Electronic Colloquium on Computational Complexity, Report No. 85 (2012)

A multi-prover interactive proof for NEXP sound against entangled provers

Tsuyoshi Ito[*] Thomas Vidick[†]

July 5, 2012

Abstract

We prove a strong limitation on the ability of entangled provers to collude in a multiplayer game. Our main result is the first nontrivial lower bound on the class MIP* of languages having multi-prover interactive proofs with entangled provers; namely MIP* contains NEXP, the class of languages decidable in non-deterministic exponential time. While Babai, Fortnow, and Lund (*Computational Complexity* 1991) proved the celebrated equality MIP $=$ NEXP in the absence of entanglement, ever since the introduction of the class MIP* it was open whether shared entanglement between the provers could weaken or strengthen the computational power of multi-prover interactive proofs. Our result shows that it does not weaken their computational power: MIP $\subseteq$ MIP*.

At the heart of our result is a proof that Babai, Fortnow, and Lund's *multilinearity test* is sound even in the presence of entanglement between the provers, and our analysis of this test could be of independent interest. As a byproduct we show that the correlations produced by any entangled strategy which succeeds in the multilinearity test with high probability can always be closely approximated using shared randomness alone, and are thus restricted to being quasi-classical.

1 Introduction

Multiprover interactive proof systems [BGKW88] are at the heart of much of the recent history of complexity theory, and the celebrated characterization MIP $=$ NEXP [BFL91] is one of the cornerstones on which the PCP theorem [AS98, ALMSS98] was built. While the key assumption on the multiple provers in an interactive proof system is that they are not allowed to communicate, traditionally this has been taken to mean that their only distributed resource was shared randomness. In a quantum universe, however, it is natural to relax this assumption and allow the provers to share *entanglement*. While still not allowing them to communicate, this increases their ability to collude against the verifier by exploiting the nonlocal correlations allowed by entanglement. The corresponding complexity class MIP* was introduced in [CHTW04], raising a fundamental question: *what is the computational complexity of entangled provers?*

[*] NEC Laboratories America, Inc., USA. Email: tsuyoshi@nec-labs.com. Supported by ARO/NSA grant W911NF-09-1-0569, as well as by grants from NSERC, CIFAR, QuantumWorks, MITACS, CFI, and ORF received while this author was a postdoctoral fellow at the Institute for Quantum Computing and David R. Cheriton School of Computer Science, University of Waterloo, Canada.

[†] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA. Email: vidick@csail.mit.edu. Supported by the National Science Foundation under Grant No. 0844626. Part of this work was done while this author was a graduate student in the Computer Science department at UC Berkeley, as well as during visits to the Perimeter Institute in Waterloo, Canada and NEC Labs America.

ISSN 1433-8092

Even before their modern re-formulation in the language of multiplayer games, starting with the work of Bell in the 1960s [Bel64] the strength of the nonlocal correlations that could be obtained from performing local measurements on entangled particles has been intensely investigated through the use of *Bell inequalities* (upper bounds on the strength of classical correlations) and *Tsirelson inequalities* (upper bounds on the strength of quantum correlations). Games, or proof systems, generalize this setup by introducing an additional layer of *interaction*: in this new context, we think of the experimenter (the verifier) as interacting with the physical devices (the provers) through the specific choice of settings (questions) that he makes, and the outcomes (answers) that he observes. The arbitrary state and measurements that are actually made inside the devices are reflected in the provers' freedom in choosing their strategy. The fundamental observation that quantum mechanics violates certain Bell inequalities translates into the fact that there exists interactive proof systems in which entangled provers can have a strictly higher success probability than could any classical, non-entangled provers.

A dramatic demonstration of this possibility is given by the Magic Square game [Mer90, Per90], a simple one-round game for which the maximum success probability of classical provers is $8/9$, but there exists a *perfect* winning strategy for entangled provers. Cleve, Høyer, Toner, and Watrous [CHTW04] were the first to draw complexity-theoretic consequences from such *non-local* properties of entanglement. They study the class $\oplus$MIP of languages having two-prover interactive proofs in which there is a single round of interaction, each of the provers is restricted to answering a single bit, and the verifier only bases his accept/reject decision on the parity of the two bits that he received. While it follows from work of Håstad [Hås01] that this class equals NEXP (and is thus as powerful as the whole of MIP) for an appropriate setting of completeness and soundness parameters, Cleve et al. show that the corresponding entangled-prover class $\oplus\mathrm{MIP}^*$ *collapses* to EXP for any choice of completeness and soundness parameters that are separated by an inverse polynomial gap.[1]

Despite intense efforts, for a long time little more was known, and prior to our work the best lower bound on MIP^* resulted from the trivial observation that multiple entangled provers are at least as powerful as a single prover, hence $\mathrm{IP} = \mathrm{PSPACE} \subseteq \mathrm{MIP}^*$, where the first equality is due to [LFKN92, Sha92].[2] The main difficulty in improving this trivial lower bound is the following: while the PCP theorem gives us a variety of two-prover interactive proof systems for NEXP-complete problems, there is strong indication (see e.g. the Magic Square game, which has a very similar structure to that of basic proof systems for MAX-3-XOR, or the aforementioned collapse of $\oplus\mathrm{MIP}^*$) that these may no longer be sound in the presence of entanglement. The fact that entanglement, as a shared resource, is poorly understood is also reflected in the complete absence of reasonable *upper bounds* on the complexity class MIP^*: while the inclusion $\mathrm{MIP} \subseteq \mathrm{NEXP}$ is straightforward, we do not know of any limits on the *dimension* of entanglement that may be useful to the provers in a given interactive proof system, and as a result their maximum success probability is not even known to be computable (see [SW08, DLTW08, NPA08] for more on this aspect).

Since existing protocols may no longer be sound in the presence of entanglement between the provers, previous work has focused on finding ways to *modify* a given protocol in a way that would make it *entanglement resistant*; that is, honest provers can convince the verifier without shared en-

[1] This was later improved [Weh06] to the inclusion of $\oplus\mathrm{MIP}^*$ in the class of two-message single-prover interactive proofs $\mathrm{QIP}(2) \subseteq \mathrm{PSPACE}$ [JUW09].

[2] It was recently shown that quantum messages are no more powerful than classical messages in *single-prover* interactive proof systems [JJUW11]: $\mathrm{QIP} = \mathrm{PSPACE}$. That result, however, has no direct relationship with our work: in our setting the messages remain classical; rather the "quantumness" manifests itself in the presence of *entanglement* between the provers, which is a notion that only arises when more than one prover is present.

tanglement while dishonest provers cannot convince the verifier with high probability even with shared entanglement. This was the route taken in [KKMTV11, IKPSY08, IKM09], which introduced techniques to limit the provers' use of their entanglement. They proved non-trivial lower bounds on variants of the class MIP*, but with error bounds that are weaker than the standard definitions allow for. These relatively weak bounds came as a result of the "rounding" technique developed in these works: by adding additional constraints to the protocol, one ensures that optimal entangled strategies are in a sense close to classical, un-entangled strategies. This closeness, however, was shown using a rounding procedure that had a certain "local" flavor, inducing a large loss in the quality of the approximation.[3]

In addition, [IKM09], based on [KKMTV11], showed that PSPACE has two-prover *one-round* interactive proofs with entangled provers, with perfect completeness and exponentially small soundness error. Prior to our work, this was the best lower bound known on single-round multi-prover interactive proof systems with entanglement.

Other related work. Given the apparent difficulty of proving good lower bounds on the power of multi-prover interactive proof systems with entangled provers, researchers have studied a variety of related models. Maybe the most natural extension of MIP* consists in giving the verifier more power by allowing him to run in quantum polynomial-time, and exchange quantum messages with the provers. The resulting class is called QMIP* (the Q stands for "quantum verifier", while the * stands for "entangled provers"), and it was formally introduced in [KM03], where it was shown that QMIP* contains MIP* (indeed, the verifier can always force classical communication by systematically measuring the provers' answers in the computational basis). Little more is known of QMIP*; in fact it is believed to equal MIP* [BFK10]. Ben-Or et al. [BHP08] introduced a model in which the verifier is quantum and the provers are allowed communication but no entanglement, and showed that the resulting class contains NEXP. Other works attempt to characterize the power of MIP* systems using *tensor norms* [RT07, JPPVW10]; so far however such norms have either led to computable, but very imprecise, approximations, or have remained (to the best of our knowledge) intractable.

1.1 Results

Let MIP$^*(k, m, c, s)$ be the class of languages that can be decided by an m-round interactive proof system with k (possibly entangled) provers and with completeness c and soundness error s.[4] Our main result is the following.

Theorem 1. *All languages in* NEXP *have a four-prover poly-round interactive proof system with perfect completeness and exponentially small soundness error against entangled provers. That is, for every $q \in$* poly*, it holds that*

$$\mathrm{NEXP} \subseteq \mathrm{MIP}^*(4, \mathrm{poly}, 1, 2^{-q}).$$

Theorem 1 resolves a long-standing open question [KM03], showing that entanglement does not weaken the power of multi-prover interactive proof systems: together with the celebrated result NEXP $=$ MIP [BFL91], it implies that MIP $\subseteq$ MIP*. We note that the proof system in Theorem 1 does not require honest provers to use any entanglement in order to achieve perfect completeness in the case of a YES-instance. In other words, if we denote by MIP$^{\mathrm{er}}$ the class of languages having entanglement resistant multi-prover interactive proof systems with bounded

[3] See the "almost-commuting implies nearly-commuting" conjecture in [KKMTV11] for more on this aspect.

[4] We refer to Section 2.2 for a more complete definition of the class MIP*.

error, our proof of Theorem 1 shows that NEXP $\subseteq$ MIP$^{\text{er}}$. Because MIP$^{\text{er}}$ $\subseteq$ MIP by definition, this implies MIP$^{\text{er}}$ = NEXP.

The interactive proof system used in the proof of Theorem 1 uses four provers and a polynomial number of rounds of interaction. We do not know if the number of provers can be reduced; however if one is willing to increase it by one then the amount of interaction required can be reduced to a single round, i.e. one message from the verifier to each prover, and one message from each prover to the verifier. Indeed, our proof system has the additional property of being *non-adaptive*: the verifier can select his questions for all the rounds before interacting with any of the provers. It is shown in [Ito11] that a non-adaptive entanglement-resistant protocol may be parallelized to a single round of interaction at the cost of adding an extra prover. Applying this result to Theorem 1 gives the following corollary.

Corollary 2. *All languages in* NEXP *have a five-prover one-round interactive proof system with perfect completeness and soundness error against entangled provers bounded away from* 1 *by an inverse polynomial, that is:*

$$\text{NEXP} \subseteq \text{MIP}^*(5, 1, 1, 1 - 1/\operatorname{poly}).$$

Prior results on the complexity of multi-prover interactive proofs with entangled provers have often been stated using the languages of *games* [CHTW04, KKMTV11, KRT10]. The main difference, in terms of computational complexity, is in the way the input size is measured. In the case of games the input is an explicit description of the game, including a list of all possible questions and valid answers, while in the setting of proof systems the messages may be described implicitly: it is their *length* that is polynomial in the input size.

Because of this difference in scaling, our results do not immediately imply any NP-hardness result in the setting of multi-player games with entangled players. Nevertheless, by adapting the proof of Theorem 1 and using the PCP theorem one can show the following. There is a constant $\kappa > 1$ and a procedure that, given as input an arbitrary 3-SAT formula with n variables and $m = \text{poly}(n)$ clauses, runs in time $2^{O(\log^\kappa n)}$ and produces an explicit description of a four-player game of size $S = 2^{O(\log^\kappa n)}$ (i.e. the number of rounds of interaction and the total number of questions and answers that can be sent and received is at most S). The game has the property that, if the 3-SAT formula was satisfiable, then there is a perfect strategy for the players, which does not require any entanglement. If, however, the 3-SAT formula was not satisfiable, then there is no strategy for the players, even using entanglement, that succeeds with probability greater than 1/2.

If one could show the above with constant $\kappa = 1$ then it would follow that finding a constant-factor approximation to the maximum success probability of four entangled players in a game with polynomially many rounds and questions is NP-hard; however our result is limited to obtaining some possibly large $\kappa > 1$. The main point, however, is that the hardness of approximation is up to *constant* factors. This is in contrast to all previous results which were limited to hardness of approximation up to factors approaching 1 very quickly as the input size grew (even after arbitrary sequential or even parallel repetition).[5]

At the heart of the proof of Theorem 1 is a soundness analysis of Babai, Fortnow and Lund's *multilinearity test* in the presence of entanglement between the provers: we show that it is in a sense "immune" to the strong non-local correlations that entangled provers may in general afford. We believe that this analysis should be of wider interest, and we explain the test and the main ideas behind its analysis in the presence of entanglement in Section 1.3 below. We first briefly outline

[5] Cleve, Gavinsky, and Jain [CGJ09] obtained a constant-factor hardness result for games with constant answer size, but in which the number of questions sent by the verifier is *exponential*.

the overall structure of our proof system in Section 1.2. It is very similar to the one introduced by Babai, Fortnow, and Lund [BFL91] to prove NEXP $\subseteq$ MIP; our contribution consists in proving its soundness against entangled provers.

1.2 Proof outline

Our interactive proof system, just as the one by Babai et al.,[6] verifies membership in a specific NEXP-complete language, *Oracle-3-satisfiability* (see Problems 1 and 2 in Section 2.3 for a definition). We give a four-prover, poly-round interactive protocol for it that has perfect completeness and soundness error bounded away from 1 by an inverse-polynomial in the input size. (Theorem 1 is obtained by sequentially repeating this interactive proof system.)

Simplifying a little bit (we refer the reader to Section 3 for details), the verifier in our protocol is given as input two integers n, N in unary (think of N as much larger than n, but still polynomial), a description of a finite field $\mathbb{F}$ of size N, and a low-degree polynomial $f : (\mathbb{F}^n)^3 \times (\mathbb{F})^3 \to \mathbb{F}$. His goal is to verify whether there exists a multilinear function $g : \mathbb{F}^n \to \mathbb{F}$ such that $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, g(\boldsymbol{x}), g(\boldsymbol{y}), g(\boldsymbol{z})) = 0$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \{0,1\}^n \subset \mathbb{F}^n$. If this is the case then the input is a YES-instance, whereas if for all functions g that are "close" to multilinear functions at least one of the constraints $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, g(\boldsymbol{x}), g(\boldsymbol{y}), g(\boldsymbol{z})) = 0$ is not satisfied then it is a NO-instance. The difficulty, of course, is that there are exponentially many constraints to verify, and *all* must be satisfied for the instance to be a YES-instance.

The protocol is divided into two distinct parts, which only weakly interact with each other. In the first part of the protocol, the verifier performs a polynomial-round *low-degree sum-check test* with a single prover, say the last prover (see Lemma 9 for an explicit formulation). This test is based on ideas already introduced by Lund, Fortnow, Karloff, and Nisan [LFKN92] and can be used to verify that a low-degree function defined over $\mathbb{F}^k$ vanishes on all of $\{0,1\}^k$. We will apply it to the low-degree function $h : (\mathbb{F}^n)^3 \to \mathbb{F}$ defined by $h(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, g(\boldsymbol{x}), g(\boldsymbol{y}), g(\boldsymbol{z}))$. An important point for us is that, in the LFKN protocol, the verifier eventually only needs to evaluate h at a *single* point $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in (\mathbb{F}^n)^3$ chosen uniformly at random.

Of course, the verifier only knows f, not g, and the goal of the second part of the protocol is for the verifier to learn the three values $g(\boldsymbol{x}), g(\boldsymbol{y}), g(\boldsymbol{z})$. Note that here the function g is arbitrary (we are trying to verify its existence), *except that it has to be multilinear*. Hence the verifier will perform one of two tests with the three remaining provers: either directly ask them for the values $g(\boldsymbol{x}), g(\boldsymbol{y}), g(\boldsymbol{z})$, or perform a certain "multilinearity test", which enforces that, however the provers answer their queries, it must be according to a function that is close to a multilinear function. The two tests will be indistinguishable from the point of view of the provers because the marginal distribution on the question to each prover is uniform over $\mathbb{F}^n$ in both cases.

Our contribution consists in showing that this protocol is sound even in the presence of entanglement between the provers. Precisely, assuming four entangled provers succeed with probability that is polynomially close to 1, we wish to conclude that the instance given as input to the verifier is a YES-instance.

Note that provers successful in the overall protocol must, in particular, succeed with high probability in the multilinearity test. The key step in the analysis consists in showing the following: Any three entangled provers that succeed in the multilinearity test with high probability are "indistinguishable" from *classical* provers who use shared randomness to jointly sample a multilinear

[6] We emphasize that the proof system we use is not new, as it is essentially the same as the one introduced in [BFL91]. We nevertheless outline it because there is a small difference in how the "oracle" in [BFL91] is simulated by provers, which is the reason our protocol, unlike the one in [BFL91], requires more than two provers.

function g, and then answer question $\boldsymbol{x}$ with $g(\boldsymbol{x})$. This step is the one that requires the most work, and we explain it in more detail in the next section. (In particular, we will clarify what is meant by "indistinguishable".)

Assuming this informal statement holds, it is not too hard to conclude the analysis of the protocol. Indeed, having replaced the three provers used in the multilinearity test by three classical provers, there is only a single "quantum" prover left, the one used to perform the sum-check test in the first part of the protocol. But entanglement cannot be useful to a single prover, and hence we may also assume that this last prover behaves classically. Since all provers are now classical, we have reduced our analysis to the classical setting and can appeal to the results in [BFL91] to conclude. We refer to Section 3 for a more detailed presentation and soundness analysis of the protocol.

1.3 The multilinearity game

The key step in the proof of Theorem 1 is the analysis of the multilinearity test of [BFL91], which generalizes the celebrated *linearity test* of Blum, Luby, and Rubinfeld [BLR93] and is essential in constructing a protocol for NEXP that has messages of polynomial length.[7] The test can be formulated as a game played between the verifier and three players. The game is parametrized by a finite field $\mathbb{F}$ and an integer n. In the game, the verifier performs either of the following with probability $1/2$ each:

- *Consistency test.* The verifier chooses $\boldsymbol{x} \in \mathbb{F}^n$ uniformly at random and sends the same question $\boldsymbol{x}$ to all three players. He expects each of them to answer with an element of $\mathbb{F}$, and accepts if and only if all the answers are equal.

