the form $X = X_1 \times X_2 \times \cdots \times X_k$ for some k, and $Y = \bigcup_{i=1}^k X_i$, where the sets X_i are disjoint. The support of the probability distribution π includes only pairs $(x, y) \in X \times Y$ for which y is x_i for some $i \in \{1, 2, \ldots, k\}$, where x_i is the ith component of the question $x = (x_1, x_2, \ldots, x_k)$. The answer sets are of the form $A = A_1 \times A_2 \times \cdots \times A_k$ for the same k as above, and $B = \bigcup_{i=1}^k A_i$, where the sets A_i are disjoint. The acceptance predicate V is a conjunction of two parts $V = V_1 \wedge V_2$. One is a predicate $V_1(X, A)$ that ignores the communication with P_2. The other is $V_2(X, Y, A, B)$, which accepts only if the answer of P_2 is precisely the ith component of the answer of P_1, where i is the same coordinate on which the question y is a projection of the question x. We observe that games with the projection property may be uniform, but cannot be free (unless the cardinality of X is one).

We are now ready to define *miss-match* games. These are games with the projection property, also satisfying the following requirements. The question set to prover P_1 is $X = X_1 \times X_2 \times \{\lambda\}$, where λ is a character which the reader can interpret as saying "a miss." Hence the question set to P_2 is $Y = X_1 \bigcup X_2 \bigcup \{\lambda\}$. The probability distribution π is arbitrary on $X_1 \times X_2$, and uniform with respect to selecting which of the three components of $x \in X$ to choose as y. Hence we shall often refer to π as a probability distribution over $X_1 \times X_2$. The answer set for prover P_1 is $A = A_1 \times A_2 \times \{\lambda\}$. Hence the answer set for P_2 is $B = A_1 \bigcup A_2 \bigcup \{\lambda\}$. Since the third component of the question to P_1 is $\{\lambda\}$ and the third component of his answer is $\{\lambda\}$, the acceptance predicate V_1 is defined only on $(X_1 \times X_2, A_1 \times A_2)$. The full acceptance predicate is $V = V_1 \wedge V_2$, where V_2 is the predicate for the projection property. Observe that the projection property forces P_2 to answer λ with a λ.

When the question to P_2 is λ we say that it is a *miss*. When the question to P_2 is in $X_1 \bigcup X_2$ we say that it is a *match*. When a miss-match game is repeated many times in parallel, some of the rounds will be miss rounds and the others will be match rounds. Though there is no advantage in introducing a miss when the game is played only once, our analysis of the error of the repeated game will make use of the presence of the miss rounds.

3. Preliminaries and main results.

Though miss-match games are a restricted form of games, any other MIP(2,1) game can be easily transformed into a miss-match game. Let $G(X, Y, Q, \pi, A, B, V)$ be an arbitrary game. Then its miss-match version $G' = G'(X', Y', Q', \pi', A', B', V')$ is as follows:

- $X' = X \times Y \times \{\lambda\}$;
- π' is identical to π over $X \times Y$;

- $A' = A \times B \times \{\lambda\}$.

Recall that $V' = V_1 \wedge V_2$. V_1 is identical to V on (X, Y, A, B). The rest of the parameters (Y', Q', B', V_2) are implicit from G' being a miss-match game. Note that if G is trivial, then G' is trivial.

PROPOSITION 3.1. *For the games G and G' as described above, $\omega(G') \leq \frac{\omega(G)+2}{3}$.*

Proof. Recall that in G', prover P_2 receives questions $y' \in X \bigcup Y \bigcup \{\lambda\}$. Fix an arbitrary deterministic strategy for P_2 in G'. (P_2's strategy can be made deterministic without loss of generality.) For every question $x \in X$, this fixes a unique answer $a \in A$, and for every question $y \in Y$, this fixes a unique answer $b \in B$. Observe that the provers in the original game G can follow this strategy, and then the probability that V is satisfied is at most $\omega(G)$. Now fix an arbitrary strategy for P_1, who receives questions $x' \in X \times Y \times \{\lambda\}$. Use the strategies of P_1 and P_2 in order to partition all questions x' to P_1 into two classes. The first class contains those x' for which P_1's answers in $A \times B$ agree with P_2's fixed strategy. The second class contains those x' for which at least one of P_1's two answers in $A \times B$ disagrees with P_2's fixed strategy. Let q denote the probability that x' is from the first class; x' is from the second class with probability $(1 - q)$. The joint probability that x' is from the first class and V_1 is satisfied is at most q and is also at most $\omega(G)$. (Otherwise, the provers for the original game can win with probability higher that $\omega(G)$.) Hence this joint probability is at most $\min[q, \omega(G)] \leq (\omega(G) + 2q)/3$. The joint probability that x' is from the second class and V_2 is satisfied is at most $\frac{2}{3}(1 - q)$: whenever x' is from the second class (probability $(1 - q)$), there is a 1/3 chance of the inconsistent answer being detected. Thus, the acceptance probability is at most $(2 + \omega(G))/3$, the sum of these upper bounds. □

For a miss-match game G', we say that prover P_1 has a *projection strategy* if, for every two questions $x = (x_1, x_2, \lambda)$ and $x' = (x'_1, x'_2, \lambda)$ and their respective answers $a = (a_1, a_2, \lambda)$ and $a' = (a'_1, a'_2, \lambda)$, it holds that $a_1 = a'_1$ whenever $x_1 = x'_1$, and $a_2 = a'_2$ whenever $x_2 = x'_2$. Let the *basic error* $p(G')$ be the maximum probability that V_1 is satisfied, where the maximum is taken over all projection strategies of P_1. For miss-match games, we will find it more convenient to work with p rather than ω. Note that when transforming an arbitrary game G to a miss-match game G', there is a correspondence between deterministic strategies for G and projection strategies for G'. We thus obtain the following simple relation between $p(G')$ and $\omega(G)$.

PROPOSITION 3.2. *For G and G' as above, $p(G') = \omega(G)$.*

Propositions 3.1 and 3.2 imply that $\omega(G') \leq \frac{p(G')+2}{3}$.

The following theorem is our main result.

THEOREM 3.3. *Let G' be an arbitrary miss-match game with basic error $p < 1$. Then $\omega((G')^n) \leq \epsilon$, whenever $n \geq c/((1-p)\epsilon)^c$, where $c \geq 0$ is a universal constant independent of G', p, n, and ϵ.*

Using Proposition 3.2 and the fact that for any miss-match game G', $p(G') \leq \omega(G')$, we can restate Theorem 3.3 in terms of the original win probabilities.

COROLLARY 3.4. *For G' an arbitrary miss-match game, $\omega((G')^n) \leq \epsilon$ when $n \geq c/((1 - \omega(G'))\epsilon)^c$, where $c \geq 0$ is some universal constant independent of G', p, and ϵ. Furthermore, for G an arbitrary nontrivial MIP(2,1) game, $\omega((G')^n) \leq \epsilon$ when $n \geq c/((1 - \omega(G))\epsilon)^c$, where G' is obtained by transforming G as described above.*

We remark that the proof of our theorem is robust enough to support simple modifications to the notion of a miss-match game. (These modifications change only the value of the constant c.) In particular, the requirement that $X = X_1 \times X_2 \times \{\lambda\}$ can be relaxed to $X = X_1 \times X_2 \times \cdots \times X_k \times \{\lambda\}$ for any fixed k, and the requirement that the question to P_2 is chosen uniformly from the $k+1$ parts of the question to P_1 can be relaxed to any other distribution with full support.

To appreciate Theorem 3.3, one should contrast it with the following theorem of [14].

Feige–Verbitsky theorem (*see* [14]). There exists an infinite family $\mathcal{G}$ of free games such that $\omega(G) \leq 3/4$ for every $G \in \mathcal{G}$, and for any n there is some $G \in \mathcal{G}$ with $\omega(G^n) \geq 1/8$.

Hence Theorem 3.3 does not hold for arbitrary games, and there is something special in the miss-match form that makes it work. Miss-match games have two ingredients—the miss (sending an occasional λ to P_2) and the match (the projection property). It turns out that the miss property alone does not suffice in order to prove Theorem 3.3. The constant 1/2 in the following proposition is arbitrary and can be replaced with any other constant.

Proposition 3.5. *Let $\lambda \circ G$ denote a modification of a game G in which, with probability* 1/2, *the question to P_2 is replaced by λ, and the verifier accepts. For any* free *game G, and for any integer $n \geq 1$, $\omega((\lambda \circ G)^n) \geq \omega(G^n)$.*

Proof. Fix an optimal strategy S for P_1 and P_2 in G^n. Based on S, we design a randomized strategy for the provers in $(\lambda \circ G)^n$. P_1 deterministically answers each question of $(\lambda \circ G)^n$ with the answer that P_1 would have given on the same question using strategy S. P_2 uses the following randomized strategy. Independently, in each coordinate on which P_2 receives a λ in $(\lambda \circ G)^n$, prover P_2 replaces λ by a question $y \in Y$ chosen uniformly at random. As a result, the question that P_2 receives in $(\lambda \circ G)^n$ is transformed into a randomly distributed question in G^n. Moreover, the assumption that G is a free game implies that the question pair that P_1 and P_2 now hold is distributed uniformly at random over all question pairs of the game G^n. Hence if the provers use strategy S, then the probability that the verifier of G^n accepts (taken over the choice of question to P_1 and transformed question to P_2) is at least $\omega(G^n)$. Indeed, P_2 answers the transformed question according to strategy S. □

Corollary 3.6. *There exists an infinite family $\mathcal{G}$ of free games such that $\omega(\lambda \circ G) \leq 7/8$ for every $G \in \mathcal{G}$, and for any n there is some $G \in \mathcal{G}$ with $\omega((\lambda \circ G)^n) \geq 1/8$.*

Proof. Apply Proposition 3.5 to the family of games from the Feige–Verbitsky theorem [14] cited above. □

Corollary 3.6 implies that there are families of games that have the miss property yet for which parallel repetition does not yield the error reduction obtained in Theorem 3.3. We were unable to resolve the question of whether the projection property alone makes Theorem 3.3 work or whether the combination of projection and miss is required.