- *Linearity test.* The verifier chooses $i \in \{1, \ldots, s\}$, $\boldsymbol{x} \in \mathbb{F}^n$ and $y_i, z_i \in \mathbb{F}$ uniformly at random, and sets $y_j = z_j = x_j$ for every $j \in \{1, \ldots, n\} \setminus \{i\}$. He sends $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ to the three players, receives $a, b, c \in \mathbb{F}$, and accepts if and only if

$$\frac{b-a}{y_i - x_i} = \frac{c-b}{z_i - y_i} = \frac{c-a}{z_i - x_i}.$$

Babai, Fortnow, and Lund show that, if any three *deterministic* players are accepted by the verifier with probability at least $1 - \varepsilon$ in this game, then the functions they each apply to their questions in order to determine their respective answers are close to a single *multilinear* function $g : \mathbb{F}^n \to \mathbb{F}$ (see Theorem 4.16 in [BFL91] for an analysis of a variant of the test over the integers). That is, for all but at most a fraction roughly $O(n^2\varepsilon)$ (provided $|\mathbb{F}|$ is large enough) of $\boldsymbol{x} \in \mathbb{F}^n$, the players' answer to question $\boldsymbol{x}$ is precisely $g(\boldsymbol{x})$.

A major hurdle in proving a similar statement in case the players are allowed to use quantum mechanics already arises in *formulating* the statement to be proven: even in the case of players restricting their use of entanglement as shared randomness, what meaning should one ascribe to their strategies being "close to multilinear"? Indeed, it could be that the answer of each player to a fixed question, when taken in isolation, is uniformly random: the whole substance of the strategy is in the *correlations* between the answers of different players. This difficulty is usually set aside by "fixing the randomness". Quantum entanglement, however, cannot be "fixed", and this forces us to face even the presumably simpler case of randomized strategies head on. We show that the

[7] One can devise a protocol based on the linearity test alone, but it requires the verifier to send messages with exponential length to the provers. Such use of the linearity test was already key in establishing the early result NP $=$ PCP$(\text{poly}, 1)$; see e.g. Theorem 2.1.10 in [ALMSS98].

following is an appropriate formulation of Babai et al.'s multilinearity test in the general setting of entangled (or even just randomized) players (see Theorem 11 for a precise statement).

Theorem 3 (Informal). *Suppose that three entangled players who share a permutation-invariant state $|\Psi\rangle$ succeed in the multilinearity game with probability $1-\varepsilon$ where each player uses the set of measurements $\{A_{\boldsymbol{x}}^a\}_{a\in\mathbb{F}}$ to determine his answer to the verifier's question $\boldsymbol{x}\in\mathbb{F}^n$.*

Then there exists a single *measurement $\{V^g\}$, independent of any question and with outcomes in the set of all multilinear functions $g:\mathbb{F}^n\to\mathbb{F}$, such that, in the multilinearity game, each player's action is* indistinguishable *from that of player whom, upon receiving his question $\boldsymbol{x}$, would*

1. *Measure his share of $|\Psi\rangle$ with $\{V^g\}$, obtaining a multilinear function g as an outcome,*
2. *Answer his question $\boldsymbol{x}$ with $g(\boldsymbol{x})$.*

Moreover, the multilinear functions used by the three players are identical with high probability.

In case the players are classical, but may use shared randomness, the theorem makes the following simple statement: players successful in the multilinearity game are "indistinguishable" from players who would first look up their random string, based on that alone select a multilinear function g, and finally answer their respective questions $\boldsymbol{x}_i$ with $g(\boldsymbol{x}_i)$. While such a statement is a direct corollary of Babai, Fortnow, and Lund's analysis, our contribution is to prove it without first "fixing the randomness" — and to show that it also holds for the case of players using entanglement.

An appropriate notion of distance on entangled-prover strategies. Crucial to the applicability of Theorem 3 is the precise notion of "indistinguishability" used. Indeed, while there is no hope of making statements on the players' measurements or their shared entangled state themselves (since the verifier has no direct access to them throughout the protocol), one still needs to use a notion that is strong enough to be meaningful even when the multilinearity game is executed as a building block in the larger protocol explained in the previous section.

The measure we use is based on the notion of *consistency* between two measurements, and it may be useful to introduce it here in a simplified setting (precise definitions are given in Section 2.1). Let $\{A^i\}_{i\in I}$ and $\{B^i\}_{i\in I}$ be two quantum measurements of the same dimension and indexed by the same set of outcomes: $A^i, B^i\geq 0$ for all $i\in I$, and $\sum_i A^i=\sum_i B^i=\mathrm{Id}$. Let $|\Psi\rangle$ be a bipartite state that is invariant under permutation of its two subsystems, and ρ its reduced state on either. We say that A and B are *ε-consistent* if the following holds:

$$\textsc{con}(A,B) := \sum_i \langle\Psi|A^i\otimes B^i|\Psi\rangle \geq 1-\varepsilon. \tag{1}$$

This definition has an operational interpretation: the two measurements A and B, when performed on the two subsystems of $|\Psi\rangle$, give the same outcome except with probability ε . The key fact about consistent measurements is the following. Suppose that A and A, B and B, and A and B are all ε-consistent. Then A and B are *indistinguishable* in the sense that

$$\sum_i \left\|\sqrt{A^i}\rho\sqrt{A^i}-\sqrt{B^i}\rho\sqrt{B^i}\right\|_1 = O(\sqrt{\varepsilon}). \tag{2}$$

This last expression corresponds to a more familiar notion of closeness of two measurements: they are close if the post-measurement states resulting from applying either are close in trace distance. The fact that (1) essentially implies (2) relies on Winter's "gentle measurement" lemma [Win99, Lemma 9] (see also Aaronson's "almost as good as new" lemma [Aar05, Lemma 2.2]), a key tool in our analysis.

In this paper we will consider two measurements to be close whenever they are consistent, having the assurance that this notion of closeness implies the more traditional one expressed by (2). In particular, it is not hard to verify that (2) implies that either measurement may be "replaced" by the other even in a wider context; see the proof of Claim 12 in Section 3 for more details on how this can be done. The advantage of using this measure is that constraints on the consistency of measurements arise naturally from the analysis of the multilinearity game, and it is a notion that is very convenient to work with.

Analysis of the multilinearity game: rounding entangled strategies. Theorem 3 states that success in the multilinearity game forces even entangled players to make a trivial use of their entanglement: since the measurement $\{V^g\}$ is independent of their respective questions, they might as well perform it before the game starts, in which case they are not using their entanglement at all. Hence the theorem implies that entangled players are no more powerful than classical players in that game. A key insight of our work, however, is to avoid any attempt to prove such a statement *directly*. Instead, our proof technique consists in progressively manipulating the players' strategies themselves, without *explicitly* trying to relate them to a classical strategy.

Our goal is to show how the measurement $\{V^g\}$ can be extracted from the initial set of measurements $\{A_{\boldsymbol{x}}^a\}$ which depend on $\boldsymbol{x} \in \mathbb{F}^n$.[8] More precisely, we show how, starting from the original measurements $\{A_{\boldsymbol{x}}^a\}$, one may remove the dependence of $\{A_{\boldsymbol{x}}^a\}$ on $\boldsymbol{x} \in \mathbb{F}^n$ one coordinate at a time — eventually reaching the measurement $\{V^g\}$. Towards this we construct a sequence of measurements $\{B_{x_{k+1},\dots,x_n}^g\}_g$, for $k = 1, \dots, n$, with outcomes g in the set of multilinear functions $\mathbb{F}^k \to \mathbb{F}$. Each of these measurements has the following key property: the respective strategies corresponding to (i) measuring according to $\{A_{\boldsymbol{x}}^a\}$ and answering a or (ii) measuring according to $\{B_{x_{k+1},\dots,x_n}^g\}$ and answering $g(x_1, \dots, x_k)$ are *consistent*, in the sense described in Eq. (1): two distinct players using either strategy will obtain the same answer with high probability (provided they started with the same question).

This sequence of measurements is defined by induction, and we only explain the one-dimensional case here. Our construction is intuitive: $\{B^g\}$ corresponds to measuring using $\{A_{x_1}^a\}$ twice, *in succession*, using two randomly chosen values of x_1, and returning the unique linear function g which interpolates between the two outcomes obtained. This can be interpreted as a quantum analogue of the reconstruction procedure already used in the linearity test of Blum, Luby, and Rubinfeld: to recover a linear function it suffices to evaluate it at two random points, and then interpolate. The construction of the measurements $\{B_{x_{k+1},\dots,x_n}^g\}$ for the one dimensional case is given in Claim 15, and in the general case in Lemma 18, which states a quantum analogue of Babai et al.'s "pasting lemma" [BFL91, Lemma 5.11].

An additional hurdle arises as a result of the induction: the quality of the approximation between the original measurements $\{A_{\boldsymbol{x}}^a\}$ and the constructed measurements $\{B_{x_{k+1},\dots,x_n}^g\}$ blows up *exponentially* with k. In order to control this error, one has to perform an additional step of *self-improvement*. This step was a key innovation in the work of Babai, Fortnow, and Lund, and extending it to the setting of entangled strategies requires substantially more work. While for the case of deterministic strategies Babai et al. were able to show, using the expansion properties of the hypercube, that any "reasonably good" k-linear approximation g at any point in the induction was automatically "extremely good", in our case we need to actively update the measurements

[8] While we do give an explicit, inductive algorithmic procedure showing how $\{V^g\}$ can be constructed, this is not necessary: the point is only in proving its *existence*.

through a self-correction procedure, obtaining the "improved" measurements as the optimum of a certain convex optimization problem. The need for such active correction is not a limitation of our approach, but rather reflects a fundamental difference between the quantum and the classical, deterministic settings: while two binary-valued functions either fully agree or fully disagree at any point, two quantum measurements can produce outcomes according to distinct but arbitrarily close distributions (think of one of the measurements as being obtained from the other by a small perturbation, such as an arbitrarily small rotation). It is this kind of "error" that needs to be corrected, and we explain our method to do so in more detail in Section 5.1.

1.4 Discussion and open questions

Improving the parameters in Theorem 1 and Corollary 2 is an open problem. For example, it might be possible to reduce the number of provers to two, and the number of rounds of interaction to one, while still preserving exponentially small soundness error, resulting in the inclusion NEXP $\subseteq$ MIP$^*(2,1,1,2^{-q})$ for every polynomial q. This would be an analogue of the known containment NEXP $\subseteq$ MIP$(2,1,1,2^{-q})$ [FL92]. Our overall protocol for NEXP requires four provers, and five provers if we would like to parallelize it by using [Ito11]. We leave the problem of reducing the number of provers to fewer than four for future work. It may also be possible to improve the soundness guarantees in Corollary 2 by using the parallel repetition techniques from [KV11], but we have not explored this possibility.

In comparison to the PCP theorem, there are important parameters which are not explicit in Theorem 1 and Corollary 2: the amount of randomness used by the verifier and the total answer length. In our constructions, both of them are just bounded by a polynomial in the input length for NEXP, and they are poly-logarithmic for the scaled-down version corresponding to verification of languages in NP. If these numbers are respectively reduced to a logarithm and a constant for NP with a constant soundness, the result will be an analogue of the PCP theorem in presence of entanglement. Obtaining such a result may require extending our analysis of the multilinearity test to the more powerful *low-degree* tests that were key to establishing the "scaled-down" version of the PCP theorem.

Honest provers in our protocol do not need entanglement in order to achieve completeness 1 in the case of a YES-instance. It remains open whether entanglement can have any positive use in this context: is MIP* strictly larger than MIP $=$ NEXP?

Organization of the paper. After giving some necessary preliminaries, Section 3 describes the protocol used to prove Theorem 1, and shows how the theorem follows from a claim about the multilinearity game in the presence of entangled provers. Section 4 introduces a more technical claim about the analysis of the multilinearity game, which is suitable to a proof by induction on the number n of variables in the verifier's questions in the game. The actual analysis is given in Section 5.

Acknowledgments. Tsuyoshi Ito thanks John Watrous for helpful discussions. Thomas Vidick thanks Umesh Vazirani for many inspiring discussions throughout the time that this work was being carried out, and in particular for first suggesting to adapt Babai et al.'s multilinearity test to the entangled-prover setting. The authors also thank Scott Aaronson, Dmitry Gavinsky, and an anonymous referee for helpful suggestions.

2 Preliminaries

In the remainder of the paper we assume that the reader is familiar with computational complexity theory [Gol08, AB09], as well as with basic notions in quantum information [NC01, KSV02] such as density matrices, POVM measurements, quantum channels, and the trace distance. For more on quantum computational complexity we refer the reader to a recent survey by Watrous [Wat09].

2.1 Notation

For a field $\mathbb{F}$, a linear function $g\colon \mathbb{F} \to \mathbb{F}$ is a function such that there exists $a, b \in \mathbb{F}$, $g(x) = ax + b$. A multilinear function $g\colon \mathbb{F}^k \to \mathbb{F}$ is a function that is linear in each of its coordinates. $\mathrm{ML}(\mathbb{F}^k, \mathbb{F})$ will denote the set of all multilinear functions from $\mathbb{F}^k$ to $\mathbb{F}$. We will denote tuples using bold symbols such as $\boldsymbol{x}$ and $\boldsymbol{b}$. Given a tuple $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $k \in [n]$, we let $\boldsymbol{x}_{\leq k} := (x_1, \ldots, x_k)$, $\boldsymbol{x}_{>k} := (x_{k+1}, \ldots, x_n)$ and $\boldsymbol{x}_{\neg k} := (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)$.

Given a positive matrix ρ and an arbitrary matrix A, we let $\mathrm{Tr}_\rho(A) := \mathrm{Tr}(A\rho)$. In case ρ is a matrix on the tensor product of two Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, and A is a matrix acting on $\mathcal{H}$, we will sometimes abuse notation and write $\mathrm{Tr}_\rho(A)$ for $\mathrm{Tr}_\rho(A \otimes \mathrm{Id}_{\mathcal{H}'})$. If $|\Psi\rangle \in \mathcal{H}^{\otimes k} \otimes \mathcal{H}'$ is a state that is invariant under permutation of the first k registers, we will often abuse notation further and use the symbol ρ to denote the reduced density of $|\Psi\rangle$ on either of the first k registers, or even any pair of registers among the first k, etc. Hence any expression of the form $\mathrm{Tr}_\rho(A \otimes B)$ should really be read as

$$\langle\Psi| A \otimes B \otimes \mathrm{Id}_{\mathcal{H}} \otimes \cdots \otimes \mathrm{Id}_{\mathcal{H}} \otimes \mathrm{Id}_{\mathcal{H}'} |\Psi\rangle,$$

where the position of A and B among the first k registers is immaterial by permutation-invariance. For any $\rho \geq 0$, we let

$$\|A\|_\rho^2 := \mathrm{Tr}\big(AA^\dagger\rho\big),$$

and observe that $A \mapsto \|A\|_\rho$ is a semi-norm (it is definite if ρ is invertible). It satisfies the following Cauchy-Schwarz inequality: for any A, B,

$$\mathrm{Tr}_\rho\big(AB^\dagger\big) \leq \|A\|_\rho \, \|B\|_\rho.$$

Measurements. In this paper, a measurement is a collection of non-negative matrices $\{P^a\}_{a\in A}$ such that $\sum_a P^a = \mathrm{Id}$ (this is usually called a *Positive Operator-Valued Measure*, or POVM). The set A is the set of *outcomes* of the measurement; outcomes will always appear as superscripts. The measurement is said *projective* if P^a is a projector, i.e. $(P^a)^2 = P^a$, for every a. A *sub*-measurement is a collection of non-negative matrices $\{P^a\}_{a\in A}$ such that $\sum_a P^a \leq \mathrm{Id}$. For integers $0 \leq k \leq n$ we will also consider families of sub-measurements, indexed by $x \in \mathbb{F}^{n-k}$ and with outcomes in the set $\mathrm{ML}(\mathbb{F}^k, \mathbb{F})$. Such a family $P = \{P^g_{\boldsymbol{x}_{>k}}\}$ will be called a *family of sub-measurements of arity k* (the parameter n will often be left implicit). A family of sub-measurements of arity n is thus a single sub-measurement with outcomes in $\mathrm{ML}(\mathbb{F}^n, \mathbb{F})$. Given a family of sub-measurements $P = \{P^g_{\boldsymbol{x}_{>k}}\}$ of arity k, we will often use the notation

$$P_{\boldsymbol{x}_{>k}} := \sum_g P^g_{\boldsymbol{x}_{>k}} \qquad \text{and} \qquad P_{\boldsymbol{x}_{\geq \ell}} := \mathrm{E}_{x_k, \ldots, x_{\ell-1}} P_{\boldsymbol{x}_{\geq k}}$$

for any $k \leq \ell \leq n$, where the expectation is taken with respect to the uniform distribution on $\mathbb{F}^{\ell-k}$. Given two families of sub-measurements P and Q with arities $k \leq \ell$ respectively, we define their *consistency*

$$\mathrm{CON}(P, Q) := \mathrm{E}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{f, g\colon g_{|x_{k\cdots\ell-1}} = f} \mathrm{Tr}_\rho\big(P^f_{\boldsymbol{x}_{\geq k}} \otimes Q^g_{\boldsymbol{x}_{\geq \ell}}\big),$$

where $g_{|\boldsymbol{x}_{k\cdots\ell-1}}$ is the $(n-\ell)$-linear function obtained by restricting g's $(\ell-k)$ first variables to $\boldsymbol{x}_{k\cdots\ell-1}$, and their *inconsistency*

$$\textsc{inc}(P,Q) := \mathrm{E}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{f,g:\, g_{|x_{k\cdots\ell-1}}\neq f} \mathrm{Tr}_\rho\big(P^f_{\boldsymbol{x}_{\geq k}} \otimes Q^g_{\boldsymbol{x}_{\geq \ell}}\big),$$

where ρ is a density matrix which will always be clear from the context. If $k > \ell$ then we define $\textsc{con}(P,Q) := \textsc{con}(Q,P)$ and $\textsc{inc}(P,Q) := \textsc{inc}(Q,P)$. We will also use shorthands $\textsc{con}(P) = \textsc{con}(P,P)$ and $\textsc{inc}(P) = \textsc{inc}(P,P)$. Note that if P is a complete family of measurements, i.e. $\sum_f P^f_{\boldsymbol{x}_{\geq k}} = \mathrm{Id}$ for every $\boldsymbol{x}_{\geq k}$, then

$$\textsc{con}(P,Q) + \textsc{inc}(P,Q) \;=\; \mathrm{E}_{\boldsymbol{x}} \sum_g \mathrm{Tr}_\rho\big(Q^g_{\boldsymbol{x}_{\geq\ell}}\big) \;=\; \mathrm{Tr}_\rho(Q),$$

which equals 1 if Q is also complete.

2.2 Multi-prover interactive proofs

In this section we define the complexity classes that this work is concerned with: multi-prover interactive proof systems (MIP systems) and multi-prover interactive proof systems *with entanglement* (MIP* systems).

Let $k(n)$ be an integer, denoting the number of provers, and $m(n)$ an integer denoting the number of rounds. Both $k(n)$ and $m(n)$ are from the set of polynomially bounded, polynomial-time computable functions in the input size $|x|$, denoted by poly. Further, c and s denote polynomial-time computable functions of the input size into $[0,1]$ corresponding to completeness acceptance probability and soundness error. For notational convenience in what follows we will omit the arguments of these functions.