We now give some intuition to explain why having miss rounds facilitates error reduction through parallel repetition. To win several parallel games with high probability, the provers must coordinate their strategies very carefully. A miss round disrupts this coordination: each prover has less understanding about what the other prover knows. Unfortunately, we know of no direct, easily motivated way of analyzing this effect.

We now give some intuition for the proof of Theorem 3.3. First, we actually analyze the error rate of G^{4n}, not G^n; the difference between n and $4n$ is easily swallowed up by an appropriate choice of c. In the analysis, out of the $4n$ rounds of G^{4n}, we concentrate only on n rounds, of which only $m \ll n$ are match rounds and the rest are miss rounds. (In fact, our analysis is somewhat simplified if we start directly with a proof system that has n rounds, of which m rounds chosen at random are match rounds, and $n - m$ rounds are miss rounds.) Oversimplifying, P_1 has two possible modes of behavior. One is to employ a projection strategy for all n rounds, answering each round independently, and within each round using a projection strategy. The other is to base the answer in any particular round on the questions in all other rounds. If P_1 employs a projection strategy, then he is not using the fact that rounds are repeated in parallel, and standard Chernoff bounds imply that the probability of simultaneous success on all rounds is low. If P_1 does not employ a projection strategy, then his answers on the m match rounds highly depend on the questions received on the $n - m$ miss rounds. But prover P_2 receives only a λ on the miss rounds and has no idea what P_1 receives. So P_2 cannot know how to answer the match rounds in a way that indeed produces a match.

The rest of the paper is organized as follows. The proof of the main theorem itself appears in section 6. Prior to the proof, in sections 4 and 5, we present some general properties of multivariate functions. In section 4 we study the influence of a small number of random variables on the value of a function. This is related to the information available to the second prover, who sees only the questions on m match rounds, regarding the output of the first prover on these m rounds, which is a function of the questions of all n rounds. In section 5 we characterize the "gray area" between the two extreme strategies for the first prover in the simplified overview above: the projection strategy and a strategy in which the answer in each round is highly influenced by the questions in all other rounds. We show that in some exact sense, there is no gray area in between.

4. The influence of random variables on a function.

The results of this section were developed with the help of Leonid Gurvits.

Let Q be a finite set and let $x_i \leftarrow \pi_i$ denote a random variable $x_i \in Q$ chosen at random according to probability distribution π_i over Q. Let $\pi = \pi_1 \times \pi_2 \times \cdots \times \pi_n$ be a product distribution over Q^n, and let $x \leftarrow \pi$ denote a random n-vector chosen according to probability distribution π. That is, for each coordinate i of x, $x_i \leftarrow \pi_i$, independently of the other coordinates of x. Let $f : Q^n \to \mathcal{R}^\ell$ be a function whose values are points in ℓ-dimensional real space, and let the *mean* $\mu[f,\pi]$ of f under π denote the expectation of f over the choice of $x \leftarrow \pi$ (which is the center of mass of the points in $\mathcal{R}^\ell$ if probability is interpreted as mass). That is,

$$\mu[f,\pi] \overset{\text{def}}{=} \mathrm{E}_{x\leftarrow\pi}[f(x)] = \sum_x \pi(x)f(x).$$

Similarly, we define the variance $\sigma^2[f,\pi]$ by

$$\sigma^2[f,\pi] \overset{\text{def}}{=} \mathrm{E}_{x\leftarrow\pi}\left[(f(x) - \mu[f,\pi])^2\right] = \mathrm{E}_{x\leftarrow\pi}[f(x)^2] - \mu[f,\pi]^2.$$

Let $[n]$ denote the set of integers from 1 to n. For $m \le n$, we will be interested in m-vectors $\vec{\imath} \in [n]^m$ in which all coordinates are distinct and ascending. Thus, $\vec{\imath}$ denotes an m-element subset of $[n]$. As a convention, $\vec{\imath} \leftarrow [n]^m$ denotes a vector chosen uniformly at random from all vectors of $[n]^m$ with distinct ascending entries. Let $\vec{q}$ denote a vector in Q^m. Let $x[\vec{\imath}]$ denote the projection of x to the coordinates indexed by $\vec{\imath}$ ($x[i]$, for scalar i, is interpreted in the obvious manner).

Definition 4.1. *An m-block $B = (\vec{\imath},\vec{q})$ is a sequence of m pairs (i_j, q_j) (for $1 \le j \le m$), where $q_j \in Q$, $i_j \in [n]$, and $i_j < i_k$ when $j < k$. In our intended use, an m-block specifies the values of m of the input variables for an n-variate function.*

A vector $x \in Q^n$ *satisfies* the m-block B if for every pair $(i_j, q_j) \in B$, $x_{i_j} = q_j$. For a fixed m-block B, let the conditional mean $\mu[f,\pi|B]$ denote the expectation of

$f(x)$ when x is chosen at random according to π, conditioned on x satisfying B. Let $B \overset{m}{\leftarrow} \pi$ denote an m-block $(\vec{\imath}, \vec{q})$ chosen at random, where $\vec{\imath} \leftarrow [n]^m$, and each entry of q_j of $\vec{q}$ is chosen independently according to π_{i_j}.

Clearly, for any function f, the expectation of the conditional mean is the same as the mean. That is, $\mathrm{E}_{B \overset{m}{\leftarrow} \pi} \mu[f, \pi | B] = \mu[f, \pi]$. We shall show that if m is small compared to n, then for most choices of an m-block B, the conditional mean is close to the mean. In other words, the value of x on m random coordinates is unlikely to significantly influence the mean of $f(x)$. Our analysis is essentially tight.

We define the *inner product* of two vectors u and v by $\langle u, v\rangle \overset{\text{def}}{=} \sum u_i v_i$ and the *norm* of vector x by $\|x\| \overset{\text{def}}{=} \sqrt{\langle x, x\rangle}$.

DEFINITION 4.2. *For a function $f : Q^n \to \mathcal{R}^\ell$ and probability distribution π, the* m-variance $\sigma^2[f, \pi, m]$ *is the mean square distance between the mean of f and the conditional mean of f. Formally,*

$$\sigma^2\left[f, \pi, m\right] \overset{def}{=} \mathrm{E}_{B \overset{m}{\leftarrow} \pi}\left[\|\mu\left[f, \pi | B\right] - \mu\left[f, \pi\right]\|^2\right].$$

By the linearity of expectation and of the inner product operation, it can be easily verified that $\mathrm{E}_{B \overset{m}{\leftarrow} \pi}[\|\mu[f, \pi|B] - \mu[f, \pi]\|^2] = \mathrm{E}_{B \overset{m}{\leftarrow} \pi}[\|\mu[f, \pi|B]\|^2] - \|\mu[f, \pi]\|^2$, giving an alternative formulation of the variance which we shall often use. To simplify notation, we use the convention that v^2 is interpreted as $\|v\|^2$ whenever v is a vector.

The m-variance generalizes the notion of the variance in the sense that when $m = n$ (that is, the input as a whole is revealed), $\sigma^2[f, \pi, n] = \mathrm{E}_{x \leftarrow \pi}[\|f(x)\|^2] - \|\mu[f, \pi]\|^2 = \sigma^2[f, \pi]$. Intuitively, the m-variance is related to the influence that a random set of m variables has on the value of a function.

For simple examples of the concept of m-variance, consider the following Boolean functions, where $n > 1$, π is uniform over $\{0, 1\}^n$, and x_i denotes the ith variable (coordinate) of x. The parity function $f(x) = \sum_{i=1}^n x_i (\text{mod } 2)$ has zero m-variance for all $m \le n - 1$, since as long as one variable remains unknown, the parity function remains balanced. The majority function, $f(x) = 1$ iff $\sum_{i=1}^n x_i \ge n/2$, has 1-variance $\Theta(1/n)$, because knowing the value of one variable biases the majority function by $\Theta(1/\sqrt{n})$. More generally, the k-majority function, $f(x) = 1$ iff $\sum_{i=1}^k x_i \ge k/2$, also has 1-variance $\Theta(1/n)$ for any k, because with probability k/n the sampled variable is among the k variables on which the majority is computed, and then it biases their majority by $\Theta(1/\sqrt{k})$.

Note that if the function f depends only on one variable, then $\sigma^2[f, \pi, m] = m\sigma^2[f, \pi]/n$, since a block B influences the value of f if it contains the distinguished variable (which happens with probability m/n), and then it completely determines the value of the function. The following theorem generalizes this to arbitrary functions.

THEOREM 4.3. *Let $f : Q^n \to \mathcal{R}^\ell$ be an arbitrary function and let π be an arbitrary product probability distribution on Q^n. Then for any m, where $1 \le m \le n$, the m-variance of f satisfies*

$$\sigma^2\left[f, \pi, m\right] \le \frac{m}{n}\sigma^2\left[f, \pi\right].$$

Proof. We first concentrate on $f : Q^n \to \mathcal{R}$; we relax this restriction later.

In our computations of expectations and variances, x_i is always distributed according to π_i. We will at times specify a set of variables in the subscript of an expectation {variance}; this denotes taking the expectation {variance} over the choice of all the variables in the subscript, where each variable x_i is independently distributed according to π_i.

Given a subset of the variable indices, $S = \{i_1, \dots, i_k\}$, we define f_S by

$$f_S(x_{i_1}, \dots, x_{i_k}) = \mathrm{E}_{x_j : j \notin S}[f(x_1, \dots, x_n)].$$

That is, f_S gives the expected value of f conditioned on the values of $x_{i_1}, \dots, x_{i_k}$. We define $\pi_S = (\pi_{i_1}, \dots, \pi_{i_k})$ and $\sigma^2[f, \pi, S]$ by

$$\sigma^2[f, \pi, S] \overset{\text{def}}{=} \sigma^2[f_S, \pi_S] = \mathrm{E}_{x_i : i \in S}[f_S(x_{i_1}, \dots, x_{i_k})^2] - \mu[f, \pi]^2, \tag{4.1}$$

where the latter equality follows from $\mu[f, \pi] = \mu[f_S, \pi_S]$.

By definition, $\sigma^2[f, \pi, m] = \mathrm{E}_S[\sigma^2[f, \pi, S]]$, where S is distributed uniformly over m-element subsets of $[n]$. The crux of our proof is to upper bound $\sigma^2[f, \pi, S]$ by a quantity I_S whose expected value (over the choice of S) is $(m/n)\sigma^2[f, \pi]$. Let $P = i_1, \dots, i_n$ be an arbitrary ordering of the variables. We define

$$I_{i_j} = \sigma^2[f, \pi, \{i_1, \dots, i_j\}] - \sigma^2[f, \pi, \{i_1, \dots, i_{j-1}\}]$$

and

$$I_S = \sum_{i \in S} I_i.$$

When P is unclear we write $I_i(P)$. It follows via a telescoping sum that $\sigma^2[f, \pi] = \sum_i I_i$; by the linearity of expectation, it follows that $\mathrm{E}_S[I_S] = (m/n) \sum_i I_i$, hence $\mathrm{E}_S[I_S] = (m/n)\sigma^2[f, \pi]$. The theorem, restricted to the case where the range of f is $\mathcal{R}$, then follows immediately from Lemma 4.4 below.

To extend the proof to the case that the range of f is $\mathcal{R}^\ell$, consider an arbitrary orthonormal basis for $\mathcal{R}^\ell$. Now view f as ℓ different functions $f_1, \dots, f_\ell$, where $f_i(x)$ is the projection of $f(x)$ on the ith basis vector. For each function f_i separately, the range is just $\mathcal{R}$, and hence the theorem holds. As the square of the distance of two points in $\mathcal{R}^\ell$ is the sum of squares of their distances when projected to the ℓ basis vectors, the theorem follows also for f. □

Lemma 4.4. *For any ordering P on $\{1, \dots, n\}$, $\sigma^2[f, \pi, S] \leq I_S$.*

For our purposes, a single P that works for all S would suffice, but considering an arbitrary P will be useful in the proof. Before proving Lemma 4.4, we prove a useful, presumably known inequality on variances.

Proposition 4.5. *Let $f : Q^2 \longrightarrow R$ be an arbitrary function and let (x, y) be distributed according to an arbitrary product probability distribution on Q^2. Then*

$$\sigma_x^2[\mathrm{E}_y[f(x, y)]] \leq \mathrm{E}_y[\sigma_x^2[f(x, y)]].$$

Proof. We first observe that for any function $g(y)$, the function $f'(x, y) = f(x, y) + g(y)$ satisfies $\sigma_x^2[f'(x, y)] = \sigma_x^2[f(x, y)]$ for all y, and $\mathrm{E}_y[f'(x, y)] = \mathrm{E}_y[f(x, y)] + \mathrm{E}_y[g(y)]$ for all x. It follows that $\sigma_x^2[\mathrm{E}_y[f(x, y)]] \leq \mathrm{E}_y[\sigma_x^2[f(x, y)]]$ iff $\sigma_x^2[\mathrm{E}_y[f'(x, y)]] \leq \mathrm{E}_y[\sigma_x^2[f'(x, y)]]$, since both quantities are unchanged by this translation. We therefore assume without loss of generality that for all y, $\mathrm{E}_x[f(x, y)] = 0$; if not, translate by $g(y) = -\mathrm{E}_x[f(x, y)]$. It follows that $\mathrm{E}_x\mathrm{E}_y[f(x, y)] = 0$, and hence that

$$\sigma_x^2[\mathrm{E}_y[f(x, y)]] = \mathrm{E}_x\mathrm{E}_y[f(x, y)^2].$$

Similarly, it follows that

$$\begin{aligned} \mathrm{E}_y[\sigma_x^2[f(x, y)]] &= \mathrm{E}_y\mathrm{E}_x[f(x, y)^2] \\ &= \mathrm{E}_x\mathrm{E}_y[f(x, y)^2]. \end{aligned}$$

Fixing x, we have $\sigma_y^2[f(x,y)] = \mathrm{E}_y[f(x,y)^2] - \mathrm{E}_y[f(x,y)]^2$. Hence, by the positivity of the variance, $\mathrm{E}_y[f(x,y)]^2 \leq \mathrm{E}_y[f(x,y)^2]$. The proposition follows. ◻

Proof of Lemma 4.4. When S is a prefix of P, then by the definitions and a telescoping sum it follows that $\sigma^2[f,\pi,S] = I_S$. To show the inequality for arbitrary S, we make incremental changes $P \to P'$ such that $I_S(P') \leq I_S(P)$, finally obtaining an ordering Q that contains S as a prefix. Then, $I_S(Q) = \sigma^2[f,\pi,S]$ and $I_S(Q) \leq I_S(P)$, implying the lemma.

For ease of exposition, we assume without loss of generality that $P = 1,\ldots,n$. We consider the ordering

$$P' = 1,\ldots,i-2,i,i-1,i+1,\ldots,n,$$

where $i \in S$ and $i-1 \notin S$. That is, we move a variable indexed by S one step closer to the beginning of the ordering, moving it past a variable that is not in S. Clearly, a sequence of such operations can be used to move all the variables of S to the beginning of the sequence. It remains to show that $I_S(P') \leq I_S(P)$. Clearly, $I_j(P') = I_j(P)$ for $j \notin \{i-1,i\}$. It thus suffices to show that $I_i(P') \leq I_i(P)$, since I_{i-1} does not affect I_S.

Let $A = \{1,\ldots,i\}$, $B = \{1,\ldots,i-1\}$, $C = \{1,\ldots,i-2,i\}$, and $D = \{1,\ldots,i-2\}$. Expanding out the definitions, we obtain,

$$\begin{aligned} I_i(P) = &\left(\mathrm{E}_{x_1,\ldots,x_{i-2}}\mathrm{E}_{x_{i-1}}\mathrm{E}_{x_i}[f_A(x_1,\ldots,x_i)^2] - \mu[f,\pi]^2\right)\\ &- \left(\mathrm{E}_{x_1,\ldots,x_{i-2}}\mathrm{E}_{x_{i-1}}[f_B(x_1,\ldots,x_{i-1})^2] - \mu[f,\pi]^2\right). \end{aligned}$$

Given $x_1,\ldots,x_{i-2}$, we define $g(x_{i-1},x_i) = f_A(x_1,\ldots,x_i)$. Noting that

$$f_B(x_1,\ldots,x_{i-1}) = \mathrm{E}_{x_i}[f_A(x_1,\ldots,x_i)],$$

we obtain,

$$I_i(P) = \mathrm{E}_{x_1,\ldots,x_{i-2}}\mathrm{E}_{x_{i-1}}[\sigma_{x_i}^2[g(x_{i-1},x_i)]].$$

By similar manipulations, we obtain

$$\begin{aligned} I_i(P') = &\left(\mathrm{E}_{x_1,\ldots,x_{i-2}}\mathrm{E}_{x_i}[f_C(x_1,\ldots,x_{i-2},x_i)^2] - \mu[f,\pi]^2\right)\\ &- \left(\mathrm{E}_{x_1,\ldots,x_{i-2}}[f_D(x_1,\ldots,x_{i-2})^2] - \mu[f,\pi]^2\right)\\ = &\ \mathrm{E}_{x_1,\ldots,x_{i-2}}[\sigma_{x_i}^2[\mathrm{E}_{x_{i-1}}[g(x_{i-1},x_i)]]]. \end{aligned}$$

However, by Proposition 4.5 it follows that for all $x_1,\ldots,x_{i-2}$,

$$\sigma_{x_i}^2[\mathrm{E}_{x_{i-1}}[g(x_{i-1},x_i)]] \leq \mathrm{E}_{x_{i-1}}[\sigma_{x_i}^2[g(x_{i-1},x_i)],$$

implying the lemma. ◻

Theorem 4.3 has the following useful corollary, that we shall use repeatedly in subsequent parts of the paper.