Multi-prover interactive proof systems (MIP **systems):** Let $k, m, l \in$ poly. A k-prover interactive proof system consists of a verifier V and k provers $P_1, \ldots, P_k$. The verifier is a probabilistic polynomial-time Turing machine, and the provers are computationally unbounded. Each of them has a read-only input tape and a private work tape. Each prover has a communication tape. The verifier has a random tape. The verifier also has k communication tapes, one for each prover, each of which is l bits long.

The input tape for every party contains the same input string x. The protocol consists of $m(|x|)$ rounds. In each round, first the verifier runs for a polynomial amount of time, updating the work and communication tapes. After that, the content of the ith communication tape is sent to the ith prover for each $i = 1, \ldots, k(|x|)$. Each prover reads this string, updates the content of his own work tape, and decides a reply to the verifier. The reply from the ith prover is written in the ith communication tape, and this round completes. After $m(|x|)$ rounds of interaction, the verifier produces a special output bit, designating acceptance or rejection. The operations by provers are instantaneous and do not have to be even computable; the provers are assumed to be able to "compute" any function.

For simplicity, we assume that each message between the verifier and the provers in each round is exactly l bits long for the purpose of a formal definition, but it is not hard to modify the definition to incorporate the more general case which does not satisfy this assumption. Formally, a *strategy* for $P_1, \ldots, P_k$ in a k-prover m-round interactive proof system consists of the length $l' \in \mathbb{N}$ of a work tape, and km mappings $f_{ij}\colon \{0,1\}^l \times \{0,1\}^{l'} \to \{0,1\}^l \times \{0,1\}^{l'}$ for $1 \leq i \leq k$ and $1 \leq j \leq m$. Each mapping f_{ij} specifies the operation which prover i performs in round j:

$f_{ij}(q, w) = (q', w')$ means that if the message from the verifier in this round is q and the work tape contains string w before the operation by the prover, then the message to the verifier in this round is q' and the work tape contains string w after the operation.

Definition 4. *Let $k, m \colon \mathbb{N} \to \mathbb{N}$, and let $c, s \colon \mathbb{N} \to [0, 1]$ such that $c(n) > s(n)$ for all $n \in \mathbb{N}$. A language L is in* $\mathrm{MIP}(k, m, c, s)$ *if and only if there exists an m-round polynomial-time verifier V for a k-prover interactive proof system such that, for every input x:*

(Completeness) *if $x \in L$, there exists a strategy for provers $P_1, \ldots, P_k$ such that the interaction protocol of V with $(P_1, \ldots, P_k)$ results in the verifier accepting with probability at least c,*

(Soundness) *if $x \notin L$, for any strategy for provers $P'_1, \ldots, P'_k$, the probability that the interaction protocol of V with $(P_1, \ldots, P_k)$ results in the verifier accepting is at most s.*

In this formulation, the provers are deterministic, but this is not a limitation because it is well-known that the power of the model does not change if we allow the provers to share a random source.

If some of the parameters k, m, c, and s are sets of functions instead of single functions, the class is interpreted to be the union over all choices in the sets. For example,

$$\mathrm{MIP}(6, 1, 1, 1 - 1/\operatorname{poly}) = \bigcup_{f \in \mathrm{poly}} \mathrm{MIP}(6, 1, 1, 1 - 1/f).$$

We denote $\mathrm{MIP}(\mathrm{poly}, \mathrm{poly}, 2/3, 1/3)$ simply by MIP.

Multi-prover interactive proof systems with entanglement (MIP^* **systems):** First introduced in [CHTW04], MIP^* systems are defined analogously to MIP systems. The only difference is that now the provers are allowed to be *quantum*, while the verifier (and communication) remains bounded in classical probabilistic polynomial-time. This implies that the provers may share an arbitrary entangled state $|\Psi\rangle$ among themselves before the protocol starts and that each prover may use his part of the entangled state to determine his reply to the verifier. In each round, the provers individually receive the messages from the verifier in a message register, perform a quantum operation on this register together with their share of the entangled state, measure the message register in the computational basis, and send back the outcome to the verifier.

Formally, an *entangled strategy* for $P_1, \ldots, P_k$ in a k-prover m-round interactive proof system with entanglement consists of the length $l' \in \mathbb{N}$ of a work tape, km quantum channels Φ_{ij} from a quantum register of $l + l'$ qubits to itself for $1 \leq i \leq k$ and $1 \leq j \leq m$, and the initial quantum state $|\Psi\rangle$ of the work tape, which is a kl'-qubit state. Each channel Φ_{ij} specifies the operation which prover i performs in round j: the first l qubits in the state correspond to the message from and to the verifier, and the last l' qubits represent the content of the work tape. After the prover's operation, the first l qubits are measured in the computational basis and sent to the verifier.

Definition 5. *A language L is in* $\mathrm{MIP}^*(k, m, c, s)$ *if and only if there exists an m-round polynomial-time verifier V for k-prover interactive proof systems such that, for every input x:*

(Completeness) *if $x \in L$, there exists an entangled strategy for provers $P_1, \ldots, P_k$ such that the interaction protocol of V with $(P_1, \ldots, P_k)$ results in the verifier accepting with probability at least c,*

(Soundness) *if $x \notin L$, for any entangled strategy for provers $P'_1, \ldots, P'_k$, the probability that the interaction protocol of V with $(P_1, \ldots, P_k)$ results in the verifier accepting is at most s.*

In certain cases, we can simplify part of the definition of entangled strategies. Suppose that the verifier interacts with certain prover P_i only once; i.e., the verifier is guaranteed to send P_i the empty string (or a fixed string) in rounds other than round j, and is guaranteed to ignore the reply from P_i in rounds other than round j. In this case, instead of specifying m quantum channels to describe the behavior of P_i in the m rounds, we may just specify measurements $A_q = (A_q^r)$ for each message q from the verifier, where the outcome of each measurement gives a reply to the verifier.[9] Since all the interactive proof systems considered in this paper have the property that the verifier interacts with each prover only once except for one prover, we use this simplified formulation in many places.

Note that we do not assume any upper bound on the size l' of the work tape used by each prover (in particular, we do not assume that $l' \in \mathrm{poly}$; the model with this restriction is considered in [KM03]). However, we do assume that they only use a finite-dimensional Hilbert space. A more general definition is commuting-operator provers, considered by Tsirelson [Tsi80] in the context of Bell inequalities and later in [SW08, DLTW08, NPA08, IKPSY08]. Although we expect that our results remain valid with minor modifications to the proofs even if dishonest provers are allowed to use arbitrary commuting-operator strategies, we have not explored this possibility.

Symmetry. We will make an important use of symmetry in the protocols that we introduce. It will be a useful simplifying assumption in two respects: first it lets one assume that the set of measurements used by all provers is the same. Second, and most important, it implies that the provers' shared entangled state is also permutation-invariant.

Definition 6. *Let $(P_1, \ldots, P_k, |\Psi\rangle)$ be a k-prover strategy.*[10] *We say that this strategy is symmetric, or permutation-invariant, if $P_1 = \cdots = P_k$ and $|\Psi\rangle$ is invariant with respect to any permutation of the subsystems corresponding to each prover.*

The following simple lemma (which already appears in [KKMTV11, Lemma 4]) shows that one can always assume without loss of generality that if a game has a certain symmetry then there is an optimal strategy for the provers which reflects that symmetry.

Lemma 7. *Suppose an* MIP* *proof system is given such that the protocol treats provers $P_1, \ldots, P_k$ symmetrically (i.e. the protocol is invariant under permutation of their questions and corresponding inverse-permutation of their answers). Then given any strategy $P_1, \ldots, P_k$ with entangled state $|\Psi\rangle$ that succeeds with probability p, there exists a strategy $P'_1, \ldots, P'_k$ with entangled state $|\Psi'\rangle$ and success probability p such that $P'_1 = \cdots = P'_k$ and $|\Psi'\rangle$ is permutation-invariant.*

Proof. By appropriately padding with extra qubits, assume that all k registers of $|\Psi\rangle$ have the same dimension. Define strategies $P'_1, \ldots, P'_k$ as follows: the provers share the entangled state $|\Psi'\rangle = \sum_{\sigma \in \mathfrak{S}_k} |\sigma(1)\rangle \otimes \cdots \otimes |\sigma(k)\rangle \otimes |\Psi^\sigma\rangle$, where the register containing $|\sigma(i)\rangle$ is given to prover i and $|\Psi^\sigma\rangle$ is obtained from $|\Psi\rangle$ by permuting its registers according to σ. For $1 \leq i \leq k$ prover i measures the register containing $|\sigma(i)\rangle$ and behaves as in the strategy $P_{\sigma(i)}$. By the assumed symmetry of the protocol this new strategy has the same success probability p, and $|\Psi'\rangle$ has the required symmetry properties. □

The following claim states a trivial but useful fact about symmetric one-round strategies.

[9] Any classical post-processing by the prover can be incorporated as part of the description of his measurement.

[10] We think of P_i as an arbitrary representation of the set of all quantum channels applied by prover i throughout the protocol.

Claim 8. *Let $(P_1, \ldots, P_k, |\Psi\rangle)$ be a symmetric one-round strategy, and for every $i \in \{1, \ldots, k\}$, $\{A_i^a\}_a$ a measurement for the i-th prover in that strategy. Then for every permutation σ on $\{1, \ldots, k\}$, and every $(a_1, \ldots, a_k)$,*

$$\langle\Psi|A_1^{a_1} \otimes \cdots \otimes A_k^{a_k}|\Psi\rangle = \langle\Psi|A_{\sigma(1)}^{a_{\sigma(1)}} \otimes \cdots \otimes A_{\sigma(k)}^{a_{\sigma(k)}}|\Psi\rangle.$$

2.3 NEXP-complete problems

We will use the following NEXP-complete problem, as stated in Proposition 4.2 of Ref. [BFL91]:

Problem 1: Oracle-3-satisfiability.
Instance. Integers $r, n \in \mathbb{N}$ in unary and a Boolean formula $B(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, a_1, a_2, a_3)$ in variables $\boldsymbol{z} \in \{0,1\}^r$, $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \{0,1\}^n$ and $a_1, a_2, a_3 \in \{0,1\}$.

Question. Does there exist $A\colon \{0,1\}^n \to \{0,1\}$ such that $B(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, A(\boldsymbol{b}_1), A(\boldsymbol{b}_2), A(\boldsymbol{b}_3)) = 1$ simultaneously for all $\boldsymbol{z} \in \{0,1\}^r$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \{0,1\}^n$?

Using the standard technique of arithmetization (see e.g. Proposition 3.1 and Lemma 7.1 of Ref. [BFL91]), one can show that the following problem is also NEXP-complete.

Problem 2: Oracle-3-satisfiability, arithmetized version.
Instance. Integers $r, n \in \mathbb{N}$ in unary and an arithmetic expression[11] for a polynomial $f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, a_1, a_2, a_3)$, where $\boldsymbol{z}$ represents r variables and each of $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ represents n variables.

Yes-promise. There exists an $A\colon \{0,1\}^n \to \{0,1\}$ such that for all $\boldsymbol{z} \in \{0,1\}^r$ and all $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \{0,1\}^n$, it holds that

$$f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, A(\boldsymbol{b}_1), A(\boldsymbol{b}_2), A(\boldsymbol{b}_3)) = 0 \tag{3}$$

in $\mathbb{Z}$ (and therefore in every field).

No-promise. For every pair $(\mathbb{F}, A)$ of a field $\mathbb{F}$ and a mapping $A\colon \{0,1\}^n \to \mathbb{F}$, there exist $\boldsymbol{z} \in \{0,1\}^r$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \{0,1\}^n$ such that Eq. (3) is not satisfied in $\mathbb{F}$.

We note that the degree of the polynomial f represented by the arithmetic expression can be at most the size of the arithmetic expression, and is therefore bounded by the input size.

2.4 Summation test

Let $\mathbb{F}$ be a finite field of characteristic two.[12] If $|\mathbb{F}| = 2^k$, an encoding scheme of elements in $\mathbb{F}$ is specified by k and a primitive polynomial $f(t)$ over $\mathbb{F}_2$ of degree k. It is well-known that given 1^k, $f(t)$, and the complete factorization of $2^k - 1$ along with the certificate that each factor in the factorization is indeed a prime (such as the Pratt certificate), it is possible to check that k and $f(t)$ form a valid encoding scheme of the field $\mathbb{F}$ in polynomial time.

Consider the following promise problem, which has both an explicit and an implicit input.

Problem 3: Summation Test Problem.
Explicit input. Integers $m, d \in \mathbb{N}$ in unary, and an encoding scheme of a finite field $\mathbb{F}$ of characteristic two.

Implicit input. A mapping $h\colon \mathbb{F}^m \to \mathbb{F}$.

[11] An *arithmetic expression* is a rooted tree whose internal nodes represent either addition or multiplication and whose leaves represent either variables or an integer constant. The size of an arithmetic expression is the number of nodes plus the sum of the number of bits required to represent the integer for each constant node.

[12] The restriction to fields of characteristic two arises from the use of Theorem 43 in Appendix C.

Promise. The given encoding scheme is valid, and the mapping $h\colon \mathbb{F}^m \to \mathbb{F}$ is a polynomial function of degree at most d in each variable.

Question. Is

$$\sum_{\boldsymbol{x}\in\{0,1\}^m} h(\boldsymbol{x}) = 0 \quad (\text{in } \mathbb{F})? \tag{4}$$

In a (single-prover) interactive proof system for a problem with an implicit input, the implicit input is given to the verifier as an oracle.[13] The following variant of the summation test of Lund, Fortnow, Karloff, and Nisan [LFKN92] is a special case of Lemma 3.5 in Ref. [BFL91].

Lemma 9 (Summation test [BFL91])**.** *Suppose that $|\mathbb{F}| \geq 2dm$. Then there exists a single-prover interactive proof system for the Summation Test Problem with perfect completeness and soundness error at most $dm/|\mathbb{F}|$. Moreover, in this interactive proof system, the verifier behaves as follows. First he chooses $\boldsymbol{q} \in \mathbb{F}^m$ uniformly at random. Then he interacts with the prover. At the same time, he reads the value $h(\boldsymbol{q})$ from the implicit input. Finally he accepts or rejects depending on $\boldsymbol{q}$, $h(\boldsymbol{q})$, and the interaction with the prover.*[14]

To apply the summation test to Problem 2, we have to consider exponentially many constraints instead of one.

Problem 4: AND Test Problem.

Explicit input. Integers $k, d \in \mathbb{N}$ in unary, and an encoding scheme of a finite field $\mathbb{F}$ of characteristic two.

Implicit input. A mapping $h\colon \mathbb{F}^k \to \mathbb{F}$.

Promise. The given encoding scheme is valid, and the mapping $h\colon \mathbb{F}^k \to \mathbb{F}$ is a polynomial function of degree at most d in each variable.

Question. Is $h(\boldsymbol{i}) = 0$ (in $\mathbb{F}$) for all $\boldsymbol{i} \in \{0,1\}^k$?

The idea for the following corollary is already explained in Section 7.1 of Ref. [BFL91]. We will give a proof in Appendix C for the sake of completeness.

Corollary 10. *There exists a polynomial $q\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ for which the following holds. There exists a single-prover interactive proof system for the AND Test Problem with perfect completeness and soundness error at most $5/8 + q(k,d)/|\mathbb{F}|$. Moreover, in this interactive proof system, the verifier behaves as follows. First he chooses $\boldsymbol{i} \in \mathbb{F}^k$ uniformly and independently at random. Then he interacts with the prover. At the same time, he reads the value $h(\boldsymbol{i})$ from the implicit input. Finally he accepts or rejects depending on $\boldsymbol{i}$, $h(\boldsymbol{i})$, and the interaction with the prover.*

3 A proof system for Oracle-3-Satisfiability

In this section we prove Theorem 1, assuming the soundness analysis of the multilinearity game (see Theorem 11 below), which will be given in Sections 4 and 5. We first describe a four-prover

[13] In Ref. [BFL91], the authors refer to the interactive proof system for the Summation Test Problem as an "interactive oracle-protocol," viewing the mapping h as an exponentially long certificate string which is given to the verifier as an oracle. However, for our purposes it will be more convenient to treat h as part of the input.

[14] In particular, this implies that the verifier reads only one value $h(\boldsymbol{q})$ from the implicit input and the position $\boldsymbol{q} \in \mathbb{F}^m$ to read is chosen uniformly in $\mathbb{F}^m$. Together with the soundness guarantee, this in turn implies that if the implicit input is δ-close to a polynomial function $\tilde{h}$ of degree at most d in each variable and $\tilde{h}$ fails to satisfy the equation (4), then the verifier accepts with probability at most $\delta + dm/|\mathbb{F}|$ no matter what the prover does.

poly-round proof system for the NEXP-complete Oracle-4-satisfiability problem, Problem 2, in Section 3.1. In Section 3.2 we show that the protocol has perfect completeness with classical provers, and in Section 3.3 we show that it has soundness error at most $1 - 1/\operatorname{poly}$ with entangled provers. Theorem 1 is then obtained by repeating this protocol sequentially.

3.1 Description of the protocol

We construct a four-prover poly-round proof system for Problem 2. Our protocol follows that of [BFL91] very closely. We replace the three queries to the Oracle in their protocol by queries to three distinct provers.

Label the provers as P, X_1, X_2, X_3. The protocol will be symmetric under any permutation of the three provers X_1, X_2, X_3. Let (r, n, f) be an instance of Problem 2, as described in Section 2.3. Let d_f be the maximum degree of f in any one variable. Let $m = r + 3n$ and $d = 2d_f$. Let $0 < c_0 < 1$ be a constant defined later (in Theorem 11), and p be the smallest power of two such that $p > \max\{8q(m,d), n^{1/c_0+4}\}$, where q is the polynomial appearing in the statement of Corollary 10. Let $\mathbb{F}$ be the finite field of size p.

The verifier first receives an encoding scheme for $\mathbb{F}$ and its certificate from P (as defined in Section 2.4), and rejects if it is not valid. In the rest of the protocol, all arithmetic operations in $\mathbb{F}$ are performed using this encoding scheme. If it is valid, the verifier performs one of the following three tests chosen uniformly at random:

- *Consistency test.* He chooses $\boldsymbol{x} \in \mathbb{F}^n$ uniformly at random and sends the same question $\boldsymbol{x}$ to provers X_1, X_2, X_3. He expects each prover to answer with an element of $\mathbb{F}$, and accepts if and only if all the answers are equal.
- *Linearity test.* He chooses $i \in \{1, \ldots, n\}$, $\boldsymbol{x} \in \mathbb{F}^n$ and $y_i \neq z_i \in \mathbb{F} \backslash \{x_i\}$ uniformly at random, and sets $y_j = z_j = x_j$ for every $j \in \{1, \ldots, n\} \setminus \{i\}$. He sends $\boldsymbol{x}$ to X_1, $\boldsymbol{y}$ to X_2, and $\boldsymbol{z}$ to X_3. He receives $a, b, c \in \mathbb{F}$ from these three provers, and accepts if and only if
$$\frac{b-a}{y_i - x_i} = \frac{c-b}{z_i - y_i} = \frac{c-a}{z_i - x_i}.$$
- *Summation test.* The verifier simulates the interactive proof system from Corollary 10 with the explicit input (m, d) and prover P. When the verifier in Corollary 10 tries to read the value $h(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ in the implicit input, where $\boldsymbol{z} \in \mathbb{F}^r$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \mathbb{F}^n$, our verifier simulates this by sending $\boldsymbol{b}_1$ to X_1, $\boldsymbol{b}_2$ to X_2, and $\boldsymbol{b}_3$ to X_3. Upon obtaining answers a_1, a_2, a_3 to his queries from these three provers, he evaluates $f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, a_1, a_2, a_3)$ and uses the result as the value of $h(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$.

Note that in each of the three tests, each of the provers X_1, X_2, X_3 is asked a question $\boldsymbol{x} \in \mathbb{F}^n$ distributed uniformly at random.

3.2 Completeness

Let (r, n, f) be a yes-instance of Problem 2. Then there exists a mapping $A\colon \{0,1\}^n \to \{0,1\}$ such that Eq. (3) is satisfied for all $\boldsymbol{z} \in \{0,1\}^r$ and all $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \{0,1\}^n$ simultaneously. Let g be the unique extension of A to a multilinear function $g\colon \mathbb{F}^n \to \mathbb{F}$. Each of X_1, X_2, X_3 answers $g(\boldsymbol{b})$ on question $\boldsymbol{b} \in \mathbb{F}^n$, while P behaves as it should in the AND test. It is clear that this deterministic strategy is accepted with certainty in the consistency test and the linearity test. In the summation test, note that the value of $h(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ which the verifier uses is given by

$$h(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3) = f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, g(\boldsymbol{b}_1), g(\boldsymbol{b}_2), g(\boldsymbol{b}_3)),$$

which is a polynomial in $\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ of degree at most $2d_f = d$ in each variable. Therefore, the promise in the AND test is satisfied and prover P has a strategy which makes the verifier accept with certainty.

3.3 Soundness

The soundness analysis is divided in two parts. First we analyze the consistency and linearity tests, which only involve the three provers X_1, X_2, X_3, and show that success in those tests implies the following. (We refer the reader to Section 2 for some relevant notation and definitions.)

Theorem 11. *There exist universal constants $0 < c_0, c < 1$, $C > 1$ such that the following holds. Let $n \geq 1$ be an integer. Let $\mathbb{F}$ be a finite field, and $(|\Psi\rangle, \{A_{\boldsymbol{x}}^a\})$ a (symmetric, projective) strategy for the provers in the three-player multilinearity game in n variables over $\mathbb{F}$ (as defined in Definition 13 below) that passes both the consistency and the linearity tests with probability at least $1-\varepsilon$. Assume furthermore that $p := |\mathbb{F}| \geq n^4 \varepsilon^{-1/2}$ and $\varepsilon \leq n^{-2/c_0}$. Then there exists a sub-measurement $\{V^g\}$, indexed by multilinear $g \in \mathrm{ML}(\mathbb{F}^n, \mathbb{F})$, such that*

$$\mathrm{E}_{\boldsymbol{x}} \sum_a \mathrm{Tr}_\rho\big((A_{\boldsymbol{x}}^a - \sqrt{V_{\boldsymbol{x}}^a})^2\big) \;\leq\; C\,\varepsilon^c, \tag{5}$$

where for every $\boldsymbol{x} \in \mathbb{F}^n$ and $a \in \mathbb{F}$ we defined $V_{\boldsymbol{x}}^a := \sum_{g:\, g(\boldsymbol{x})=a} V^g$.

The proof of Theorem 11 is our main technical contribution, and it is given in Sections 4 and 5. Assuming the theorem, we prove that our proof system has soundness error at most $1 - n^{-2/c_0}/3$, provided n is larger than an absolute constant depending on c, c_0, and C.

Let (r, n, f) be a no-instance. Toward contradiction, suppose that the provers have a symmetric[15] entangled strategy S whose acceptance probability is at least $1 - \varepsilon/3$, where $\varepsilon = n^{-2/c_0}$. Let $|\Psi\rangle \in \mathcal{P} \otimes \mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{X}_3$ be the state used in the strategy S and $(A_{\boldsymbol{x}}^a)_{a\in\mathbb{F}}$ be the projective measurements used by each of the provers X_1, X_2, X_3 in the strategy S.

The verifier can be viewed as playing the multilinearity game with the players X_1, X_2, X_3 with probability $2/3$ and performing something else, namely the summation test, with probability $1/3$. Therefore, the marginal strategy of S for players X_1, X_2, X_3 has winning probability at least $1 - \varepsilon/2$ in the multilinearity test. Because $|\mathbb{F}| = p > n^{1/c_0+4} = n^4\varepsilon^{-1/2}$, Theorem 11 implies that there exists a sub-measurement $\{V^g\}_{g\in\mathrm{ML}(\mathbb{F}^n,\mathbb{F})}$ such that inequality (5) holds, where ρ is the reduced state of $|\Psi\rangle\langle\Psi|$ on $\mathcal{X}_1$. For every $\boldsymbol{x} \in \mathbb{F}^n$ and $a \in \mathbb{F}$, let

$$V_{\boldsymbol{x}}^a = \sum_{\substack{g\in\mathrm{ML}(\mathbb{F}^n,\mathbb{F})\\ g(\boldsymbol{x})=a}} V^g.$$

For $0 \leq i \leq 3$, let S_i be the entangled strategy obtained from S by replacing the measurement for the first i provers $X_1, \ldots, X_i$ by $V_{\boldsymbol{x}}^a$.[16] Note $S_0 = S$. In the strategy S_i, the provers $X_1, \ldots, X_i$ can be implemented so that they measure the prior entanglement without looking at their question. In

[15] Lemma 7 shows that we may assume this holds without loss of generality.

[16] Since V is a sub-measurement, the $V_{\boldsymbol{x}}^a$ may not sum to identity. In that case we introduce an additional outcome "fail", corresponding to the element $\mathrm{Id} - \sum_a V_{\boldsymbol{x}}^a$. Whenever a prover obtains that outcome he aborts the protocol.

particular, in the strategy S_3, every prover except for P measures the prior entanglement without looking at the question, and therefore S_3 can be implemented using shared randomness alone.

For $0 \leq i \leq 3$, let p_i be the probability that the strategy S_i is accepted in the four-prover protocol. By definition, $p_0 \geq 1 - \varepsilon/3$. We prove the following.

Claim 12. *For $i = 1, \ldots, 3$, it holds that $|p_{i-1} - p_i| \leq \sqrt{C}\varepsilon^{c/2}$.*

Proof. The only difference between strategies S_{i-1} and S_i is the measurements used by prover X_i. We call the message from the verifier to X_i as register $\mathcal{A}$, and call everything other than $\mathcal{A}$ and the private space $\mathcal{X}_i$ for prover X_i as register $\mathcal{B}$. Register $\mathcal{A}$ is classical, but we treat it as a quantum register which always contains a state in the computational basis. Let σ be the global state before the prover X_i performs his measurement, and σ_A (resp. σ_V) be the global state after the prover X_i performs the measurement A_x (resp. V) on his share of the state, and then discards the post-measurement state. Since the marginal distribution on the question to X_i is uniform, the state σ has the following form:

$$\sigma = \mathrm{E}_{\boldsymbol{x} \in \mathbb{F}^n} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|_{\mathcal{A}} \otimes \sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}},$$

where $\mathrm{Tr}_{\mathcal{B}}\sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}} = \sigma^{\mathcal{X}_i} = \rho$ is independent of $\boldsymbol{x}$. We want to bound $(1/2)\|\sigma_W - \sigma_M\|_1$, where

$$\sigma_W = \mathrm{Tr}_{\mathcal{X}_i}\left[\mathrm{E}_{\boldsymbol{x} \in \mathbb{F}^n} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|_{\mathcal{A}} \otimes \sum_{a \in \mathbb{F}} |a\rangle\langle a|_{\mathcal{C}} \otimes (A_{\boldsymbol{x}}^a \otimes I_{\mathcal{B}})\sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}}(A_{\boldsymbol{x}}^a \otimes I_{\mathcal{B}})\right],$$
$$\sigma_M = \mathrm{Tr}_{\mathcal{X}_i}\left[\mathrm{E}_{\boldsymbol{x} \in \mathbb{F}^n} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|_{\mathcal{A}} \otimes \sum_{a \in \mathbb{F}} |a\rangle\langle a|_{\mathcal{C}} \otimes (\sqrt{V_{\boldsymbol{x}}^a} \otimes I_{\mathcal{B}})\sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}}(\sqrt{V_{\boldsymbol{x}}^a} \otimes I_{\mathcal{B}})\right],$$

and $\mathcal{C}$ denotes the register used for prover i's answers. For $\boldsymbol{x} \in \mathbb{F}^n$, define isometries $U_{\boldsymbol{x}}, V_{\boldsymbol{x}} \colon \mathcal{X}_i \otimes \mathcal{B} \to \mathcal{X}_i \otimes \mathcal{B} \otimes \mathcal{C}$ by

$$U_{\boldsymbol{x}} = \sum_{a \in \mathbb{F}} A_{\boldsymbol{x}}^a \otimes I_{\mathcal{B}} \otimes |a\rangle_{\mathcal{C}},$$
$$V_{\boldsymbol{x}} = \sum_{a \in \mathbb{F}} \sqrt{V_{\boldsymbol{x}}^a} \otimes I_{\mathcal{B}} \otimes |a\rangle_{\mathcal{C}}.$$

Then,

$$\begin{aligned}
&\|\sigma_W - \sigma_M\|_1 \\
&\leq \left\|\mathrm{E}_{\boldsymbol{x} \in \mathbb{F}^n} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|_{\mathcal{A}} \otimes \sum_{a \in \mathbb{F}} |a\rangle\langle a|_{\mathcal{C}} \otimes \big((A_{\boldsymbol{x}}^a \otimes I_{\mathcal{B}})\sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}}(A_{\boldsymbol{x}}^a \otimes I_{\mathcal{B}}) - (\sqrt{V_{\boldsymbol{x}}^a} \otimes I_{\mathcal{B}})\sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}}(\sqrt{V_{\boldsymbol{x}}^a} \otimes I_{\mathcal{B}})\big)\right\|_1 \\
&\leq \mathrm{E}_{\boldsymbol{x} \in \mathbb{F}^n}\left\|\sum_{a \in \mathbb{F}} |a\rangle\langle a|_{\mathcal{C}} \otimes \big((A_{\boldsymbol{x}}^a \otimes I_{\mathcal{B}})\sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}}(A_{\boldsymbol{x}}^a \otimes I_{\mathcal{B}}) - (\sqrt{V_{\boldsymbol{x}}^a} \otimes I_{\mathcal{B}})\sigma_{\boldsymbol{x}}^{\mathcal{X}_i\mathcal{B}}(\sqrt{V_{\boldsymbol{x}}^a} \otimes I_{\mathcal{B}})\big)\right\|_1 \\
&\leq 2\mathrm{E}_{\boldsymbol{x} \in \mathbb{F}^n}\sqrt{\sum_{a \in \mathbb{F}} \mathrm{Tr}\big((A_{\boldsymbol{x}}^a - \sqrt{V_{\boldsymbol{x}}^a})^2\rho\big)} \\
&\leq 2\sqrt{\mathrm{E}_{\boldsymbol{x} \in \mathbb{F}^n}\sum_{a \in \mathbb{F}} \mathrm{Tr}\big((A_{\boldsymbol{x}}^a - \sqrt{V_{\boldsymbol{x}}^a})^2\rho\big)} \\
&\leq 2\sqrt{C\varepsilon^c},
\end{aligned}$$

where the third inequality is by Lemma 35, the fourth is by convexity and the last by (5). Therefore, we have that $|p_{i-1} - p_i| \leq (1/2)\|\sigma_W - \sigma_M\|_1 \leq \sqrt{C}\varepsilon^{c/2}$ as claimed. □

By the triangle inequality, Claim 12 implies that $|p_0 - p_3| \le 3\sqrt{C}\varepsilon^{c/2}$, and therefore

$$p_3 \ge p_0 - 3\sqrt{C}\varepsilon^{c/2} \ge 1 - \varepsilon/3 - 3\sqrt{C}\varepsilon^{c/2} \ge 1 - 4\sqrt{C}\varepsilon^{c/2},$$

where the last inequality uses $c \le 1$ and $C \ge 1$.

If not all of the provers choose the same multilinear function, then they pass in the consistency test with probability at most $n/|\mathbb{F}| \le 1/6$ by the Schwartz-Zippel lemma [Sch80, Zip79] (see Lemma 33 in Appendix A for a statement). In the strategy S_3, they pass in the consistency test with probability at least $1 - 12\sqrt{C}\varepsilon^{c/2}$. Therefore, they choose the same multilinear function with probability at least $1 - 12\sqrt{C}\varepsilon^{c/2}/(1-1/6) \ge 1 - 15\sqrt{C}\varepsilon^{c/2}$. This implies that if an oracle chooses a multilinear function in the same way as the prover X_1 and uses it for all three queries, the distribution on their answers will differ by at most $2 \cdot 15\sqrt{C}\varepsilon^{c/2} = 30\sqrt{C}\varepsilon^{c/2}$ in statistical distance. Therefore, this oracle (which always implements a multilinear function) together with the prover P is accepted in the interactive proof system of Corollary 10 with probability at least $1 - 12\sqrt{C}\varepsilon^{c/2} - 30\sqrt{C}\varepsilon^{c/2} = 1 - 42\sqrt{C}\varepsilon^{c/2}$.

Because (r, n, f) is a no-instance of Problem 2 and $|\mathbb{F}| = p > 8q(m, d)$, the acceptance probability in the interactive proof system of Corollary 10 is less than 3/4. Comparing this with the lower bound in the previous paragraph, we obtain

$$1 - 42\sqrt{C}\varepsilon^{c/2} < \frac{3}{4},$$

which implies

$$\varepsilon > \frac{1}{(168^2 \cdot C)^{1/c}},$$

contradicting the definition $\varepsilon = n^{-c_0/2}$ as soon as n is large enough. Since we obtained this contradiction from the assumption that there exists an entangled strategy with acceptance probability at least $1 - \varepsilon/3$, we have proved the claimed soundness guarantee against entangled provers.

4 The multilinearity game

In this section we analyze the combination of the consistency test and the linearity test described in Section 3 as a stand-alone game played between a referee and $r \ge 3$ players, which we call the *r-player multilinearity game in n variables over* $\mathbb{F}$. The game is parametrized by an integer n and a finite field $\mathbb{F}$ of arbitrary size $p = |\mathbb{F}|$ (which is not necessarily a prime), and it is performed with r players $X_1, \ldots, X_r$ treated symmetrically. The referee performs either of the following two tests with probability 1/2 each:

- *Consistency test.* The referee chooses $\boldsymbol{x} \in \mathbb{F}^n$ uniformly at random and sends the same question $\boldsymbol{x}$ to all players $X_1, \ldots, X_r$. He expects each player to answer with an element of $\mathbb{F}$, and accepts if and only if all the answers are equal.
- *Linearity test.* The referee chooses $i \in \{1, \ldots, n\}$, $\boldsymbol{x} \in \mathbb{F}^n$ and $y_i \ne z_i \in \mathbb{F} \setminus \{x_i\}$ uniformly at random, and sets $y_j = z_j = x_j$ for every $j \in \{1, \ldots, n\} \setminus \{i\}$. He sends $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ to 3 out of the r players chosen at random, receives $a, b, c \in \mathbb{F}$, and accepts if and only if

$$\frac{b-a}{y_i - x_i} = \frac{c-b}{z_i - y_i} = \frac{c-a}{z_i - x_i}.$$

We now define explicitly what we mean by a *strategy* for the players in the multilinearity game.

Definition 13. *A* strategy for the players in the r-player multilinearity game in n variables over $\mathbb{F}$ *is given by the following. Finite-dimensional Hilbert spaces $\mathcal{X}_1, \ldots, \mathcal{X}_r$ and $\mathcal{P}$, a state $|\Psi\rangle \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_r \otimes \mathcal{P}$, and for every $i \in [r]$ and $\boldsymbol{x} \in \mathbb{F}^n$ a measurement $\{(A_i)_{\boldsymbol{x}}^a\}_{a\in\mathbb{F}}$ on $\mathcal{X}_i$. It is understood that, upon receiving question $\boldsymbol{x}_i \in \mathbb{F}^n$, player i measures register X_i corresponding to his share of $|\Psi\rangle$ using the measurement $\{(A_i)_{\boldsymbol{x}_i}^a\}_{a\in\mathbb{F}}$, sending the outcome a back to the verifier as his answer.*

We will say that a strategy is symmetric *if $\mathcal{X}_1 \simeq \cdots \simeq \mathcal{X}_r$, $(A_1)_{\boldsymbol{x}}^a = \cdots = (A_r)_{\boldsymbol{x}}^a$ for every $\boldsymbol{x}$ and a (in which case we will simply call the resulting measurement $\{A_{\boldsymbol{x}}^a\}$), and $|\Psi\rangle$ is invariant with respect to arbitrary permutation of the registers $X_1, \ldots, X_r$.*

Finally, a strategy will be called projective *if all measurements $\{(A_i)_{\boldsymbol{x}}^a\}_{a\in\mathbb{F}}$ are projective.*

In case a strategy is symmetric, we will often abuse notation and use the symbol ρ to denote the reduced density of $|\Psi\rangle$ on any $\otimes_{i\in S}\mathcal{X}_i$, for $S \subseteq [r]$, without specifying explicitly which registers are understood: by symmetry only the number of registers matters, and this will always be clear in context.

The main result of this section is the following. We refer to Section 2.1 for definitions of the quantities appearing in the theorem, and to Lemma 7 for a proof that the symmetry assumption made in the theorem is without loss of generality.