COROLLARY 4.6. *Let Q and A be finite sets, let $f : Q^n \to A$ be an arbitrary function, and let π be an arbitrary product probability distribution on Q^n. Then the influence of a random m-block on the probability that $f(x)$ attains any particular value $a \in A$ is bounded as follows:*

$$\mathrm{E}_{B \overset{m}{\leftarrow} \pi}\left[\sum_{a\in A}\left(\mathrm{Pr}_{x\leftarrow(Q^n,\pi)}\left[f\left(x\right)=a\right] - \mathrm{Pr}_{x\leftarrow(\pi|B)}\left[f\left(x\right)=a\right]\right)^2\right] \leq m/n.$$

Proof. By the linearity of expectation, it suffices to show that

$$\sum_{a\in A} \mathrm{E}_{B \overset{m}{\leftarrow} \pi} \left[\left(\Pr_{x\leftarrow\pi} \left[f\left(x\right) = a\right] - \Pr_{x\leftarrow(\pi|B)} \left[f\left(x\right) = a\right] \right)^2 \right] \leq m/n. \tag{4.2}$$

For each $a \in A$, define $f_a : Q^n \to \{0,1\}$ by $f_a(x) = 1$ if $f(x) = a$ and $f_a(x) = 0$ otherwise. For any distribution D, we have $\mu[f_a, D] = \Pr_{x\leftarrow D}[f(x) = a]$. By Definition 4.2, we can write the left-hand side of (4.2) as $\sum_{a\in A} \sigma^2[f_a, \pi, m]$. By Theorem 4.3, $\sigma^2[f_a, \pi, m] \leq \frac{m}{n}\sigma^2[f_a, \pi]$. Now, $\sigma^2[f_a, \pi] = p_a(1 - p_a) \leq p_a$, where $p_a = \Pr_{x\leftarrow\pi}[f(x) = a]$, so

$$\sum_{a\in A} \sigma^2[f_a, \pi, m] \leq \frac{m}{n} \sum_{a\in A} p_a = \frac{m}{n}. \qquad \square$$

5. Correlations within multivalued functions. Let Q and A be finite sets, let $f : Q^n \to A^n$ be a function, and let π be a probability distribution over Q, extended as a product distribution π^n over Q^n. Recall that $\vec{\imath} \in [n]^m$ denotes a vector of m distinct, ascending entries in $\{1, \ldots, n\}$, the notion of an m-block $(\vec{\imath}, \vec{q})$, and the projection notation from section 4. Let $\Pr[\vec{a}|(\vec{\imath}, \vec{q})]$ denote the probability that $f(x)[\vec{\imath}] = \vec{a}$, conditioned on $x[\vec{\imath}] = \vec{q}$ (i.e., the probability is taken over $x \leftarrow (Q^n, \pi^n|(\vec{\imath}, \vec{q}))$). This notation is extended in a natural way to $\Pr[\vec{a}|(\vec{\imath}, \vec{q})(i, q)]$ (in which case the conditioning is done on one coordinate more than the projection), and to $\Pr[\vec{a}, a|(\vec{\imath}, \vec{q})(i, q)]$ for $a \in A$. Throughout we use the convention that i has to be distinct from all coordinates of $\vec{\imath}$.

Definition 5.1. *For a parameter* $\varepsilon > 0$, *the m-triple* $(\vec{\imath}, \vec{q}, \vec{a})$ *is* alive *if* $\Pr[\vec{a}|(\vec{\imath}, \vec{q})] \geq \varepsilon$. *The m-block* $(\vec{\imath}, \vec{q})$ *is* alive *if for some* $\vec{a}$ *the m-triple* $(\vec{\imath}, \vec{q}, \vec{a})$ *is alive.*

Definition 5.2. *An m-block* $(\vec{\imath}, \vec{q})$ $(1 - \eta)$-determines *a pair* (i, q) *if for every live m-triple* $(\vec{\imath}, \vec{q}, \vec{a})$ *there exists* $a \in A$ *such that*

$$\Pr\left[\vec{a}, a|\left(\vec{\imath}, \vec{q}\right)\left(i, q\right)\right] \geq (1 - \eta) \Pr\left[\vec{a}|\left(\vec{\imath}, \vec{q}\right)\left(i, q\right)\right].$$

The above inequality may be thought of in terms of conditional probability. Conditioned on $x[\vec{\imath}, i] = \vec{q}, q$ and $f(x)[\vec{q}] = \vec{a}$, $f(x)[i] = a$ with probability at least $(1 - \eta)$.

Definition 5.3. *An m-block* $(\vec{\imath}, \vec{q})$ *is* $(1 - \eta)$-good *if it is alive and* $(1 - \eta)$-determines *a randomly chosen pair* (i, q) $(i \leftarrow ([n] - \vec{\imath})$ *and* $q \leftarrow \pi)$ *with probability at least* $1 - \eta$.

It is useful to consider the "majority answer" guess for the value of $f(x)$ at some coordinate, given partial information about x and $f(x)$.

Definition 5.4. *Given* $(\vec{\imath}, \vec{q}, \vec{a})$, *we define* $\mathrm{maj}(i, q)$ *to be the most likely (breaking ties lexicographically) value for* $f(x)[i]$, *over the choice of* x, *conditioned on* $x[\vec{\imath}] = \vec{q}$, $f(x)[\vec{\imath}] = \vec{a}$, *and* $x[i] = q$.

A good block $(\vec{\imath}, \vec{q})$ has the property that for any live $(\vec{\imath}, \vec{q}, \vec{a})$ a large fraction of the other answers are expected to agree with the majority answer defined above.

The following proposition will be used in the proof of our main theorem.

Proposition 5.5. *Suppose that* $(\vec{\imath}, \vec{q})$ *is* $(1 - \eta)$*-good, for* $\eta < \frac{1}{2}$, *and that* $(\vec{\imath}, \vec{q}, \vec{a})$ *is alive. Then* $f(x)[i] = \mathrm{maj}(i, q)$ *with probability at least* $1 - 2\eta/\varepsilon$ *in the following experiment:*

1. *Choose* x *according to* π^n, *conditioned on* $x[\vec{\imath}] = \vec{q}$ *and* $f(x)[\vec{\imath}] = \vec{a}$.
2. *Choose* i *uniformly from* $[n] - \vec{\imath}$.
3. *Let* $q = x[i]$.

Proof. Consider the experiment of choosing (i, q) by $i \leftarrow ([n] - \vec{\imath})$, $q = x[i]$, conditioned on $x[\vec{\imath}] = \vec{q}$ but *not* conditioned on $f(x)[\vec{\imath}] = \vec{a}$. Since $(\vec{\imath}, \vec{q})$ is good, by Definition 5.3, (i, q) fails to be $(1-\eta)$-determined by $(\vec{\imath}, \vec{q})$ with probability at most η. The probability that it fails to be $(1-\eta)$-determined, conditioned on $f(x)[\vec{\imath}] = \vec{a}$, is at most $\eta/\Pr(f(x)[\vec{\imath}] = \vec{a}) \leq \eta/\varepsilon$. Hence, (i, q) is $(1-\eta)$-determined with probability $1 - \eta/\varepsilon$, conditioned on $x[\vec{\imath}] = \vec{q}$ and $f(x)[\vec{\imath}] = \vec{a}$. By Definition 5.2, there is some a such that under this same conditioning (and additionally conditioning on $x[i] = q$) $f(x)[i] = a$ with probability at least $(1-\eta)$; by the definition of maj and the above, we have that $f(x)[i] = \mathrm{maj}(i, q)$ with probability at least

$$(1 - \eta/\varepsilon)(1-\eta) \geq 1 - 2\eta/\varepsilon. \qquad \square$$

Intuitively, for any function $f : Q^n \to A^n$, either the value of the projection of $f(x)$ on a random coordinate can be guessed with high confidence by looking only at the same coordinate of x, or else, when one looks at the value of x at m random coordinates, errors build up, making it improbable to simultaneously guess correctly the projection of $f(x)$ on the respective m coordinates. The following lemma builds upon this intuition.

LEMMA 5.6. *For $m < n/2$, $n > 2^{10}\eta^{-4}\varepsilon^{-8}$, $\eta \leq \frac{1}{2}$, and $m > 2^5\eta^{-2}\varepsilon^{-4}$, there exists a* good *block size k, $1 \leq k \leq m$, such that one of the following two conditions holds:*

1. *The probability that a random k-block $(\vec{\imath}, \vec{q}) \overset{k}{\leftarrow} \pi^n$ is alive is at most ε.*
2. *If one chooses $(\vec{\imath}, \vec{q})$ at random from the live k-blocks, $(\vec{\imath}, \vec{q})$ will be $(1-\eta)$-good with probability at least $(1-\varepsilon)$. That is,*

$$\Pr_{(\vec{\imath},\vec{q}) \overset{k}{\leftarrow} \pi^n}\left[(\vec{\imath}, \vec{q}) \textit{ is } (1-\eta)\textit{-good} \mid (\vec{\imath}, \vec{q}) \textit{ live}\right] \geq 1 - \varepsilon.$$

Proof. For each j, $1 \leq j \leq n/2$, let E_j denote the following expectation:

$$E_j = \mathrm{E}_{(\vec{\imath},\vec{q})}\left[\sum_{\vec{a} \in A^j} \Pr\left[\vec{a} \middle| (\vec{\imath}, \vec{q})\right]^2\right],$$

where the expectation is taken over the choice of random j-block $(\vec{\imath}, \vec{q})$. For every j, $0 < E_j \leq 1$.

Let us give some insight on what E_j is measuring. Given $(\vec{\imath}, \vec{q})$, the summation term measures the predictability of $\vec{a}$: if $\vec{a}$ is completely determined by $(\vec{\imath}, \vec{q})$, the summation will be 1; if $\vec{a}$ is uniformly distributed with N possible values, the summation will be $1/N$. There are two competing considerations governing the relationship between E_j and E_{j+1}. Having more questions, and hence more answers, tends to fragment the set of possible answer vectors, causing E_j to be larger than E_{j+1}. However, seeing more questions may give so much information about the answers that the answer vectors are less fragmented. For example, if questions (q_1, q_2) are answered by (q_2, q_1), then E_1 may be quite small and $E_2 = 1$. We will show that, for our choice of parameters, if j is not a good block size, then the former consideration dominates; quantitatively, $E_j - E_{j+1} \geq 1/m$. Hence if there is no good block size between 1 and m, then $E_1 - E_{m+1} \geq 1$, a contradiction.

In order to lower bound $E_j - E_{j+1}$, we introduce a hybrid quantity:

$$E_j^* = \mathrm{E}_{(\vec{\imath},\vec{q})(i,q)}\left[\sum_{\vec{a} \in A^j} \Pr\left[\vec{a} \middle| (\vec{\imath}, \vec{q})\,(i, q)\right]^2\right],$$

where the expectation is taken over the choice of random j-block $(\vec{\imath}, \vec{q})$ and random pair (i, q) $(i \notin \vec{\imath})$.

Let us give some more insight. E_j^* measures the average predictability of $\vec{a}$ given $(\vec{\imath}, \vec{q})$ and an additional pair (i, q). Our proof proceeds as follows. First, we show that E_j^* is not much larger that E_j. The idea is that a random (i, q) is not likely to have much influence on $\vec{a}$, and hence knowing (i, q) will not make $\vec{a}$ that much more predictable on average. We then show that if j is not a good block size, then E_{j+1} is significantly smaller than E_j^*. That is, given $(\vec{\imath}, \vec{q})$ and (i, q), $(\vec{a}, a)$ is significantly harder to predict on average than just $\vec{a}$. Combining these relations between E_j, E_j^*, and E_{j+1}, we show that E_{j+1} is significantly smaller than E_j when $j \leq n/2$ is not a good block size.

It follows from the above definitions that

$$E_j^* - E_j = \mathrm{E}_{(\vec{\imath},\vec{q})}\left[\mathrm{E}_{(i,q)}\left[\sum_{\vec{a}\in A^j} \Pr\left[\vec{a}|\left(\vec{\imath},\vec{q}\right)(i,q)\right]^2 - \Pr\left[\vec{a}|\left(\vec{\imath},\vec{q}\right)\right]^2\right]\right] \tag{5.1}$$

and

$$E_j^* - E_{j+1} = \mathrm{E}_{(\vec{\imath},\vec{q})(i,q)}\left[\sum_{\vec{a}\in A^j}\left(\Pr\left[\vec{a}|\left(\vec{\imath},\vec{q}\right)(i,q)\right]^2 - \sum_{a\in A}\Pr\left[\vec{a},a|\left(\vec{\imath},\vec{q}\right)(i,q)\right]^2\right)\right]. \tag{5.2}$$

Proposition 5.7. *For any block size $j \leq n/2$ (whether good or not),*

$$E_j^* - E_j \leq \frac{2}{\sqrt{n-j}} \leq \sqrt{\frac{8}{n}}$$

.

Proof. Fix a j-block $(\vec{\imath}, \vec{q})$. The value of $\vec{a}$ is now a function of $n - j$ variables. Let ℓ denote the cardinality of A and consider the vector $v \in \mathcal{R}^{\ell^j}$ of probabilities $\Pr[\vec{a}|(\vec{\imath},\vec{q})]$. Then $\|v\|^2 = \sum_{\vec{a}\in A^j}(\Pr[\vec{a}|(\vec{\imath},\vec{q})])^2 \leq 1$. Now fix also a pair (i, q) and consider the vector $u \in \mathcal{R}^{\ell^j}$ of probabilities $\Pr[\vec{a}|(\vec{\imath},\vec{q})(i,q)]$. Then $\|u\|^2 = \sum_{\vec{a}\in A^j}\Pr[\vec{a}|(\vec{\imath},\vec{q})(i,q)]^2 \leq 1$, and by the triangle inequality, $\|u + v\| \leq 2$. Moreover,

$$\begin{aligned}\sum_{\vec{a}\in A^j}\left(\Pr\left[\vec{a}|\left(\vec{\imath},\vec{q}\right)(i,q)\right]^2 - \Pr\left[\vec{a}|\left(\vec{\imath},\vec{q}\right)\right]^2\right) &= \|u\|^2 - \|v\|^2 = \langle(u+v),(u-v)\rangle \\ &\leq \|u+v\|\cdot\|u-v\| \leq 2\|u-v\|.\end{aligned}$$

By convexity and Corollary 4.6 (with $m = 1$) we obtain

$$\mathrm{E}_{(i,q)}\left[\|u-v\|\right]^2 \leq \mathrm{E}_{(i,q)}\left[\|u-v\|^2\right] \leq \frac{1}{n-j}.$$

Combining this with the previous inequality yields

$$\mathrm{E}_{(i,q)}\left[\sum_{\vec{a}\in A^j}\left(\Pr\left[\vec{a}|\left(\vec{\imath},\vec{q}\right)(i,q)\right]^2 - \Pr\left[\vec{a}|\left(\vec{\imath},\vec{q}\right)\right]^2\right)\right] \leq \frac{2}{\sqrt{n-j}}.$$

The proposition follows by taking expectations over $(\vec{\imath}, \vec{q})$. □

PROPOSITION 5.8. *If block size $j \leq n/2$ is not good, then $E_j^* - E_{j+1} \geq \eta^2\varepsilon^4/8$.*

Proof. Clearly, for any j-block $(\vec{\imath}, \vec{q})$ and any pair (i, q) with $i \notin \vec{\imath}$, and any $\vec{a} \in A^j$,

$$\Pr\left[\vec{a} \middle| (\vec{\imath}, \vec{q})\,(i,q)\right]^2 \;\geq\; \sum_{a \in A} \Pr\left[\vec{a}, a \middle| (\vec{\imath}, \vec{q})\,(i,q)\right]^2 . \tag{5.3}$$

When(i, q) is not $(1-\eta)$-determined by $(\vec{\imath}, \vec{q})$, then strict inequality holds. We quantify the effect of these inequalities. First, we show that with a reasonably large probability, some useful events occur.

PROPOSITION 5.9. *Suppose that $j \leq n/2$ is not a good block size. Let $(\vec{\imath}, \vec{q})$ be a randomly chosen j-block and let (i, q) be a randomly chosen pair with $i \notin \vec{\imath}$. Then with probability at least $\eta\varepsilon^2/2$, there will exist some $\vec{a}$ such that*

1. *(For all $a \in A$)*$\Pr[\vec{a}, a|(\vec{\imath}, \vec{q})(i, q)] \leq (1-\eta)\Pr[\vec{a}|(\vec{\imath}, \vec{q})(i, q)]$*, and*
2. $\Pr[\vec{a}|(\vec{\imath}, \vec{q})(i, q)] \geq \varepsilon/2$.

We defer the proof of Proposition 5.9 and proceed to analyzing its consequences. We first use a straightforward convexity argument to show that for $(\vec{\imath}, \vec{q})$, (i, q), and $\vec{a}$ satisfying the conclusions of Proposition 5.9,

$$\sum_{a \in A} \Pr\left[\vec{a}, a \middle| (\vec{\imath}, \vec{q})\,(i,q)\right]^2 \leq \left(1 - 2\eta + 2\eta^2\right) \cdot \Pr\left[\vec{a} \middle| (\vec{\imath}, \vec{q})\,(i,q)\right]^2 .$$

In general, if one wants to maximize $X_1^2 + \cdots + X_k^2$ subject to $X = X_1 + \cdots + X_k$, where all quantities are positive, it always helps to subtract from smaller values in order to add to larger values. Given the added constraint that $X_i \leq (1-\eta)X$, the maximum is attained by setting $X_1 = (1-\eta)X$, $X_2 = \eta X$, and all the rest equal to 0, giving $X_1^2 + \cdots + X_k^2 = (1 - 2\eta + 2\eta^2)X^2$.

This implies (when $\eta \leq 1/2$) that

$$\Pr\left[\vec{a} \middle| (\vec{\imath}, \vec{q})\,(i,q)\right]^2 - \sum_{a \in A} \Pr\left[\vec{a}, a \middle| (\vec{\imath}, \vec{q})\,(i,q)\right]^2 \geq \eta \cdot \Pr\left[\vec{a} \middle| (\vec{\imath}, \vec{q})\,(i,q)\right]^2 \geq \eta \cdot (\varepsilon/2)^2. \tag{5.4}$$

By (5.3), (5.1), and (5.4), each $((\vec{\imath}, \vec{q}), (i, q))$ that satisfies the condition of Proposition 5.9 contributes at least $\eta(\varepsilon/2)^2 = \eta\varepsilon^2/4$ times the probability of the event occurring to $E_j^* - E_{j+1}$. Thus the good events contribute at least $(\eta\varepsilon^2/2)(\eta\varepsilon^2/4) = \eta^2\varepsilon^4/8$ to $E_j^* - E_{j+1}$. (Recall that all the other $((\vec{\imath}, \vec{q}), (i, q))$ terms contribute nonnegatively, by (5.3).) □

Proof of Proposition 5.9. The statement of the proposition considers $\vec{a}$ such that (among other properties) $\Pr[\vec{a}|(\vec{\imath}, \vec{q})(i, q)] \geq \varepsilon/2$. We first show that if we modify this property, asking instead that $\Pr[\vec{a}|(\vec{\imath}, \vec{q})] \geq \varepsilon$, then such $\vec{a}$ exist with sufficiently high probability. We then show it unlikely that an $\vec{a}$ exists that has the modified property but not the original property.