Theorem 14. *There exists universal constants $0 < c_0 < 1$, $C_0 > 1$ such that the following holds. Let $(|\Psi\rangle, \{A_{\boldsymbol{x}}^a\}_a)$ be a permutation-invariant projective strategy for $r \geq 3$ players in the r-player multilinearity game in n variables over $\mathbb{F}$ with success probability at least $1 - \varepsilon/2$. Assume furthermore that $p = |\mathbb{F}| \geq n^4\varepsilon^{-1/2}$ and $\varepsilon \leq n^{-2/c_0}$. Then there exists a sub-measurement $\{V^g\}_{g\in\mathrm{ML}(\mathbb{F}^n,\mathbb{F})}$, indexed by multilinear $g : \mathbb{F}^n \to \mathbb{F}$, such that*

1. *V is consistent with A: $\mathrm{INC}(V, A) \leq C_0\, \varepsilon^{c_0}$,*
2. *$\mathrm{Tr}_\rho(V) \geq 1 - C_0\, \varepsilon^{c_0}$.*

The two items in the conclusion of the theorem intuitively state the following. Suppose that one of the players in the multilinearity game was to receive a question $\boldsymbol{x} \in \mathbb{F}^n$, measure his share of the entangled state $|\Psi\rangle$ according to the projective measurement $\{A_{\boldsymbol{x}}^a\}$, and answer the outcome he obtains (as he would in the original game). Now, suppose further that another player, upon receiving the same question $\boldsymbol{x} \in \mathbb{F}^n$, instead of measuring her own share of $|\Psi\rangle$ according to $\{A_{\boldsymbol{x}}^a\}$, was to perform the measurement $\{V^g, \mathrm{Id} - V\}$, where $V = \sum_g V^g$ (which is independent of $\boldsymbol{x}$!). If she obtains the last outcome then she aborts the experiment. If, however, she obtains an outcome $g \in \mathrm{ML}(\mathbb{F}^n, \mathbb{F})$, then she answers her question $\boldsymbol{x}$ with $g(\boldsymbol{x})$. Item 1. above states that, on average over the choice of $\boldsymbol{x}$, the probability that both players eventually produce different outcomes (conditioned on the second player not aborting) is at most $O(\varepsilon^{c_0})$. Item 2. guarantees that, in the hypothetical scenario we just described, the second player does not abort too often: the probability that she obtains the outcome "$\mathrm{Id} - V$" is at most $C_0\, \varepsilon^{c_0}$.

We will show that Theorem 14 implies Theorem 11 in Section 4.2, while Theorem 14 will be proved in Section 5. In the following section we prove a weaker version of the multilinearity test, the "linearity test", which implies Theorem 14 for $n = 1$.

4.1 Preliminary analysis: the linearity test

Let $(|\Psi\rangle, \{A_{\boldsymbol{x}}^a\}_a)$ be a symmetric projective strategy for the players in the multilinearity game, as defined in Definition 13. The following relations translate the assumption that the players succeed in the consistency test with probability $1-\varepsilon$, and in the linearity test with probability $1-\varepsilon$.

$$\mathrm{E}_{\boldsymbol{x}} \sum_a \mathrm{Tr}_\rho\big(A_{\boldsymbol{x}}^a \otimes A_{\boldsymbol{x}}^a\big) \;\geq\; 1-\varepsilon, \tag{6}$$

$$\forall i \in [n], \qquad \mathop{\mathrm{E}}_{x_i \neq x_i' \neq x_i'', \boldsymbol{x}_{\neg i}} \sum_{\frac{a'-a}{x_i'-x_i} = \frac{a''-a'}{x_i''-x_i'} = \frac{a''-a}{x_i''-x_i}} \mathrm{Tr}_\rho\big(A^a_{x_i,\boldsymbol{x}_{\neg i}} \otimes A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} \otimes A^{a''}_{x_i'',\boldsymbol{x}_{\neg i}}\big) \;\geq\; 1-n\varepsilon \;\geq\; 1-\sqrt{\varepsilon}, \tag{7}$$

where all expectations are taken under the uniform distribution over the sets in which their indices range, and the last inequality follows from our assumption that $n \leq \varepsilon^{-c_0/2} \leq \varepsilon^{-1/2}$.

The following claim proves the "linearity" part of the multilinearity test, thereby establishing the base case for the induction that will be performed in Section 5. It also illustrates some of the key techniques, in terms of the manipulation of measurement operators, that will be used throughout the paper. (The interested reader may thus wish to gain good familiarity with the proof of the claim before moving on to later sections, in which proofs will not always be as detailed.)

Claim 15. *Let $i \in [n]$, and $\varepsilon \geq p^{-1}$. Suppose that $(|\Psi\rangle, \{A_{\boldsymbol{x}}^a\})$ is a (symmetric, projective) strategy passing the consistency test with probability at least $1-\varepsilon$, and the linearity test in the i-th direction with probability at least $1-\sqrt{\varepsilon}$. Then there exists a family of measurements $\big\{B^\ell_{\boldsymbol{x}_{\neg i}}\big\}_{\ell \in \mathrm{ML}(\mathbb{F},\mathbb{F})}$ of arity 1 such that*

$$\mathrm{E}_{\boldsymbol{x}} \sum_a \Big\| A_{\boldsymbol{x}}^a - \sum_{\ell:\, \ell(x_i)=a} B^\ell_{\boldsymbol{x}_{\neg i}} \Big\|_\rho^2 \;=\; O\big(\sqrt{\varepsilon}\big). \tag{8}$$

We will often use the notation $B_{\boldsymbol{x}}^a := \sum_{\ell:\, \ell(x_i)=a} B^\ell_{\boldsymbol{x}_{\neg i}}$, leaving the dependence on i implicit. We note for future use that the bound (8) implies that

$$\mathrm{CON}(A,B) \;\geq\; 1 - O(\sqrt{\varepsilon}) \qquad \text{and} \qquad \mathrm{CON}(B) \;\geq\; 1-O(\sqrt{\varepsilon}).$$

These inequalities can be deduced directly from (8), but they will also be apparent from the proof of Claim 15, which we now give.

Proof. For any $\ell \in \mathrm{ML}(\mathbb{F},\mathbb{F})$, define

$$B^\ell_{\boldsymbol{x}_{\neg i}} \;:=\; \mathrm{E}_{x_i \neq x_i'}\, A^{\ell(x_i)}_{x_i,\boldsymbol{x}_{\neg i}} A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}} A^{\ell(x_i)}_{x_i,\boldsymbol{x}_{\neg i}}.$$

Then $\big\{B^\ell_{\boldsymbol{x}_{\neg i}}\big\}_\ell$ is a well-defined measurement: each operator is non-negative, and since for fixed $x_i \neq x_i'$, as ℓ ranges over $\mathrm{ML}(\mathbb{F}^2,\mathbb{F})$ both $\ell(x_i)$ and $\ell(x_i')$ independently range over $\mathbb{F}$, they sum to $\sum_a (A_{\boldsymbol{x}}^a)^2 = \mathrm{Id}$ since, by assumption, for every $\boldsymbol{x}$ and a the measurement operator $A_{\boldsymbol{x}}^a$ is a projector. Using the definition of $\|\cdot\|_\rho$, we can expand

$$\begin{aligned}\mathrm{E}_{\boldsymbol{x}} \sum_a \Big\| A_{\boldsymbol{x}}^a - \sum_{\ell:\ell(x_i)=a} B^\ell_{\boldsymbol{x}_{\neg i}} \Big\|_\rho^2 &= \mathrm{E}_{\boldsymbol{x}} \sum_a \mathrm{Tr}_\rho\big(A_{\boldsymbol{x}}^a\big) + \mathrm{E}_{\boldsymbol{x}} \sum_{\substack{\ell,\ell' \\ \ell(x_i)=\ell'(x_i)}} \mathrm{Tr}_\rho\big(B^\ell_{\boldsymbol{x}_{\neg i}} B^{\ell'}_{\boldsymbol{x}_{\neg i}}\big) \\ &\quad - 2\,\mathrm{E}_{\boldsymbol{x}} \sum_{a,\ell:\, \ell(x_i)=a} \mathrm{Tr}_\rho\big(A_{\boldsymbol{x}}^a B^\ell_{\boldsymbol{x}_{\neg i}}\big).\end{aligned} \tag{9}$$

We first lower bound the last term above. Applying Lemma 40 from Appendix B with $T_{\boldsymbol{x}}^h = A_{\boldsymbol{x}}^a$ and $Z_{\boldsymbol{x}}^h = B_{\boldsymbol{x}}^a$, we get

$$\Big| \mathrm{E}_{\boldsymbol{x}} \sum_{a,\ell:\, \ell(x_i)=a} \mathrm{Tr}_\rho\big(A_{\boldsymbol{x}}^a B^\ell_{\boldsymbol{x}_{\neg i}}\big) - \mathrm{E}_{\boldsymbol{x}} \sum_{a,\ell:\, \ell(x_i)=a} \mathrm{Tr}_\rho\big(A_{\boldsymbol{x}}^a \otimes B^\ell_{\boldsymbol{x}_{\neg i}}\big) \Big| \;=\; O\big(\mathrm{INC}(A)^{1/2}\big) \;=\; O\big(\sqrt{\varepsilon}\big) \tag{10}$$

by (6), hence it will suffice to show a lower bound on $\mathrm{E}_{\boldsymbol{x}} \sum_{a,\ell:\,\ell(x_i)=a} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes B^\ell_{\boldsymbol{x}_{\neg i}}\big)$. Using the definition of $B^\ell_{\boldsymbol{x}_{\neg i}}$, we have

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{x}} \sum_{a,\ell:\,\ell(x_i)=a} & \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes B^\ell_{\boldsymbol{x}_{\neg i}}\big) \\
&= \mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{a,\ell:\,\ell(x_i)=a} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}} A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}}\big) \\
&= \mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{a,\ell:\,\ell(x_i)=a} \sum_{a'} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}} A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}} \otimes A^{a'}_{x_i',\boldsymbol{x}_{\neg i}}\big) \\
&\leq \mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{a,\ell:\,\ell(x_i)=a} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}} A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}}\big) + \varepsilon \\
&\leq \mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{a,\ell:\ell(x_i)=a} \sum_{a'} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}}\big) + \varepsilon, \qquad (11)
\end{aligned}$$

where the first equality simply uses that the $A^{a'}_{x_i',\boldsymbol{x}_{\neg i}}$ sum to identity over a', the first inequality uses (6) on the last two registers (together with $A^a_{\boldsymbol{x}} \leq \mathrm{Id}$), and the last is by positivity. Let $\sigma := \rho^{(3)}$ be the reduced density of $|\Psi\rangle$ on any 3 of the provers, and apply Claim 37 to the POVM $\{A^a_{\boldsymbol{x}}\}_a$ for every $\boldsymbol{x}$. Eq. (6) implies that this POVM is consistent, hence

$$\mathrm{E}_{\boldsymbol{x}} \Big\| \sum_a \big(A^a_{\boldsymbol{x}} \otimes \mathrm{Id}\big)\, \rho^{(2)}\, \big(A^a_{\boldsymbol{x}} \otimes \mathrm{Id}\big) - \rho^{(2)} \Big\|_1 \;=\; O\big(\sqrt{\varepsilon}\big),$$

where we used that the $A^a_{\boldsymbol{x}}$ are projectors. Hence

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{a,\ell:\ell(x_i)=a} & \Big| \sum_{a'} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes \big(A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} - A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}}\big) \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}}\big) \Big| \\
&= \mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{a,\ell:\,\ell(x_i)=a} \Big| \mathrm{Tr}\Big(\big(A^a_{\boldsymbol{x}} \otimes A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}}\big) \cdot \\
&\qquad\qquad \Big(\sum_{a'} \big(\mathrm{Id} \otimes A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} \otimes \mathrm{Id}\big)\, \rho\, \big(\mathrm{Id} \otimes A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} \otimes \mathrm{Id}\big) - \rho \Big)\Big) \Big| \\
&\leq \mathrm{E}_{\boldsymbol{x},x_i'} \Big\| \sum_{a'} A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} \rho A^{a'}_{x_i',\boldsymbol{x}_{\neg i}} - \rho \Big\|_1 \;=\; O\big(\sqrt{\varepsilon}\big),
\end{aligned}$$

where for the inequality we used that for every $\boldsymbol{x}$ and $x_i \neq x_i''$, $\sum_{a,\ell:\ell(x_i)=a} A^a_{\boldsymbol{x}} \otimes A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}} \leq \mathrm{Id}$, and monotonicity of the trace distance. Combining this last bound with (11), we obtain

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{x}} \sum_{a,\ell:\,\ell(x_i)=a} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes B^\ell_{\boldsymbol{x}_{\neg i}}\big) &= \mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{a,\ell:\,\ell(x_i)=a} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}}\big) + O\big(\sqrt{\varepsilon}\big) \\
&= \mathrm{E}_{\boldsymbol{x},x_i'\neq x_i''} \sum_{\ell} \mathrm{Tr}_\rho\big(A^{\ell(x_i)}_{\boldsymbol{x}} \otimes A^{\ell(x_i'')}_{x_i'',\boldsymbol{x}_{\neg i}} \otimes A^{\ell(x_i')}_{x_i',\boldsymbol{x}_{\neg i}}\big) + O\big(\sqrt{\varepsilon}\big).
\end{aligned}$$

If $x_i = x_i'$ or $x_i = x_i''$, the last summation above evaluates to 1. Hence the expectation is at least as large as the probability that the $\{A^a_{\boldsymbol{x}}\}$ pass the linearity test along the i-th coordinate, which is at least $1 - \sqrt{\varepsilon}$ by (7), hence

$$\mathrm{E}_{\boldsymbol{x}} \sum_{a,\ell:\,\ell(x_i)=a} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes B^\ell_{\boldsymbol{x}_{\neg i}}\big) \;\geq 1 - O\big(\sqrt{\varepsilon}\big).$$

Combining this inequality with (10) and using that the first two terms in (9) are at most 1 each proves the claim. □

4.2 Proof of Theorem 11

In this section we show how Theorem 11, which is the result we need in order to analyze the overall protocol from Section 3, follows from Theorem 14. Theorem 14 is proved in Section 5.

Proof of Theorem 11. Let $\{V^g\}_{g\in \mathrm{ML}(\mathbb{F}^n,\mathbb{F})}$ be the sub-measurement guaranteed by Theorem 14. Expanding

$$\begin{aligned}
\mathrm{E}_x \sum_a \mathrm{Tr}_\rho\big((A_x^a - \sqrt{V_x^a})^2\big) &= \mathrm{E}_x \sum_a \Big(\mathrm{Tr}_\rho\big((A_x^a)^2\big) + \mathrm{Tr}_\rho\big(V_x^a\big) - 2\,\mathrm{Tr}_\rho\big(A_x^a\sqrt{V_x^a}\big)\Big)\\
&\leq 2 - 2\,\mathrm{E}_x \sum_a \mathrm{Tr}_\rho\big(A_x^a\sqrt{V_x^a}\big), \qquad (12)
\end{aligned}$$

it will suffice to show that this last expectation is close to 1. By applying Lemma 40 from Appendix B with $T_x^h = A_x^a$ and $Z_x^h = \sqrt{V_x^a}$ we obtain that

$$\Big|\mathrm{E}_x \sum_a \mathrm{Tr}_\rho\big(A_x^a\sqrt{V_x^a}\big) - \mathrm{E}_x \sum_a \mathrm{Tr}_\rho\big(A_x^a \otimes \sqrt{V_x^a}\big)\Big| \;=\; O\big(\textsc{inc}(A)^{1/2}\big) \;=\; O\big(\sqrt{\varepsilon}\big)$$

by (6). Hence to upper-bound the right-hand-side of (12) it suffices to lower-bound

$$\begin{aligned}
\mathrm{E}_x \sum_a \mathrm{Tr}_\rho\big(A_x^a \otimes \sqrt{V_x^a}\big) &\geq \mathrm{E}_x \sum_a \mathrm{Tr}_\rho\big(A_x^a \otimes V_x^a\big)\\
&\leq 1 - C_0\varepsilon^{c_0} - \mathrm{E}_x \sum_{g,a\neq g(x)} \mathrm{Tr}_\rho\big(V^g \otimes A_x^a\big)\\
&\geq 1 - 2C_0\varepsilon^{c_0},
\end{aligned}$$

where the second inequality uses item 2. from Theorem 14, and the last inequality follows from item 1. and the definition of $\textsc{inc}(V,A)$. Combined with (12), this proves Theorem 11. □

5 Soundness analysis of the multilinearity game

In this section we prove our main result on the analysis of the multilinearity game in the presence of entanglement between the provers, Theorem 14. The proof proceeds by induction, and the key inductive step is summed up in the following proposition. (We refer to section 2.1 for a definition of the quantities that appear in the proposition.)

Proposition 16. *There exists a universal constant $0 < c_1 < 1/2$ such that the following holds. Suppose that $(|\Psi\rangle, \{A_x^a\}_a)$ is a symmetric projective strategy for the players in the 3-player multilinearity game in n variables over $\mathbb{F}$ that is accepted with probability at least $1-\varepsilon$ in both the linearity test and the consistency test, for some $\varepsilon > 0$. Let $p := |\mathbb{F}|$ and $\delta > 0$, and assume that $n^{-8/c_1^2} \geq \delta \geq \sqrt{n}\varepsilon^{1/8} \geq np^{-1/4}$. Let $1 \leq k \leq n-1$ and T be a given family of sub-measurements of arity k such that $\textsc{inc}(T,A) \leq \delta$. Then there exists a family of sub-measurements V of arity $k+1$ such that*

1. $\textsc{inc}(V,A) = O(\varepsilon^{c_1})$,

2. *For any family of sub-measurements P of arity at least $k+1$,*

$$\big|\textsc{con}(P,V) - \textsc{con}(P,T)\big| \;=\; O\big(\delta^{c_1} + \textsc{inc}(P,A)^{1/2}\big),$$

3. *For any family of sub-measurements P, of arbitrary arity,*

$$\big|\textsc{con}(P,V) - \textsc{con}(P,T)\big| \;=\; O\big(\delta^{c_1} + \big|\textsc{con}(T,T) - \mathrm{Tr}_\rho(T)\big|^{1/2}\big).$$

We first show that Theorem 14 follows from Proposition 16.

Proof of Theorem 14. Starting from $V_0 = A$, let $V_1, \ldots, V_n$ be the sequence of measurements of increasing arity $1, \ldots, n$ given by Proposition 16. By item 1, for every $i \in [n]$ we have $\textsc{inc}(V_i, A) \leq C_1 \varepsilon^{c_1}$ for some universal constant C_1. Applying item 2 to $P = V_i$ and $V = V_i, V_{i-1}, \ldots, V_0$, an easy induction shows that

$$\big|\textsc{con}(V_i,V_i) - \textsc{con}(V_i,A)\big| \;=\; O\big(i\,\big(\varepsilon^{{c_1}^2} + \varepsilon^{c_1/2}\big)\big).$$

Hence using item 1. and $\textsc{inc}(V,A) + \textsc{con}(V,A) = \mathrm{Tr}_\rho(V)$, since A is a complete family of measurements, we also get

$$\big|\textsc{con}(V_i,V_i) - \mathrm{Tr}_\rho(V_i)\big| \;=\; O\big(i\,\varepsilon^{{c_1}^2}\big),$$

where we used $c_1 < 1/2$. Applying item 3 with $P = A$, an immediate induction then gives

$$\big|\textsc{con}(V_n,A) - \textsc{con}(A,A)\big| \;=\; O\big(n\sqrt{n}\,\varepsilon^{{c_1}^2/2}\big).$$

But $\textsc{con}(A,A) \geq 1 - \varepsilon$ by (6), and using $\mathrm{Tr}_\rho(V_n) = \textsc{con}(V_n,A) + \textsc{inc}(V_n,A)$ once more the theorem is proved for an appropriate choice of the constants c_0, C_0. □

The proof of Proposition 16 itself proceeds by induction, and is based on two lemmas. The first is a quantum analogue of the "self-improvement lemma" [BFL91, Lemma 5.10]. It shows that, if a family of sub-measurements $\{R^g_{\mathbf{x}_{\geq k}}\}$ is weakly consistent with $\{A^a_{\mathbf{x}}\}$, *and* it passes the consistency and linearity tests with high probability, then there exists an "improved" family of sub-measurements $\{T^g_{\mathbf{x}_{\geq k}}\}$ that are highly consistent with $\{A^a_{\mathbf{x}}\}$. (Item 3 in the conclusion of the lemma is not ultimately needed, but is required to combine Lemma 17 with Lemma 18 in the proof of Proposition 16.)