If j is not a good block size, then the probability that $(\vec{\imath}, \vec{q})$ is alive is at least ε, and conditioned on being alive, the probability that $(\vec{\imath}, \vec{q})$ is not good is again at least ε. Thereafter, the probability that $(\vec{\imath}, \vec{q})$ does not $(1-\eta)$-determine a random (i, q) is at least η. Assume that all three of these events occur (this has probability at least $\eta\varepsilon^2$). Then because $(\vec{\imath}, \vec{q})$ is alive and $(\vec{\imath}, \vec{q})$ does not $(1-\eta)$-determine (i, q), there is some $\vec{a}$ such that $\Pr[\vec{a}|(\vec{\imath}, \vec{q})] \geq \varepsilon$ and, for every $a \in A$,

$$\Pr\left[\vec{a}, a \middle| (\vec{\imath}, \vec{q})\,(i,q)\right] \leq (1-\eta) \Pr\left[\vec{a} \middle| (\vec{\imath}, \vec{q})\,(i,q)\right] .$$

It remains to bound the probability that for some $\vec{a}$, $\Pr[\vec{a}|(\vec{\imath},\vec{q})] \geq \varepsilon$ but $\Pr[\vec{a}|(\vec{\imath},\vec{q})(i,q)] < \varepsilon/2$. We say that $((\vec{\imath},\vec{q}),(i,q))$ is *discardable* if, for some $\vec{a}$, $\Pr[\vec{a}|(\vec{\imath},\vec{q})] - \Pr[\vec{a}|(\vec{\imath},\vec{q})(i,q)] \geq \varepsilon/2$. Note that if $\Pr[\vec{a}|(\vec{\imath},\vec{q})] \geq \varepsilon$ and $((\vec{\imath},\vec{q}),(i,q))$ is not discardable, then $\Pr[\vec{a}|(\vec{\imath},\vec{q})(i,q)] \geq \varepsilon/2$. We show that discardable $((\vec{\imath},\vec{q}),(i,q))$ occur with probability at most $\eta\varepsilon^2/2$; combined with the above bound, this implies the proposition.

Consider a live j-block $(\vec{\imath},\vec{q})$ and define a function $g : Q^{n-j} \to A^j$ that receives as an argument the value of an input x on the coordinates other than $\vec{\imath}$, and outputs the value of $f(x)$ on the coordinates indexed by $\vec{\imath}$. By Corollary 4.6,

$$\mathrm{E}_{(i,q)\leftarrow(Q,\pi)}\left[\sum_{\vec{a}\in A^j}\left(\Pr_{x\leftarrow(Q^n,\pi^n|(\vec{\imath},\vec{q}))}\left[g\left(x\right)=\vec{a}\right]-\Pr_{x\leftarrow(Q^n,\pi^n|(\vec{\imath},\vec{q})(i,q))}\left[g\left(x\right)=\vec{a}\right]\right)^2\right]$$
$$\leq 1/\left(n-j\right) \leq 2/n.$$

Using Markov's inequality and simple manipulations, it follows that the probability (over the choice of (i,q)) that there is any $\vec{a}$ with $\Pr[\vec{a}|(\vec{\imath},\vec{q})]-\Pr[\vec{a}|(\vec{\imath},\vec{q})(i,q)] \geq \varepsilon/2$ is at most $(2/n)/(\varepsilon/2)^2 = 8/\varepsilon^2 n$. For $n > 16/\eta\varepsilon^4$ this probability is at most $\eta\varepsilon^2/2$. □

Finally, we conclude the proof of Lemma 5.6. From Propositions 5.7 and 5.8 we obtain that $E_j - E_{j+1} \geq \eta^2\varepsilon^4/8 - \sqrt{\frac{8}{n}}$. For $j \leq m \leq n/2$ and $n > 2^{10}\eta^{-4}\varepsilon^{-8}$, $E_j - E_{j+1} \geq \eta^2\varepsilon^4/32$. For $m > 2^5\eta^{-2}\varepsilon^{-4}$, this difference is more than $1/m$ and the lemma follows. □

6. Proof of the main theorem.

For notational convenience we write the initial miss-match game as G rather than G'. For ease of exposition, rather than consider G^n, we shall consider a related game that we denote by $\hat{G}$, which is derived from G^n by conditioning that there be exactly m match rounds. Hence $\hat{G}$ contains n parallel rounds of G, with the restriction of having exactly m match rounds and $n-m$ miss rounds, where $m < n/2$ will be determined later. The locations of the match rounds is random. Proposition 6.1 implies that upper bounding the error for $\hat{G}$ will suffice to prove our main theorem, though with a different value for c (that "swallows up" the difference between $4n$ and n).

Proposition 6.1. *For $\hat{G}$ as described above, $\omega(\hat{G}) \geq \omega(G^{4n}) - 4e^{-n/32}$.*

Proof. We describe a probability measure on strategies $P = (p_1, p_2)$ for the provers in $\hat{G}$. Each strategy corresponds to a particular way in which the game $\hat{G}$ can be completed to a game G^{4n}. Let p_k denote the probability that G^{4n} has exactly k match rounds, conditioned on it having at least m match rounds and at least $n-m$ miss rounds (the latter condition is equivalent to having at most $3n+m$ match rounds).

We can view such a completion as the following random process:

1. The provers uniformly choose a set $S \subset [4n]$ indexing the n rounds into which they place the rounds of $\hat{G}$.
2. The provers choose $k \in [m, 3n+m]$ according to the probability measure $\Pr[k] = p_k$.
3. The provers uniformly choose a set $M \subseteq [4n] \setminus S$ of size $k-m$. For each round indexed by M the provers agree on the questions for match rounds, generated according to G.
4. For all the remaining rounds (not indexed by S or M), the provers agree on the questions for miss rounds, generated according to G.

The provers, prior to the start of the protocol, jointly sample one member P from the space of completion strategies.

When a prover receives a list of n questions in the game $\hat{G}$, the prover uses the agreed upon completion to embed these as part of a list of $4n$ questions in the game G^{4n}; then the prover employs the optimal strategy of the game G^{4n} and sends back only the n answers that correspond to the original questions of $\hat{G}$. It can readily be seen that the probability that V accepts is at least the probability that the optimal strategy for G^{4n} wins, given the distribution on the questions. This distribution is the "usual" distribution for G^{4n}, conditioned on there being at least m match rounds and $n - m$ miss rounds. By standard Chernoff bounds, the probability that there are less than $n - m$ miss rounds or less than m match rounds out of $4n$ rounds is at most $4e^{-n/32}$; this probability is negligible compared to ϵ when n is sufficiently large. Hence the expected probability of winning is at least $\omega(G^{4n}) - 4e^{-n/32}$, where expectation is taken over the random choice of completion strategy and random questions for $\hat{G}$.

Finally, by averaging, there is at least one deterministic strategy P that ensures probability of acceptance of at least $\omega(G^{4n}) - 4e^{-n/32}$ in the game $\hat{G}$. □

We now prove Theorem 3.3 by showing that $\omega(\hat{G}) \leq \epsilon$.

Proof. Fix a (deterministic) strategy (p_1, p_2) for $\hat{G}$ with highest probability of success for the two provers. This strategy is a pair of functions $p_1 : (X_1 \times X_2)^n \to (A_1 \times A_2)^n$ and $p_2 : (X_1 \bigcup X_2 \bigcup \{\lambda\})^n \to (A_1 \bigcup A_2)^m$. Observe that once p_1 is fixed, the function p_2 is characterized using the maximum likelihood principle. That is, on seeing the sequence $\vec{q}$ of n questions, it is best for P_2 to consider all possible sequences of n questions to P_1 that are consistent with P_2 receiving $\vec{q}$, compute the answers of P_1 on the match rounds for each of these sequences, and then answer the match rounds of $\vec{q}$ in a way identical to the most likely answer sequence by P_1 (breaking ties arbitrarily).

Let $Q = X_1 \times X_2 \times \{1, 2\}$, let $A = A_1 \bigcup A_2$, and let π' be a probability distribution over Q that agrees with π over $X_1 \times X_2$ and is uniform over $\{1, 2\}$ (that is, choose the first two coordinates according to π and then choose the last coordinate uniformly from $\{1, 2\}$). Based on p_1 we define the function $f : Q^n \to A^n$, described as follows. For an input in Q^n, first strip off the last component of each coordinate, obtaining an input in $(X_1 \times X_2)^n$. Then compute the output of p_1 on the stripped input. This will be a vector in $(A_1 \times A_2)^n$. Finally, for each coordinate in this output vector, save only one of its two components (either the one in A_1 or the one in A_2), based on the respective value of the $\{1, 2\}$ component of the original input vector. The function f corresponds to what P_2 would need to respond if all rounds were match rounds. In $\hat{G}$ some rounds are miss rounds, and our analysis will concentrate on coordinates of f that correspond to match rounds.

Recall the notion of a k-block from section 4. Assume that $m < n/2$. (We will enforce this condition in our choice of m.) Applying Lemma 5.6 with respect to f and probability distribution $(\pi')^n$, we have a *good* block of size $k \leq m$ in the sense defined in Lemma 5.6. (We shall expand on this below.) Fix this k for the rest of the proof.