Lemma 17 (Self-improvement lemma)**.** *Let $(|\Psi\rangle, \{A^a_{\mathbf{x}}\}_a)$ be a (symmetric, projective) strategy for* 3 *players in the multilinearity game, and $n^{-8} \geq \delta \geq \sqrt{n}\varepsilon^{1/8} \geq 1/p$ such that the following hold:*

1. *The strategy $(|\Psi\rangle, \{A^a_{\mathbf{x}}\}_a)$ is accepted with probability at least $1 - \varepsilon/2$ in the multilinearity game,*

2. *There exists a family of sub-measurements R of arity k such that $\textsc{inc}(R,A) \leq \delta$.*

Then there exists a family of sub-measurements T of arity k, together with, for every $\mathbf{x} \in \mathbb{F}^n$, a family of matrices $\{\hat{S}^g_{\mathbf{x}}\}_g$, indexed by $g \in \mathrm{ML}(\mathbb{F}^{k-1}, \mathbb{F})$, such that the following hold:

1. $\textsc{inc}(T,A) = O(\varepsilon^{1/16})$,

2. *For any family of sub-measurements P, of arbitrary arity, $\big|\textsc{con}(P,R) - \textsc{con}(P,T)\big| = O\big(\sqrt{\delta}\big)$,*

3. *For every $\mathbf{x}$ and a, $\sum_{g: g(\mathbf{x}_{<k}) = a} \hat{S}^g_{\mathbf{x}} \big(\hat{S}^g_{\mathbf{x}}\big)^\dagger \leq A^a_{\mathbf{x}}$, and for every $\mathbf{x}_{\geq k}$ and g, $T^g_{\mathbf{x}_{\geq k}} = \big(\mathrm{E}_{\mathbf{x}_{<k}} \hat{S}^g_{\mathbf{x}}\big)\big(\mathrm{E}_{\mathbf{x}_{<k}} \hat{S}^g_{\mathbf{x}}\big)^\dagger$ and*

$$\mathrm{E}_{\mathbf{x}} \sum_g \Big\| \hat{S}^g_{\mathbf{x}} - \sqrt{T^g_{\mathbf{x}_{\geq k}}} \Big\|_\rho^2 \;\leq\; \delta.$$

The second lemma is an analogue of the "pasting lemma" [BFL91, Lemma 5.11]. It shows how, starting from a family of sub-measurements T of arity k that is consistent with A, one may construct a family of sub-measurements V of increased arity $k+1$ that is still somewhat consistent with A, as expressed in item 1 below. Items 2 and 3 are important to ensure that the new sub-measurement V is not "too incomplete", which would render item 1 trivial.

Lemma 18 (Pasting lemma)**.** *There exists a universal constant $0 < c_2 < 1$ such that the following holds. Let $\varepsilon, \delta > 0$ be such that $np^{-1} \leq \varepsilon \leq \delta^2$. Let $(|\Psi\rangle, \{A_x^a\}_a)$ be a (symmetric, projective) strategy for 3 players that is accepted with probability at least $1 - \varepsilon/2$ in the multilinearity game. Let $1 \leq k \leq n-1$ and T a family of sub-measurements of arity k such that $\textsc{inc}(T, A) \leq \delta$, and T satisfies item 3. in the conclusion of Lemma 17. Then there exists a family of sub-measurements V of arity $k+1$ such that*

1. *V is consistent with A: $\textsc{inc}(V, A) = O\big(\delta^{c_2}\big)$,*
2. *For any family of sub-measurements P of arity at least $k+1$,*
$$\big|\textsc{con}(P, V) - \textsc{con}(P, T)\big| \;=\; O(\delta^{c_2} + \textsc{inc}(P, A)^{1/2}),$$
3. *For any family of sub-measurements P, of arbitrary arity,*
$$\big|\textsc{con}(P, V) - \textsc{con}(P, T)\big| \;=\; O\big(\delta^{c_2} + \big|\textsc{con}(T, T) - \mathrm{Tr}_\rho(T)\big|^{1/2}\big).$$

Proposition 16 follows almost immediately by combining the two lemmas.

Proof of Proposition 16. Let T be the family of sub-measurements given in the statement of the proposition. First apply Lemma 18 to T, obtaining a family of sub-measurements R (called V in the lemma) of arity $k+1$ such that items 1, 2 and 3 in the conclusion of the lemma hold. Next apply Lemma 17 to R, obtaining a family of sub-measurements V of arity $k+1$ (called T in the lemma) such that items 1 and 2 hold, where given our assumption $\textsc{inc}(T, A) \leq \delta$ and item 1 from Lemma 18 the bound in item 2 is $O(\delta^{c_2/2})$. Item 1 from Lemma 17 implies item 1 in the proposition (provided c_1 is chosen small enough), and item 2 (resp. item 3) follows from combining item 2 from Lemma 17 with item 2 (resp. item 3) from Lemma 18. □

5.1 The self-improvement lemma

In this section we prove Lemma 17. Before proceeding with the details, we give some intuition and a high-level overview of how we will proceed.

Consider the following simplified situation in $n = 2$ dimensions. Although we will eventually require p to be a large power of 2, for the purposes of this overview it is sufficient to think about the case $p = 2$, so that the players' answers are simply bits. For every $x \in \mathbb{F}^2$ we are given a two-outcome projective measurement (A_x^0, A_x^1): picture two orthogonal "planes" of dimension $d/2$ each, where d is the dimension of either players' private space and can be arbitrarily large. Our goal is to find a global "refinement" of these planes: a single measurement $\{T^g\}$, with outcomes in the set of bilinear functions $g : \mathbb{F}^2 \to \mathbb{F}$, such that at every x the approximation $A_x^a \approx_\varepsilon \sum_{g:\, g(x)=a} T^g$ holds.[17] In order to achieve this, we make two additional assumptions:

1. There exists another measurement $\{R^g\}$ which achieves an approximation of weaker quality, up to some $\delta \gg \varepsilon$, than the one we are looking for,

[17] At this point we are being vague as to how the approximation is measured — it will eventually be expressed solely in terms of the consistency between the two measurements.

2. The $\{A_x^a\}$ are very close to *linear*: for every axis-parallel line $(x_1,\cdot)$ (resp. $(\cdot,x_2)$) there is a measurement $\{B_{x_1}^\ell\}_\ell$ (resp. $\{B_{x_2}^\ell\}_\ell$) with outcomes in the set of linear functions $\ell:\mathbb{F}\to\mathbb{F}$ such that $A_{(x_1,x_2)}^a \approx_\varepsilon \sum_{\ell:\,\ell(x_2)=a} B_{x_1}^\ell$ (resp. $A_{(x_1,x_2)}^a \approx_\varepsilon \sum_{\ell:\,\ell(x_1)=a} B_{x_2}^\ell$).

The goal is to use the high quality of the approximation along lines to improve the quality of the overall "bilinear" approximation. Let's trust that an ideal measurement $\{T^g\}$, achieving an approximation of order ε, exists, and think of $\{R^g\}$ as an adversarially "corrupted" version of $\{T^g\}$. There are two main ways in which $\{T^g\}$ can be corrupted: the first is by applying an arbitrary (but not too large) rotation on the whole space. The second is by "mislabeling" some of the measurement elements: e.g. for some g, a subspace of the space on which the ideal operator T^g projects could have been labeled as a subspace of $R^{g'}$ for some $g'\neq g$. Note that the first type of error is unique to the quantum setting, and did not arise in the setting of Babai et al.'s "self-improvement" lemma [BFL91]. Indeed, while quantum measurements are subject to arbitrarily small perturbations that may add up over time, nothing short of flipping the output of a binary function will suffice to corrupt it.

We devise a procedure which recovers from the first type of perturbation, but not the second. This appears unavoidable: if some components of the measurement $\{R^g\}$ are mis-labeled (say by completely re-shuffling the part of each measurement element that falls in a small-dimensional subspace of the whole space), there is no generic way to recover the corresponding ideal measurement elements. This is the main reason why the measurements we construct "shrink" at every step of the induction, and we have to work with sub-measurements instead: any "mislabeled" portions of space will have to be ignored. Since we cannot recover from such errors, it is crucial that they do not add up to too much throughout the whole induction process.

To correct the first type of error, we introduce the following procedure:

1. For every x, find the measurement $\{S_x^g\}_g$ which is closest to $\{R^g\}$ while being *perfectly* consistent with $\{A_x^a\}$: that is, $\sum_{g:g(x)=a} S_x^g = A_x^a$. This is possible only because the elements S_x^g are allowed to depend on x. We define the $\{S_x^g\}$ as the optimum solution to a specific convex program (see (13) below). Intuitively, S_x^g is obtained as the "projection" of R^g on the subspace $A_x^{g(x)}$.

2. Show that $\{S_x^g\}_g$ in fact only depends on x up to some error depending on ε only (and not δ), so that defining $T^g := \mathrm{E}_x S_x^g$ leads to the consistent measurement we are looking for.

The second step is crucial: why would the $\{S_x^g\}$ be (almost) independent of x? Here the linearity relations satisfied by the $\{A_x^a\}$ come into play. Using the perfect consistency of S and A, together with the linearity of A, we are able to conclude that the $\{S_x^g\}$ should not vary too much *along any axis-parallel line*. That is, $S_{(x_1,x_2)}^g \approx_\varepsilon S_{(x_1,x_2')}^g$ for any x_1 and x_2,x_2' (and similarly in the other direction). This step depends on the specific optimization problem that was introduced in order to define $\{S_x^g\}_g$ (see (13) below). This invariance along axis-parallel lines can then be combined with the (reasonably) good expansion properties of the hypercube to conclude that the $\{S_x^g\}$ are in fact globally invariant, leading to the "corrected" measurement $\{T^g\}$. (We note that the fact that invariance along axis-parallel lines implies global invariance was already used in [BFL91].)

We proceed with the details. In the following section we introduce the optimization procedure that is used to define the operators $\left\{S_x^g\right\}_g$. In Section 5.1.2 we show that the $\{S_x^g\}$ are close to being independent of x, leading to the definition of the family of sub-measurements $\{T_{x_{\geq k}}^g\}$. In Section 5.1.3 we show that T satisfies the conclusions of Lemma 17.

5.1.1 A convex optimization problem

Let $\{R^g_{x_{\geq k}}\}_g$ be the family of sub-measurements promised in the assumptions of Lemma 17. Let $\{\hat{S}^g_x\}_g$, where $x \in \mathbb{F}^n$ and $g \in \mathrm{ML}(\mathbb{F}^{k-1}, \mathbb{F})$, be an optimal solution to the following convex optimization problem:

Convex program for self-improvement

$$\omega := \min \mathrm{E}_x \sum_g \Big\| \hat{S}^g_x - \sqrt{R^g_{x_{\geq k}}} \Big\|^2_\rho \tag{13}$$

$$\forall x, a, \sum_{g: g(x_{<k}) = a} \hat{S}^g_x (\hat{S}^g_x)^\dagger \leq A^a_x,$$

where $\sqrt{R^g_{x_{\geq k}}}$ is the positive square root of $R^g_{x_{\geq k}}$. Let $S^g_x := \hat{S}^g_x(\hat{S}^g_x)^\dagger$.[18] Our first claim shows that the optimum of (13) is bounded as a function of the inconsistency of R and A.

Claim 19. *Suppose that the* $\{R^g_{x_{\geq k}}\}_g$ *satisfy the assumptions of Lemma 17. Then the optimum* ω *of* (13) *is at most* $\textsc{inc}(A, R) + O(\sqrt{\varepsilon})$.

Proof. We construct a feasible solution achieving the claimed value. Let $\hat{S}^g_x := A^{g(x_{<k})}_x \sqrt{R^g_{x_{\geq k}}}$. Then by definition $\{\hat{S}^g_x\}$ is a feasible solution to (13). To upper-bound its value, we first evaluate

$$\begin{aligned}
\mathrm{E}_x \sum_g \Big(\mathrm{Tr}_\rho\big(\hat{S}^g_x \sqrt{R^g_{x_{\geq k}}}\big) - \mathrm{Tr}_\rho\big(R^g_{x_{\geq k}}\big) \Big) &= \mathrm{E}_x \sum_g \mathrm{Tr}_\rho\big(\big(A^{g(x_{<k})}_x - \mathrm{Id}\big) R^g_{x_{\geq k}}\big) \\
&= \mathrm{E}_x \sum_a \mathrm{Tr}_\rho\Big(A^a_x \Big(\sum_{g: g(x) \neq a} R^g_{x_{\geq k}} \Big)\Big) \\
&= \mathrm{E}_x \sum_g \mathrm{Tr}_\rho\big(R^g_{x_{\geq k}} \otimes A^{g(x_{<k})}_x\big) + O\big(\textsc{inc}(A)^{1/2}\big),
\end{aligned}$$

where the second equality uses that $\sum_a A^a_x = \mathrm{Id}$ for every x, and the last follows from an application of Lemma 40. A similar calculation shows that

$$\mathrm{E}_x \sum_g \mathrm{Tr}_\rho\big(\hat{S}^g_x (\hat{S}^g_x)^\dagger\big) = \mathrm{E}_x \sum_g \mathrm{Tr}_\rho\big(R^g_{x_{\geq k}} \otimes A^{g(x_{<k})}_x\big) + O\big(\textsc{inc}(A)^{1/2}\big).$$

To conclude, expand $\big\| \hat{S}^g_x - \sqrt{R^g_{x_{\geq k}}} \big\|^2_\rho$ and use

$$\mathrm{E}_x \sum_g \mathrm{Tr}_\rho\big(R^g_{x_{\geq k}} \otimes A^{g(x_{<k})}_x\big) = \mathrm{Tr}_\rho(R) - \textsc{inc}(A, R)$$

by definition, together with the bound $\textsc{inc}(A) \leq \varepsilon$ from (6). □

[18] We will usually use a hat, as in $\hat{S}$, to denote matrices which we think of as factorizations of positive semidefinite matrices, but are not necessarily positive themselves. In general, the relation between $\hat{X}$ and X will always be that $X = \hat{X}\hat{X}^\dagger$.

5.1.2 Constructing a family of sub-measurements independent of $\boldsymbol{x}_{<k}$

As a first step in showing that any optimal solution to (13) must be close to one that does not depend on $\boldsymbol{x}_{<k}$, we show that such an optimal solution must be close to another feasible solution which is furthermore close to being invariant along the direction of any axis-parallel line in direction $i < k$. Precisely, we have the following.

Claim 20. *Assume $p^{-1} \leq \varepsilon$. For every $i < k$ there exists a feasible solution $\big\{\hat{Z}_{\boldsymbol{x}}^g\big\}_g$ to (13), with objective value at most $\omega + O\big(\varepsilon^{1/4}\big)$, such that*

$$\mathrm{E}_{\boldsymbol{x}} \sum_g \big\| \hat{Z}_{\boldsymbol{x}}^g - \mathrm{E}_{x_i'} \hat{Z}_{\boldsymbol{x}_{\neg i}, x_i'}^g \big\|_\rho^2 \;=\; O\big(\sqrt{\varepsilon}\big).$$

Proof. Let $\{\hat{S}_{\boldsymbol{x}}^g\}$ be an optimal solution to (13), and for any $i < k$ let

$$\hat{Y}_{\boldsymbol{x}_{\neg i}}^g \;:=\; B_{\boldsymbol{x}_{\neg i}}^{g|\ell_i(\boldsymbol{x})} \big(\mathrm{E}_{x_i} \hat{S}_{\boldsymbol{x}}^g\big),$$

where $\ell_i(\boldsymbol{x})$ is the line going through $\boldsymbol{x}$ and parallel to the i-th axis, and $\{B_{\boldsymbol{x}_{\neg i}}^\ell\}_\ell$ is the "lines" family of measurements introduced in Claim 15. We first claim that the $\hat{Y}_{\boldsymbol{x}_{\neg i}}^g$, while not strictly feasible, achieve an objective value in (13) of at most $\omega + O(\varepsilon^{1/4})$.

Towards proving this, we first show that $B_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})} \hat{S}_{\boldsymbol{x}}^g$ is close to $\hat{S}_{\boldsymbol{x}}^g$. Recall the definition of $B_{\boldsymbol{x}}^a = \sum_{\ell:\, \ell(x_i) = a} B_{\boldsymbol{x}_{\neg i}}^\ell$. Using the fact that, since $\{S_{\boldsymbol{x}}^g\}$ is feasible, $A_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})} \hat{S}_{\boldsymbol{x}}^g = \hat{S}_{\boldsymbol{x}}^g$, we get

$$\begin{aligned} \mathrm{E}_{\boldsymbol{x}} \sum_g \big\| B_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})} \hat{S}_{\boldsymbol{x}}^g - \hat{S}_{\boldsymbol{x}}^g \big\|_\rho^2 &= \mathrm{E}_{\boldsymbol{x}} \sum_g \mathrm{Tr}_\rho\Big(\big(B_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})} - A_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})}\big) S_{\boldsymbol{x}}^g \big(B_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})} - A_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})}\big)\Big) \\ &\leq \mathrm{E}_{\boldsymbol{x}} \sum_a \big\| B_{\boldsymbol{x}}^a - A_{\boldsymbol{x}}^a \big\|_\rho^2 \\ &= O\big(\sqrt{\varepsilon}\big) \end{aligned} \tag{14}$$

by Claim 15. Using the triangle inequality and convexity, the following (not necessarily feasible) operators

$$\tilde{Y}_{\boldsymbol{x}_{\neg i}}^g \;:=\; \mathrm{E}_{x_i} B_{\boldsymbol{x}}^{g(\boldsymbol{x}_{\leq k})} \hat{S}_{\boldsymbol{x}}^g$$

also achieve a value $\omega + O(\sqrt{\varepsilon})$ in (13).

Next we show that the $\tilde{Y}_{\boldsymbol{x}_{\neg i}}^g$ are close to the $\hat{Y}_{\boldsymbol{x}_{\neg i}}^g := B_{\boldsymbol{x}_{\neg i}}^{g|\ell_i(\boldsymbol{x})} \mathrm{E}_{x_i} \hat{S}_{\boldsymbol{x}}^g$. From the definition,

$$\tilde{Y}_{\boldsymbol{x}_{\neg i}}^g \;=\; B_{\boldsymbol{x}_{\neg i}}^{g|\ell_i(\boldsymbol{x})} \big(\mathrm{E}_{x_i} \hat{S}_{\boldsymbol{x}}^g\big) + \mathrm{E}_{x_i} \sum_{\substack{\ell:\, \ell(x_i) = g(\boldsymbol{x}_{\leq k}) \\ \ell \neq g_{|\ell_i(\boldsymbol{x})}}} B_{\boldsymbol{x}_{\neg i}}^\ell \hat{S}_{\boldsymbol{x}}^g.$$

The norm of the second term can be expanded as follows:

$$\begin{aligned} &\mathrm{E}_{\boldsymbol{x}_{\neg i}} \sum_g \Big\| \mathrm{E}_{x_i} \sum_{\substack{\ell:\, \ell(x_i) = g(\boldsymbol{x}_{\leq k}) \\ \ell \neq g_{|\ell_i(\boldsymbol{x})}}} B_{\boldsymbol{x}_{\neg i}}^\ell \hat{S}_{\boldsymbol{x}}^g \Big\|_\rho^2 \\ &\qquad = \mathrm{E}_{\boldsymbol{x}_{\neg i}} \sum_g \mathrm{E}_{x_i, y_i} \sum_{\substack{\ell:\, \ell(x_i) = g(\boldsymbol{x}_{\leq k}) \\ \ell \neq g_{|\ell_i(\boldsymbol{x})}}} \;\sum_{\substack{\ell':\, \ell'(y_i) = g(\boldsymbol{x}_{\leq k}) \\ \ell' \neq g_{|\ell_i(\boldsymbol{x})}}} \mathrm{Tr}_\rho\big(B_{\boldsymbol{x}_{\neg i}}^\ell \hat{S}_{\boldsymbol{x}_{\neg i}, x_i}^g (\hat{S}_{\boldsymbol{x}_{\neg i}, y_i}^g)^\dagger B_{\boldsymbol{x}_{\neg i}}^{\ell'}\big). \end{aligned}$$

Eq. (29) from Lemma 40 shows that the contribution of all terms such that $\ell \neq \ell'$ is at most $O\big(\sqrt{\textsc{inc}(B)}\big) = O\big(\varepsilon^{1/4}\big)$ by Claim 15. But the only possibility for $\ell = \ell'$ is that also $x_i = y_i$, since two distinct linear functions on $\mathbb{F}$ intersect in at most one point. Hence we have that

$$\begin{aligned}\mathrm{E}_{\boldsymbol{x}_{\neg i}} \sum_g \Big\| \mathrm{E}_{x_i} \sum_{\substack{\ell\colon \ell(x_i)=g(\boldsymbol{x}_{\leq k})\\ \ell\neq g_{|\ell_i(\boldsymbol{x})}}} B^{\ell}_{\boldsymbol{x}_{\neg i}} \hat{S}^g_{\boldsymbol{x}} \Big\|_\rho^2 &= \mathrm{E}_{\boldsymbol{x}_{\neg i}} \sum_g \frac{1}{p} \mathrm{E}_{x_i} \sum_{\substack{\ell\colon \ell(x_i)=g(\boldsymbol{x}_{\leq k})\\ \ell\neq g_{|\ell_i(\boldsymbol{x})}}} \mathrm{Tr}_\rho\big(B^{\ell}_{\boldsymbol{x}_{\neg i}} S^g_{\boldsymbol{x}} B^{\ell}_{\boldsymbol{x}_{\neg i}}\big) + O\big(\varepsilon^{1/4}\big)\\ &\leq \frac{4}{p} + O\big(\varepsilon^{1/4}\big).\end{aligned}$$

Given our assumption on p, this implies

$$\mathrm{E}_{\boldsymbol{x}_{\neg i}} \sum_g \big\| \tilde{Y}^g_{\boldsymbol{x}_{\neg i}} - \hat{Y}^g_{\boldsymbol{x}_{\neg i}} \big\|_\rho^2 \,=\, O\big(\varepsilon^{1/4}\big),$$

and hence the $\hat{Y}^g_{\boldsymbol{x}_{\neg i}}$, while still not necessarily feasible, achieve an objective value in (13) of $\omega + O\big(\varepsilon^{1/4}\big)$.

Finally, define $\hat{Z}^g_{\boldsymbol{x}} := A^{g(\boldsymbol{x}_{\leq k})}_{\boldsymbol{x}} B^{g_{|\ell_i(\boldsymbol{x})}}_{\boldsymbol{x}_{\neg i}} \big(\mathrm{E}_{x_i} \hat{S}^g_{\boldsymbol{x}}\big)$. Then the $\big\{\hat{Z}^g_{\boldsymbol{x}}\big\}$ are feasible in (13), and the fact that

$$\mathrm{E}_{\boldsymbol{x}} \sum_g \big\| \hat{Z}^g_{\boldsymbol{x}} - \hat{Y}^g_{\boldsymbol{x}_{\neg i}} \big\|_\rho^2 \,=\, O\big(\sqrt{\varepsilon}\big) \tag{15}$$

follows from arguments similar to those used in the proof of Claim 19. Hence the $\big\{\hat{Z}^g_{\boldsymbol{x}}\big\}$ are a feasible solution to (13) with objective value at most $\omega + O\big(\varepsilon^{1/4}\big)$. Finally, by convexity (15) implies that

$$\mathrm{E}_{\boldsymbol{x}_{\neg i}} \sum_g \big\| \mathrm{E}_{x_i} \hat{Z}^g_{\boldsymbol{x}} - \hat{Y}^g_{\boldsymbol{x}_{\neg i}} \big\|_\rho^2 \,=\, O\big(\sqrt{\varepsilon}\big),$$

which together with the triangle inequality and (15) shows that the $\{\hat{Z}^g_{\boldsymbol{x}}\}$ are close to their expectation on any axis-parallel line in the i-th direction, proving the claim. □

Using convexity of $X \to \|X - A\|_\rho^2$ for fixed A, the following follows from Claims 19 and 20.

Claim 21. *Let* $\big\{\hat{S}^g_{\boldsymbol{x}}\big\}$ *be an optimal solution to* (13). *Then*

$$\mathrm{E}_{\boldsymbol{x}, i<k} \sum_g \|\hat{S}^g_{\boldsymbol{x}} - \mathrm{E}_{x'_i} \hat{S}^g_{\boldsymbol{x}_{\neg i} x'_i}\|_\rho^2 = O\big(\varepsilon^{1/4}\big).$$

Proof. We show that the two solutions constructed to (13), $\big\{\hat{S}^g_{\boldsymbol{x}}\big\}$ and $\big\{\hat{Z}^g_{\boldsymbol{x}}\big\}$ from Claim 20, must be close:[19]

$$\mathrm{E}_{\boldsymbol{x}, i<k} \sum_g \big\| \hat{Z}^g_{\boldsymbol{x}} - \hat{S}^g_{\boldsymbol{x}} \big\|_\rho^2 \,=\, O\big(\varepsilon^{1/4}\big). \tag{16}$$

The claim then follows by using the triangle inequality to combine this bound with the fact, proved in Claim 20, that the $\hat{Z}^g_{\boldsymbol{x}}$ themselves are close to their expectation along any axis-parallel line in the i-th direction. Hence it suffices to prove (16). Since the feasible set of (13) is convex, for any

[19] Note that $\hat{Z}^g_{\boldsymbol{x}}$ implicitly depends on i, and the following equation is measuring the distance on average over the $k-1$ different constructions of $\hat{Z}^g_{\boldsymbol{x}}$ obtained for all $1 \leq i < k$.

$0 \le t \le 1$ the elements $\{(1-t)\hat{S}^g_{\boldsymbol{x}} + t\hat{Z}^g_{\boldsymbol{x}}\}$ also constitute a feasible solution. By optimality of $\{\hat{S}^g_{\boldsymbol{x}}\}$, the resulting objective value must be at least ω: for every $0 \le t \le 1$,

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{x}} \sum_g \Big\| \hat{S}^g_{\boldsymbol{x}} - \sqrt{R^g_{\boldsymbol{x}_{\ge k}}} \Big\|_\rho^2 &\le \mathrm{E}_{\boldsymbol{x}} \sum_g \Big\| (1-t)\hat{S}^g_{\boldsymbol{x}} + t\hat{Z}^g_{\boldsymbol{x}} - \sqrt{R^g_{\boldsymbol{x}_{\ge k}}} \Big\|_\rho^2 \\
&= t^2\, \mathrm{E}_{\boldsymbol{x}} \sum_g \Big\| \hat{Z}^g_{\boldsymbol{x}} - \hat{S}^g_{\boldsymbol{x}} \Big\|_\rho^2 + \mathrm{E}_{\boldsymbol{x}} \sum_g \Big\| \hat{S}^g_{\boldsymbol{x}} - \sqrt{R^g_{\boldsymbol{x}_{\ge k}}} \Big\|_\rho^2 \\
&\qquad + 2\,t\, \mathrm{E}_{\boldsymbol{x}} \sum_g \mathrm{Tr}_\rho\Big(\big(\hat{Z}^g_{\boldsymbol{x}} - \hat{S}^g_{\boldsymbol{x}}\big)\big(\hat{S}^g_{\boldsymbol{x}} - \sqrt{R^g_{\boldsymbol{x}_{\ge k}}}\big)^\dagger \Big).
\end{aligned}$$

Using the known objective values, re-arranging and making $t \to 0$, we obtain that

$$\mathrm{E}_{\boldsymbol{x}} \sum_g \mathrm{Tr}_\rho\Big(\big(\hat{Z}^g_{\boldsymbol{x}} - \hat{S}^g_{\boldsymbol{x}}\big)\big(\sqrt{R^g_{\boldsymbol{x}_{\ge k}}} - \hat{S}^g_{\boldsymbol{x}}\big)^\dagger \Big) = O\big(\varepsilon^{1/4}\big).$$

Hence

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{x}} \sum_g \big\| \hat{S}^g_{\boldsymbol{x}} - \hat{Z}^g_{\boldsymbol{x}} \big\|_\rho^2 &= \mathrm{E}_{\boldsymbol{x}} \sum_g \Big(\Big\| \hat{Z}^g_{\boldsymbol{x}} - \sqrt{R^g_{\boldsymbol{x}_{\ge k}}} \Big\|_\rho^2 - \Big\| \hat{S}^g_{\boldsymbol{x}} - \sqrt{R^g_{\boldsymbol{x}_{\ge k}}} \Big\|_\rho^2 \\
&\qquad + 2\,\mathrm{Tr}_\rho\Big(\big(\hat{Z}^g_{\boldsymbol{x}} - \hat{S}^g_{\boldsymbol{x}}\big)\big(\sqrt{R^g_{\boldsymbol{x}_{\ge k}}} - \hat{S}^g_{\boldsymbol{x}}\big)^\dagger \Big) \Big) \\
&= O\big(\varepsilon^{1/4}\big),
\end{aligned}$$

proving (16). □

Claim 21 shows that the $\{\hat{S}^g_{\boldsymbol{x}}\}_g$ do not vary much along any axis-parallel line in the i-th direction. Using the expansion properties of the hypercube, we can deduce that the $\{\hat{S}^g_{\boldsymbol{x}}\}_g$ are close (in the squared $\|\cdot\|_\rho$ norm) to a single operator, independent of the first $(k-1)$ coordinates.

Claim 22. *For every $\boldsymbol{x}_{\ge k}$ and g, let $\hat{T}^g_{\boldsymbol{x}_{\ge k}} := \mathrm{E}_{\boldsymbol{x}_{<k}} \hat{S}^g_{\boldsymbol{x}}$. Then*

$$\mathrm{E}_{\boldsymbol{x}} \sum_g \big\| \hat{S}^g_{\boldsymbol{x}} - \hat{T}^g_{\boldsymbol{x}_{\ge k}} \big\|_\rho^2 \;=\; O\big(n\varepsilon^{1/4}\big).$$

Proof. This is a direct consequence of the expansion properties of the hypercube, as expressed in Claim 38. □

5.1.3 Proof of Lemma 18

We conclude the proof of Lemma 18 by showing that the non-negative operators

$$T^g_{\boldsymbol{x}_{\ge k}} \;:=\; \hat{T}^g_{\boldsymbol{x}_{\ge k}} \big(\hat{T}^g_{\boldsymbol{x}_{\ge k}}\big)^\dagger,$$

where for any $\boldsymbol{x}_{\ge k}$ and g the matrix $\hat{T}^g_{\boldsymbol{x}_{\ge k}}$ is defined in Claim 22 in the previous section, satisfy the conclusions of the lemma. First note that item 3 follows directly from Claim 22, so it will suffice to verify that items 1 and 2 hold. Regarding item 1, we can bound

$$\begin{aligned}\textsc{inc}(T,A) &= \mathrm{E}_{\boldsymbol{x}} \sum_{g,a\neq g(\boldsymbol{x}_{<k})} \mathrm{Tr}_\rho\big(T^g_{\boldsymbol{x}_{\geq k}} \otimes A^a_{\boldsymbol{x}}\big)\\ &= \mathrm{E}_{\boldsymbol{x}} \sum_{g,a\neq g(\boldsymbol{x}_{<k})} \mathrm{Tr}_\rho\big(S^g_{\boldsymbol{x}} \otimes A^a_{\boldsymbol{x}}\big) + O\big(\sqrt{n}\varepsilon^{1/8}\big)\\ &\leq \mathrm{E}_{\boldsymbol{x}} \sum_{a\neq b} \mathrm{Tr}_\rho\big(A^a_{\boldsymbol{x}} \otimes A^b_{\boldsymbol{x}}\big) + O\big(\sqrt{n}\varepsilon^{1/8}\big)\\ &= O\big(\sqrt{n}\varepsilon^{1/8}\big),\end{aligned}$$

where the second equality follows from Cauchy-Schwarz and Claim 22, the inequality follows from the fact that the $\hat{S}^g_{\boldsymbol{x}}$ are a feasible solution to (13), and the last uses self-consistency of A as in (6).

Item 2 is proved in a similar way. Let P be a family of sub-measurements of arity ℓ, and assume that $\ell \leq k$, the other case being treated symmetrically. By definition,

$$\begin{aligned}\big|\textsc{con}(P,T) - \textsc{con}(P,R)\big| &= \mathrm{E}_{\boldsymbol{x}} \sum_{f,g:\, g_{|x_{\ell\cdots k-1}}=f} \mathrm{Tr}_\rho\big(P^f_{\boldsymbol{x}_{\geq l}} \otimes \big(T^g_{\boldsymbol{x}_{\geq k}} - R^g_{\boldsymbol{x}_{\geq k}}\big)\big)\\ &\leq \Big(\mathrm{E}_{\boldsymbol{x}} \sum_{f,g:\, g_{|x_{\ell\cdots k-1}}=f} \mathrm{Tr}_\rho\big(P^f_{\boldsymbol{x}_{\geq l}} \otimes \big(\hat{T}^g_{\boldsymbol{x}_{\geq k}} - \sqrt{R^g_{\boldsymbol{x}_{\geq k}}}\big)\big(\hat{T}^g_{\boldsymbol{x}_{\geq k}} - \sqrt{R^g_{\boldsymbol{x}_{\geq k}}}\big)^\dagger\big)\Big)^{1/2}\\ &\qquad \cdot \Big(\mathrm{E}_{\boldsymbol{x}} \sum_{f,g:\, g_{|x_{\ell\cdots k-1}}=f} \mathrm{Tr}_\rho\big(P^f_{\boldsymbol{x}_{\geq l}} \otimes \big(\hat{T}^g_{\boldsymbol{x}_{\geq k}} + \sqrt{R^g_{\boldsymbol{x}_{\geq k}}}\big)\big(\hat{T}^g_{\boldsymbol{x}_{\geq k}} + \sqrt{R^g_{\boldsymbol{x}_{\geq k}}}\big)^\dagger\big)\Big)^{1/2}\\ &\leq \sqrt{2}\Big(\mathrm{E}_{\boldsymbol{x}} \sum_g \big\|\hat{T}^g_{\boldsymbol{x}_{\geq k}} - \sqrt{R^g_{\boldsymbol{x}_{\geq k}}}\big\|^2_\rho\Big)^{1/2}\\ &= O\big(\textsc{inc}(R,A)^{1/2} + \sqrt{n}\varepsilon^{1/8}\big),\end{aligned}$$

where the first inequality is by Cauchy-Schwarz, the second uses that $\sum_f P^f_{\boldsymbol{x}_{\geq l}} \leq \mathrm{Id}$ for every $\boldsymbol{x}_{\geq l}$, and the last follows from the bounds proved in Claim 19 and Claim 22.

5.2 The pasting lemma

In this section we prove Lemma 18. Let T be the family of sub-measurements whose existence is promised in the lemma's assumptions. For every $\boldsymbol{x}$, let $\big\{\hat{S}^h_{\boldsymbol{x}}\big\}_h$ and $\big\{T^h_{\boldsymbol{x}_{\geq k}}\big\}_h$ be as in item 3 of Lemma 17. Let δ be such that

$$\max\Big\{\textsc{inc}(T,B),\, \textsc{inc}(T,A),\, \mathrm{E}_{\boldsymbol{x}} \sum_h \Big\|\hat{S}^h_{\boldsymbol{x}} - \sqrt{T^h_{\boldsymbol{x}_{\geq k}}}\Big\|^2_\rho\Big\} \;\leq\; \delta, \tag{17}$$

where here $\{B^\ell_{\boldsymbol{x}_{\neg k}}\}_\ell$ are the "lines" measurements in the k-th direction, as defined in Claim 15. Note that Claim 15 implies that $\textsc{inc}(T,B) \leq \textsc{inc}(T,A) + O(\varepsilon^{1/4})$, which justifies including $\textsc{inc}(T,B)$ in (17).

Our goal is to define a new family of sub-measurements V, depending on one less coordinate of $\boldsymbol{x}$ than T, but such that V is still consistent with A, and moreover V is not "too small", as measured by items 2 and 3 in the lemma. The main idea is to define $\{V^g_{\boldsymbol{x}_{>k}}\}$ as (roughly) corresponding to the sequential application of $\{T^h_{\boldsymbol{x}_{\geq k}}\}$ twice, for two random choices of x_k. This will produce two $(k-1)$-multilinear functions h and h', from which a k-multilinear function g can be recovered by interpolation. This is essentially the same method as was used to define the "line" operators B from the "point" operators A in Claim 15. Here the main additional difficulty is that we are starting with a family of sub-measurements, instead of complete, projective measurements as was the case in Claim 15.

This section is organized as follows. We start with some preliminary observations in Section 5.2.1. The family of sub-measurements V is defined in Section 5.2.2. Item 1 in the conclusion of Lemma 18 is proved in Section 5.2.3, and items 2 and 3 are proved in Section 5.2.4.