For the purpose of our proof, we describe a three-step procedure by which the verifier selects the questions to the two provers in the n rounds. It can easily be verified that this three-step procedure induces the correct probability distribution on the questions.

Step 1. Select at random a k-block $(\vec{\imath}, \vec{q}) \overset{k}{\leftarrow} (\pi')^n$ for f. That is, select at random k distinct rounds, and in each such round select a question pair from $X_1 \times X_2$ with probability π (to be sent to P_1), and an index uniformly from $\{1, 2\}$ specifying which half is sent to P_2. This specifies k match rounds.

Step 2. From the remaining rounds, select at random an $(m-k)$-block, B, specifying $m-k$ additional match rounds.

Step 3. Make the remaining $n-m$ rounds miss rounds. Select the questions to P_1 for these $n-m$ rounds. (P_2 receives λ on these rounds.)

To outline the rest of the analysis, we describe five events, at least one of which will occur when the verifier accepts. For each event, we bound the probability that the event occurs and that the verifier accepts. Although these events are not independent, we can conservatively bound the probability that the verifier accepts by the sum of the individual bounds.

Consider the k-block $(\vec{\imath}, \vec{q})$ selected by the verifier in Step 1, and consider the eventual answer $\vec{a}$ that P_2 gives on his respective half questions on the rounds specified by $\vec{\imath}$. Based on $(\vec{\imath}, \vec{q})$ and on function f, consider a new function $g : Q^{n-k} \rightarrow A^k$ that gives the value of f on the k coordinates specified by $\vec{\imath}$ as a function of the remaining $n-k$ coordinates of the input. For the verifier to accept, P_2 must give the same answers on these half questions. There are two cases:

1. The answer $\vec{a}$ of P_2 was an unlikely answer for P_1, in the sense that $\Pr_{x \leftarrow (\pi')^{n-k}}[g(x) = \vec{a}] < \varepsilon$, where $\varepsilon < \epsilon$ will be chosen later. There are two subcases to consider, depending on whether $\vec{a}$ continues to be an unlikely answer even after the $(m-k)$-block B is chosen.
   (a) $\Pr_{x \leftarrow ((\pi')^{n-k}|B)}[g(x) = \vec{a}] < 2\varepsilon$. This subcase is handled by Event 1.
   (b) $\Pr_{x \leftarrow ((\pi')^{n-k}|B)}[g(x) = \vec{a}] \geq 2\varepsilon$. This subcase is handled by Event 2.
2. $\Pr_{x \leftarrow (\pi')^{n-k}}[g(x) = \vec{a}] \geq \varepsilon$. This implies that the k-block $(\vec{\imath}, \vec{q})$ is alive (with respect to f). We have two subcases.
   (a) The k-block $(\vec{\imath}, \vec{q})$ is not $(1-\eta)$-good, where the precise value of η will be chosen later. This is handled by Event 3.
   (b) The k-block $(\vec{\imath}, \vec{q})$ is $(1-\eta)$-good. There are two subcases depending on the number of the remaining coordinates that are answered by P_1 according to the majority strategy. The subcase that this number is small is handled by Event 4, and the subcase that this number is large is handled by Event 5.

We now describe and analyze these events in greater detail.

Event 1. If $\Pr_{x \leftarrow ((\pi')^{n-k}|B)}[g(x) = \vec{a}] < 2\varepsilon$, we say that Event 1 occurs and assume that the verifier accepts with probability 2ε. This is indeed an upper bound on the acceptance probability in this case, because the verifier accepts only if P_1 and P_2 answer identically these k questions, and $\Pr[g(x) = \vec{a}]$ measures the probability (over the choices of the questions to P_1 in the miss rounds) that P_1 answers the same way as P_2 does. (Note that P_2's answer $\vec{a}$ is determined by $(\vec{\imath}, \vec{q})$ and B.) The contribution of Event 1 to the total acceptance probability is 2ε. (Note that we need not bound the probability of an event if we can bound the error probability conditioned on the event.)

Event 2. The outcome of Step 1 is a k-block $(\vec{\imath}, \vec{q})$. Event 2 happens when

1. $\Pr_{x \leftarrow (\pi')^{n-k}}[g(x) = \vec{a}] < \varepsilon$, and
2. $\Pr_{x \leftarrow ((\pi')^{n-k}|B)}[g(x) = \vec{a}] \geq 2\varepsilon$.

It follows that

$$\Pr_{x \leftarrow ((\pi')^{n-k}|B)}\left[g\left(x\right) = \vec{a}\right] - \Pr_{x \leftarrow (\pi')^{n-k}}\left[g\left(x\right) = \vec{a}\right] \geq \varepsilon.$$

We show that this happens with low probability, thus bounding the probability that Event 2 occurs and the verifier accepts. (Note that we do not need to consider the probability that the verifier accepts if we can bound the probability of the event.)

For function g, we apply Corollary 4.6 to bound the influence of Step 2 (in which an additional $(m-k)$-block is selected), obtaining

$$\mathrm{E}_{B \overset{m-k}{\leftarrow} (\pi')^{m-k}} \left[\sum_{\vec{a} \in A^k} \left(\mathrm{Pr}_{x \leftarrow (\pi')^{n-k}} \left[g\left(x\right) = \vec{a} \right] - \mathrm{Pr}_{x \leftarrow ((\pi')^{n-k} | B)} \left[g\left(x\right) = \vec{a} \right] \right)^2 \right]$$
$$\leq \left(m-k\right) / \left(n-k\right) \leq m/n.$$

Now, if Event 2 occurs, then there is some $\vec{a}$ that contributes ε^2 to the above summation term. Note that all the other terms in the summation are nonnegative. By Markov's inequality, the probability of Event 2, taken over the choice of B, is at most $m/n\varepsilon^2$.

Event 3. Event 3 is the event that $(\vec{\imath}, \vec{q})$, selected in Step 1, is alive but not $(1-\eta)$-good. We upper bound the probability of Event 3 by ε as follows. Recall that k was selected to be good with respect to (f, π') in the sense of Lemma 5.6. Hence, either $(\vec{\imath}, \vec{q})$ is alive with probability at most ε or it is the case that live $(\vec{\imath}, \vec{q})$ fail to be $(1-\eta)$-good with probability at most ε. In either case, the probability that $(\vec{\imath}, \vec{q})$ is alive but not $(1-\eta)$-good is at most ε.

To define Events 4 and 5, recall the notion of *majority answer* from Definition 5.4. We introduced the notation $\mathrm{maj}(i,q)$ to denote the most likely value for $f(x)[i]$ when $x[i] = q \overset{\mathrm{def}}{=} (q_1, q_2, b)$ (also conditioned on $x[\vec{\imath}] = \vec{q}$ and $f(x)[\vec{\imath}] = \vec{a}$). In our case, this corresponds to P_1's most likely bth half answer on coordinate i when $x[i] = q$. For $j \in \{1,2\}$, we extend this notation to $\mathrm{maj}(i,q,j)$ to denote P_1's most likely jth half answer on coordinate i when $x[i] = q$. Note that j need not be equal to b.

Event 4. Event 4 occurs when

1. $(\vec{\imath}, \vec{q})$ is *alive* and *good*;
2. $\mathrm{Pr}_{x \leftarrow (\pi')^{n-k}}[g(x) = \vec{a}] \geq \varepsilon$; and
3. given $(\vec{\imath}, \vec{q}, \vec{a})$, less than a $1 - 2\eta/\delta$ fraction of the $2(n-k)$ remaining half answers of P_1 equal their respective majority answer, $\mathrm{maj}(i,q,j)$. Here, δ is a parameter that will be chosen later.

Note that for a fixed k-block $(\vec{\imath}, \vec{q})$, at most $1/\varepsilon$ different $\vec{a}$ may satisfy this condition. We shall concentrate on an arbitrary one such $\vec{a}$ and later multiply the probability of acceptance by a factor of $1/\varepsilon$.

The k-block $(\vec{\imath}, \vec{q})$ is good and the triple $(\vec{\imath}, \vec{q}, \vec{a})$ is alive. By Proposition 5.5, the following experiment succeeds with probability $\alpha \geq 1 - 2\eta/\varepsilon$.

1. Choose i uniformly from $[n] - \vec{\imath}$ and choose z according to $(\pi')^n$, conditioned on $z[\vec{\imath}] = \vec{q}$ and $f(z)[\vec{\imath}] = \vec{a}$.
2. Let $q = z[i]$. Succeed iff $f(z)[i] = \mathrm{maj}(i,q)$.

Recall that maj is defined with respect to $(\vec{\imath}, \vec{q}, \vec{a})$.

Let β denote the expected fraction of half answers given by P_1 that are equal to their respective $\mathrm{maj}(i,q,j)$, where the conditioning is as in the definition of Event 4. We will show that $\alpha = \beta$, and hence that $\beta \geq 1 - 2\eta/\varepsilon$. By Markov's inequality, the probability that P_1 gives the majority answer on less than a $1 - 2\eta/\delta$ fraction of the half questions is at most δ/ε. Summing over at most $1/\varepsilon$ possible $\vec{a}$, we have that Event 4 occurs with probability at most δ/ε^2.

It remains to show that $\alpha = \beta$. Given $z = (q_1^1, q_2^1, b^1), (q_1^2, q_2^2, b^2), \ldots,$ define $z \backslash b = (q_1^1, q_2^1), (q_1^2, q_2^2), \ldots$ and $b(z) = b^1, b^2, \ldots$. Using the linearity of expectation we can express β as the success probability of the following experiment.

1. Choose i uniformly from $[n]-\vec{\imath}$ and choose z according to $(\pi')^n$, conditioned on $z[\vec{\imath}]=\vec{q}$ and $f(z)[\vec{\imath}]=\vec{a}$.
2. Let $q=z[i]\overset{\text{def}}{=}(q_1,q_2,b)$. Choose j uniformly from $\{1,2\}$. Let $(a_1,a_2)=p_1(z\backslash b)[i]$. Succeed iff $a_j=\text{maj}(i,q,j)$.

By the definition of f, we can rewrite the second step in the experiment defining α to be the following:

2. Let $q=z[i]\overset{\text{def}}{=}(q_1,q_2,b)$. Choose $j\in\{1,2\}$ as $j=b$. Let $(a_1,a_2)=p_1(z\backslash b)[i]$. Succeed iff $a_j=\text{maj}(i,q,j)$.

Thus, the only difference between the two experiments is in the distribution on j. The experiment for β chooses j uniformly, independently of all other events, whereas the experiment for α chooses $j=b$. We now use the fact that b is not an input to P_1, and hence the answers of P_1 are independent of b.

We can imagine choosing z according to $(\pi')^n$, then conditioning on $(\vec{\imath},\vec{q})$ and then conditioning on $f(z)[\vec{\imath}]=\vec{a}$. Initially, $b(z)$ is uniformly distributed, conditioned on $z\backslash b$. After conditioning on $\vec{q}$, $b(z)[\vec{\imath}]$ is fixed, but $b(z)[[n]-\vec{\imath}]$ remains uniformly distributed over $\{0,1\}^{n-k}$. Now, $f(z)[\vec{q}]$ is completely independent of $b(z)[[n]-\vec{\imath}]$, so further conditioning based on this value has no effect on the distribution of $b(z)[[n]-\vec{\imath}]$. Hence b is uniformly distributed and so is j in the experiment for α.

We established that j is identically distributed in the two experiments. To complete the proof of $\alpha=\beta$ we observe that the correlation between j and b in the experiment defining α is irrelevant to its probability of success, because neither (a_1,a_2) nor $(\text{maj}(i,q,1),\text{maj}(i,q,2))$ depends on b.

Event 5. Event 5 occurs when

1. $(\vec{\imath},\vec{q})$ is *alive* and *good*;
2. $\Pr_{x\leftarrow(\pi')^{n-k}}[g(x)=\vec{a}]\geq\varepsilon$; and
3. given $(\vec{\imath},\vec{q},\vec{a})$, at least a $1-2\eta/\delta$ fraction of the $2(n-k)$ remaining half answers of P_1 equal their respective majority answer, $\text{maj}(i,q,j)$.

It follows that for at least a $1-4\eta/\delta$ fraction of these coordinates, P_1 answers both half questions according to a majority strategy. Which majority strategy is used depends on $\vec{a}$; there are at most $1/\varepsilon$ values of $\vec{a}$ such that $\Pr_{x\leftarrow(Q^{n-k},(\pi')^{n-k})}[g(x)=\vec{a}]\geq\varepsilon$. Fixing $\vec{a}$ and assuming that P_1 plays the majority strategy for a $1-4\eta/\delta$ fraction of the $n-k$ subgames, consider the probability that verifier accepts on each of them. This probability is bounded above by the probability that the verifier accepts on a $1-4\eta/\delta$ fraction of the $n-k$ subgames, when P_1 plays the majority strategy for all $n-k$ subgames.

Fixing a strategy for P_1, whether the verifier accepts or rejects is a function of its random coins, which are chosen independently for each subgame. Since the majority strategy is a projection strategy, for each subgame the probability that the verifier accepts is at most p, independent of any of the other games. We can therefore apply a Chernoff bound (Theorem A.4 in [1]) and conclude that the probability that the verifier accepts in more than a $(1+p)/2$ fraction of the $n-k$ coordinates is at most $e^{-(1-p)^2(n-k)/2}<e^{-(1-p)^2n/4}$. When $\eta<(1-p)\delta/8$, then $(1+p)/2<1-4\eta/\delta$, and Event 5 occurs with probability at most $e^{-(1-p)^2n/4}/\varepsilon$.

Summing up the probabilities for the five events, the verifier accepts with probability at most

$$2\varepsilon+m/n\varepsilon^2+\varepsilon+\delta/\varepsilon^2+e^{-(1-p)^2n/4}/\varepsilon.$$

The various parameters need to satisfy $m<n/2$, $\eta<\frac{1}{2}$, $n>2^{10}\eta^{-4}\varepsilon^{-8}$, and $m>2^5\eta^{-2}\varepsilon^{-4}$ from Lemma 5.6 and $\eta<(1-p)\delta/8$ from Event 5. If $n\geq c/((1-p)\epsilon)^c$, for a sufficiently large c, then the other parameters can be chosen to make the above expression at most ϵ, as desired. Needless to say, our analysis is not tight. $\square$

Remark. The main difference between the proof of the above theorem in the current version and the preliminary version [12] is in the definition of the function f. In the miss-match case, the function f is based on the strategy of prover P_1, and its definition is a bit complicated. In the confuse-or-compare case [12], the function f was simply the strategy p_2 of prover P_2. The simpler definition for f allowed subsequent arguments to be stated more simply.

Acknowledgments. We thank Leonid Gurvits for his help in proving Theorem 4.3. We thank the referees for numerous useful comments.

REFERENCES

[1] N. Alon and J. Spencer, *The Probabilistic Method*, John Wiley, New York, 1992.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, *Proof verification and hardness of approximation problems*, J. ACM, 45 (1998), pp. 501–555.

[3] M. Bellare, S. Goldwasser, C. Lund, and A. Russell, *Efficient probabilistic checkable proofs and applications to approximation*, in Proceedings of the 25th Annual ACM Symposium on the Theory of Computing, San Diego, CA, 1993, pp. 294–304.

[4] M. Bellare, O. Goldreich, and M. Sudan, *Free bits, PCPs, and nonapproximability—towards tight results*, SIAM J. Comput., 27 (1998), pp. 804–915.

[5] M. Bellare and M. Sudan, *Improved non-approximability results*, in Proceedings of the 26th Annual ACM Symposium on the Theory of Computing, Montreal, Canada, 1994, pp. 184–193.

[6] M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson, *Multi prover interactive proofs: How to remove intractability assumptions*, in Proceedings of the 20th Annual ACM Symposium on the Theory of Computing, Chicago, IL, 1988, pp. 113–131.

[7] J. Cai, A. Condon, and R. Lipton, *On bounded round multi-prover interactive proof systems*, in Proceedings of the 5th IEEE Symposium on Structure in Complexity Theory, Barcelona, Spain, 1990, pp. 45–54.

[8] J. Cai, A. Condon, and R. Lipton, *Playing games of incomplete information*, Theoret. Comput. Sci., 103 (1992), pp. 25–38.

[9] J. Cai, A. Condon, and R. Lipton, *PSPACE is provable by two provers in one round*, J. Comput. System Sci., 48 (1994), pp. 183–193.

[10] U. Feige, *On the success probability of the two provers in one round proof systems*, in Proceedings of the 6th IEEE Symposium on Structure in Complexity Theory, Chicago, IL, 1991, pp. 116–123.

[11] U. Feige, *Error Reduction by Parallel Repetition—the State of the Art*, Technical report CS95-32, Weizmann Institute, Rehovot, Israel, 1995.

[12] U. Feige and J. Kilian, *Two prover protocols—low error at affordable rates*, in Proceedings of the 26th Annual ACM Symposium on the Theory of Computing, Las Vegas, NV, 1994, pp. 172–183.

[13] U. Feige and L. Lovasz, *Two-prover one-round proof systems, their power and their problems*, in Proceedings of the 24th Annual ACM Symposium on the Theory of Computing, Victoria, Canada, 1992, pp. 733–744.

[14] U. Feige and O. Verbitsky, *Error reduction by parallel repetition—a negative example*, in Proceedings of the 11th Annual IEEE Conference on Computational Complexity, Philadelphia, PA, 1996, pp. 70–76.

[15] L. Fortnow, J. Rompel, and M. Sipser, *On the power of multi-prover interactive protocols*, Theoret. Comput. Sci., 134 (1994), pp. 545–557.

[16] J. Hastad, *Clique is hard to approximate within $n^{1-\epsilon}$*, in Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science, Burlington, VT, 1996, pp. 627–636.

[17] J. Kahn, G. Kalai, and N. Linial, *The influence of variables on Boolean functions*, in Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science, White Plains, NY, 1988, pp. 68–80.

[18] D. Lapidot and A. Shamir, *A one-round, two-prover, zero-knowledge protocol for NP*, Combinatorica, 15 (1995), pp. 203–214.

[19] D. Lapidot and A. Shamir, *Fully parallelized multi prover protocols for NEXP-time*, J. Comput. System Sci., 54 (1997), pp. 215–220.

[20] C. Lund and M. Yannakakis, *On the hardness of approximating minimization problems*, J. ACM, 41(5) (1194), pp. 960–981.

[21] R. Raz, *A parallel repetition theorem*, SIAM J. Comput., 27 (1998), pp. 763–803.

[22] O. Verbitsky, *Towards the parallel repetition conjecture,* Theoret. Comput. Sci., 157 (1996), pp. 277–282.