5.2.1 Pre-processing

In this section we prove a preliminary claim, Claim 23 below, which lets us modify the family of sub-measurements T into another family Q that has useful properties. The important property is item 3. in the claim, which establishes a form of commutation between Q and the "line" measurements B. Intuitively, that such a property would hold for Q equal to T should follow from the consistency between the families of sub-measurements defined by T and B: consistent measurements are "compatible", and by the gentle measurement lemma (cf. Lemma 35) the order in which they are performed does not matter. However, we could not show directly that item 3 below holds for the family of sub-measurements T itself; hence we need to modify it slightly.

Claim 23. *Let T be the family of sub-measurements satisfying the assumptions of Lemma 18, and δ be as in* (17). *There exists a family of sub-measurements $\{Q^h_{\boldsymbol{x}_{\geq k}}\}$ such that the following hold:*

1. $\mathrm{Tr}_\rho(Q) \geq \mathrm{Tr}_\rho(T) - O(\delta^{c_4})$,
2. *For every $\boldsymbol{x}_{\geq k}$ and h, $Q^h_{\boldsymbol{x}_{\geq k}} = B^h_{\boldsymbol{x}_{\geq k}} \tilde{Q}^h_{\boldsymbol{x}_{\geq k}} B^h_{\boldsymbol{x}_{\geq k}}$ for some family of sub-measurements $\{\tilde{Q}^h_{\boldsymbol{x}_{\geq k}}\}$ (and in particular $\textsc{inc}(Q, A) = O(\varepsilon^{1/2})$),*
3. *Let $Q_{\boldsymbol{x}_{>k}} = \mathrm{E}_{x_k} \sum_h Q^h_{\boldsymbol{x}_{\geq k}}$. For any $r \geq 1$,*

$$\mathrm{E}_{\boldsymbol{x}} \mathrm{Tr}_\rho\Big(\Big((Q_{\boldsymbol{x}_{>k}})^r - \sum_\ell B^\ell_{\boldsymbol{x}_{\neg k}} (Q_{\boldsymbol{x}_{>k}})^r B^\ell_{\boldsymbol{x}_{\neg k}}\Big)^2\Big) \;=\; O\big(r^2 \delta^{c_4}\big),$$

where $c_4 > 0$ is a universal constant.

Proof. For any $\boldsymbol{x}_{<k}$ define a "pinching" map

$$\mathcal{E}_{\boldsymbol{x}_{<k}} : \; T^h_{\boldsymbol{x}_{\geq k}} \;\mapsto\; B^{h(\boldsymbol{x}_{<k})}_{\boldsymbol{x}} T^h_{\boldsymbol{x}_{\geq k}} B^{h(\boldsymbol{x}_{<k})}_{\boldsymbol{x}}.$$

Note that $\mathcal{E}_{\boldsymbol{x}_{<k}}$ also implicitly depends on $h(\boldsymbol{x}_{<k})$, but this dependence will always be clear from the context. Let $\mathcal{E}(\cdot) := \mathrm{E}_{\boldsymbol{x}_{<k}} \mathcal{E}_{\boldsymbol{x}_{<k}}(\cdot)$. The idea for the definition of Q consists in applying the map $\mathcal{E}$ to T a certain number of times, leveraging a certain stability property that will follow after sufficiently many applications.

Let M be an integer to be fixed later, and for every $\boldsymbol{x}_{\geq k}$ and h let $R^h_{\boldsymbol{x}_{\geq k}} := \mathcal{E}^M(T^h_{\boldsymbol{x}_{\geq k}})$, where $\mathcal{E}^M$ denotes the sequential composition of $\mathcal{E}$ with itself M times. Using the Schwarz-Zippel lemma (Lemma 33) it is not hard to verify that, as long as $M \geq 1$, $\textsc{inc}(R, B) = O(\textsc{inc}(B, B) + n/p) = O(\varepsilon^{1/2})$. The proof of Claim 23 is based on the following sequence of facts.

Fact 24. *There is a choice of $M \leq \delta^{-1/4}$ for which the following holds:*

$$\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho\big(\big(R^h_{\boldsymbol{x}_{\geq k}} - \mathcal{E}_{\boldsymbol{x}_{<k}}(R^h_{\boldsymbol{x}_{\geq k}})\big)^2\big) \;=\; O(\delta^{1/4}).$$

Proof. Let $J_1 = \delta^{-1/4}$. The proof is based on the use of the potential function

$$\Phi_i \;:=\; \mathrm{E}_{\boldsymbol{x}_{\geq k}} \sum_h \mathrm{Tr}_\rho\big(\mathcal{E}^{J_1-i}\big((\mathcal{E}^i(T^h_{\boldsymbol{x}_{\geq k}}))^2\big)\big),$$

defined for all $0 \leq i \leq J_1$. Note that Φ_i is non-negative, always at most 1, and by the pinching inequality $(\mathcal{E}(X))^2 \leq \mathcal{E}(X^2)$ for any positive semidefinite X, Φ_i is non-increasing with i. Let i_1 the smallest index i for which it holds that

$$\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho\Big(\mathcal{E}^{J_1-i}\Big(\mathcal{E}_{\boldsymbol{x}_{<k}}\big((\mathcal{E}^{i-1}(T^h_{\boldsymbol{x}_{\geq k}}))^2\big) - \big(\mathcal{E}_{\boldsymbol{x}_{<k}}(\mathcal{E}^{i-1}(T^h_{\boldsymbol{x}_{\geq k}}))\big)^2\Big)\Big) \;\leq\; \delta^{1/4}. \tag{18}$$

Using operator convexity of the square function, this inequality not being satisfied for some i implies that $\Phi_{i-1} - \Phi_i > \delta^{1/4}$. Since this can happen for at most $\delta^{-1/4}$ indices i, an $0 \leq i_1 \leq \delta^{1/4}$ such that (18) is satisfied for $i = i_1$ must exist. Using self-consistency of B $(J_1 - i_1)$ times, and consistency of T and B, (18) is seen to imply

$$\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho\big(\big(\mathcal{E}^{i_1-1}(T^h_{\boldsymbol{x}_{\geq k}}) - \mathcal{E}_{\boldsymbol{x}_{<k}}\big(\mathcal{E}^{i_1-1}(T^h_{\boldsymbol{x}_{\geq k}})\big)\big)^2\big) \;\leq\; O\big(\delta^{-1/4}\textsc{inc}(B)^{1/2} + \delta^{1/4}\big).$$

To conclude, we set $M := i_1 - 1$ and use $\textsc{inc}(B) = O(\varepsilon^{1/2}) \leq \delta^{1/2}$. □

The following is a consequence of Fact 24.

Fact 25. *The following holds*

$$\mathrm{E}_{\boldsymbol{x}_{\neg k}, x_k \neq y_k} \sum_{g,\ell:\, \ell \neq g|_{\boldsymbol{x}_{<k}}} \mathrm{Tr}_\rho\big(B^\ell_{\boldsymbol{x}_{\neg k}} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} R^{g|y_k}_{y_k \boldsymbol{x}_{>k}} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} B^\ell_{\boldsymbol{x}_{\neg k}}\big) \;=\; O\big(\textsc{inc}(R,B)^{1/2} + \delta^{1/8}\big).$$

Proof. By definition of R,

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{x}_{>k}, x_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} R^{g|y_k}_{y_k \boldsymbol{x}_{>k}} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}}\big)\\
&\qquad = \mathrm{E}_{\boldsymbol{x}_{\neg k}, x_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} B^{g(\boldsymbol{x}_{<k} y_k)}_{\boldsymbol{x}_{\neg k} y_k} R^{g|y_k}_{y_k \boldsymbol{x}_{>k}} B^{g(\boldsymbol{x}_{<k} y_k)}_{\boldsymbol{x}_{\neg k} y_k} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}}\big)\\
&\qquad = \mathrm{E}_{\boldsymbol{x}_{\neg k}, x_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(\big(B^{g(\boldsymbol{x}_{\leq k})}_{\boldsymbol{x}_{\neg k} x_k} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} B^{g(\boldsymbol{x}_{\leq k})}_{\boldsymbol{x}_{\neg k} x_k}\big) B^{g(\boldsymbol{x}_{<k} y_k)}_{\boldsymbol{x}_{\neg k} y_k} R^{g|y_k}_{y_k \boldsymbol{x}_{>k}} B^{g(\boldsymbol{x}_{<k} y_k)}_{\boldsymbol{x}_{\neg k} y_k} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}}\big) + O\big(\delta^{1/8}\big),
\end{aligned}$$

where the second equality follows from Fact 24. Using that, by definition, for $x_k \neq y_k$, $B^{g(\boldsymbol{x}_{<k} x_k)}_{\boldsymbol{x}_{\neg k} x_k} B^{g(\boldsymbol{x}_{<k} y_k)}_{\boldsymbol{x}_{\neg k} y_k} = B^{g|\boldsymbol{x}_{<k}}_{\boldsymbol{x}_{\neg k}}$ and consistency of R and B, we get

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{x}_{>k}, x_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} R^{g|y_k}_{y_k \boldsymbol{x}_{>k}} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}}\big)\\
&\qquad = \mathrm{E}_{\boldsymbol{x}_{\neg k}, x_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} B^{g|\boldsymbol{x}_{<k}}_{\boldsymbol{x}_{\neg k}} R^{g|y_k}_{y_k \boldsymbol{x}_{>k}} B^{g|\boldsymbol{x}_{<k}}_{\boldsymbol{x}_{\neg k}} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}}\big) + O\big(\textsc{inc}(R,B)^{1/2} + \delta^{1/8}\big)\\
&\qquad = \mathrm{E}_{\boldsymbol{x}_{\neg k}, x_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(B^{g|\boldsymbol{x}_{<k}}_{\boldsymbol{x}_{\neg k}} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} R^{g|y_k}_{y_k \boldsymbol{x}_{>k}} R^{g|x_k}_{x_k \boldsymbol{x}_{>k}} B^{g|\boldsymbol{x}_{<k}}_{\boldsymbol{x}_{\neg k}}\big) + O\big(\textsc{inc}(R,B)^{1/2} + \delta^{1/8}\big),
\end{aligned}$$

where the last equality again follows (after a little work) from Fact 24. □

We will also use the following.

Fact 26. *Let $\{S^g_{\mathbf{x}_{>k}}\}_g$ be an arbitrary family of sub-measurements and $\mu_2 > 0$. There exists an $i_2 \leq \mu_2^{-1}$ such that*

$$\mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\Big(\big(\mathcal{E}_{\mathbf{x}_{>k}}\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big) - \mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big)^2\Big) \;=\; O\big(\mu_2 + \mu_2^{-1}\,\mathrm{INC}(B)^{1/2}\big),$$

where here we denote $\mathcal{E}(S^g_{\mathbf{x}_{>k}}) = \mathrm{E}_{\mathbf{x}_{<k}} B^{g|y_{<k}}_{\mathbf{x}_{\neg k}} S^g_{\mathbf{x}_{>k}} B^{g|y_{<k}}_{\mathbf{x}_{\neg k}}$. Moreover, for all $i \geq i_2$ it holds that

$$\mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\big(\big(\mathcal{E}^{i+1}(S^g_{\mathbf{x}_{>k}}) - \mathcal{E}^{i}(S^g_{\mathbf{x}_{>k}})\big)^2\big) \;=\; O\big(\mu_2 + i\,\mathrm{INC}(B)^{1/2}\big).$$

Proof. The proof is very similar to that of Fact 24, and is based on the use of the potential function

$$\Phi_i \;:=\; \mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\big(\mathcal{E}^{J_2-i}\big((\mathcal{E}^i(S^g_{\mathbf{x}_{>k}}))^2\big)\big),$$

defined for all $0 \leq i \leq J_2$, where $J_2 = \mu_2^{-1}$. Note that Φ_i is always at most 1, and by the pinching inequality $\mathcal{E}(X)^2 \leq \mathcal{E}(X^2)$ for any positive semidefinite X, Φ_i is non-increasing with i. Let i_2 the smallest index such that $\Phi_{i_2-1} - \Phi_{i_2} \leq \mu_2$; as long as $J_2 \geq \mu_2^{-1}$ such an $0 \leq i_2 \leq J_2$ must exist. By definition, it then holds that

$$\mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\Big(\mathcal{E}^{J_2-i_2}\Big(\mathcal{E}_{\mathbf{x}_{<k}}\big(\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big)^2\big) - \big(\mathcal{E}_{\mathbf{x}_{<k}}\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big)\big)^2\Big)\Big) \;\leq\; \mu_2.$$

Using self-consistency of B $(J_2 - i_2)$ times, we obtain

$$\mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\Big(\big(\mathcal{E}_{\mathbf{x}_{<k}}\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big) - \mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big)^2\Big) \;\leq\; \mu_2 + O(J_2\,\mathrm{INC}(B)^{1/2}).$$

To conclude the proof, it suffices to use the operator convexity of the square function to move the expectation over $\mathbf{x}_{<k}$ inside the square, and then observe that

$$\begin{aligned}
&\mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\big(\big(\mathcal{E}\big(\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big) - \mathcal{E}^2\big(\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big)\big)\big)^2\big) \\
&\qquad\qquad \leq \mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\big(\mathcal{E}\big(\big(\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big) - \mathcal{E}\big(\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big)\big)\big)^2\big) \\
&\qquad\qquad \leq \mathrm{E}_{\mathbf{x}} \sum_g \mathrm{Tr}_\rho\big(\big(\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}}) - \mathcal{E}\big(\big(\mathcal{E}^{i_2-1}(S^g_{\mathbf{x}_{>k}})\big)\big)\big)^2\big) + O\big(\mathrm{INC}(B)^{1/2}\big),
\end{aligned}$$

again using self-consistency of B. □

Let M' be an integer to be fixed later, and for every $\mathbf{x}_{\geq k}$ and h define

$$Q^h_{\mathbf{x}_{\geq k}} := \big(\mathrm{E}_{\mathbf{x}_{<k}} B^{h(\mathbf{x}_{<k})}_{\mathbf{x}}\big)^{M'} \big(R^h_{\mathbf{x}_{\geq k}}\big) \big(\mathrm{E}_{\mathbf{x}_{<k}} B^{h(\mathbf{x}_{<k})}_{\mathbf{x}}\big)^{M'}.$$

Observe that, as before, as long as $M' \geq 1$ it holds that $\mathrm{INC}(Q, B) = O(\mathrm{INC}(B, B) + n/p) = O(\varepsilon^{1/2})$. For any $g \in \mathrm{ML}(\mathbb{F}^k, \mathbb{F})$, let $Q^g_{\mathbf{x}_{>k}} := \mathrm{E}_{x_k \neq y_k} Q^{g|x_k}_{x_k \mathbf{x}_{\geq k}} Q^{g|y_k}_{y_k \mathbf{x}_{\geq k}}$. The following implies item 3. in Claim 23: for any $r \geq 1$,

$$\mathrm{E}_{\mathbf{x}} \mathrm{Tr}_\rho\Big(\Big((Q_{\mathbf{x}_{>k}})^r - \sum_g \big(B^{g|x_{<k}}_{\mathbf{x}_{\neg k}} Q^g_{\mathbf{x}_{>k}} B^{g|x_{<k}}_{\mathbf{x}_{\neg k}}\big)^r\Big)^2\Big) \;=\; O\big(r^2 \delta^{c_3}\big), \tag{19}$$

where $c_3 > 0$ is a universal constant. Eq. (19) is proved by induction on r. The case $r = 1$ is stated in the following claim.

Fact 27. *The following holds*

$$\mathrm{E}_{\boldsymbol{x}}\mathrm{Tr}_\rho\Big(\Big(Q_{\boldsymbol{x}_{>k}} - \sum_g B_{\boldsymbol{x}_{\neg k}}^{g|x_{<k}} Q_{\boldsymbol{x}_{>k}}^g B_{\boldsymbol{x}_{\neg k}}^{g|x_{<k}}\Big)^2\Big) \;=\; O\big(\delta^{1/16} + M'\delta^{1/8}\big). \tag{20}$$

Proof. Fact 25 implies that $\{Q_{\boldsymbol{x}_{>k}}^g\}$ and B are $O(\delta^{1/8})$-consistent, from which it follows that

$$\begin{aligned}\mathrm{E}_{\boldsymbol{x}}\sum_{g,g'}\mathrm{Tr}_\rho\big(B_{\boldsymbol{x}_{\neg k}}^{g|x_{<k}} Q_{\boldsymbol{x}_{>k}}^g B_{\boldsymbol{x}_{\neg k}}^{g|x_{<k}} B_{\boldsymbol{x}_{\neg k}}^{g'|x_{<k}} Q_{\boldsymbol{x}_{>k}}^{g'} B_{\boldsymbol{x}_{\neg k}}^{g'|x_{<k}}\big) &= \mathrm{E}_{\boldsymbol{x}}\sum_{g,g'}\mathrm{Tr}_\rho\big(Q_{\boldsymbol{x}_{>k}}^g B_{\boldsymbol{x}_{\neg k}}^{g|x_{<k}} B_{\boldsymbol{x}_{\neg k}}^{g'|x_{<k}} Q_{\boldsymbol{x}_{>k}}^{g'}\big) + O(\delta^{1/8})\\ &= \mathrm{E}_{\boldsymbol{x}}\sum_{g,g'}\mathrm{Tr}_\rho\big(Q_{\boldsymbol{x}_{>k}}^g Q_{\boldsymbol{x}_{>k}}^{g'} \otimes B_{\boldsymbol{x}_{\neg k}}^{g|x_{<k}} \otimes B_{\boldsymbol{x}_{\neg k}}^{g'|x_{<k}}\big) + O(\delta^{1/16} + M'\delta^{1/8})\\ &= \mathrm{E}_{\boldsymbol{x}}\mathrm{Tr}_\rho\big((Q_{\boldsymbol{x}_{>k}})^2\big) + O(\delta^{1/16} + M'\delta^{1/8}).\end{aligned}$$

Here the second equality follows by applying Fact 26 with $\mu_2 := \delta^{1/16}$ and $S_{\boldsymbol{x}}^g$ chosen as $Q_{\boldsymbol{x}_{>k}}^g$ to move the term $B_{\boldsymbol{x}_{\neg k}}^{g|x_{<k}}$ on the outside, and holds as long as $M' \geq \mu_2^{-1} = \delta^{-1/16}$. (The third uses consistency of Q and B.) Expanding out the square in (20), all four terms can be related up to $O(M'\delta^{1/8})$ by using similar arguments. □

The induction step required to prove Eq. (19) uses arguments similar to that of the proof of Fact 27, and we leave the details to the reader. Once that equation is established, choosing $M' = \delta^{-1/16}$ item 3 in Claim 23 follows. Items 1 and 2 in the claim are simple consequences of the definition of Q from R, and of R from T; again we omit the details. □

5.2.2 Construction of the pasted family of sub-measurements

In this section and for the remainder of the proof of Lemma 18 we rename the family of sub-measurements $\{Q_{\boldsymbol{x}_{\geq k}}^h\}$ constructed in the previous section into $\{T_{\boldsymbol{x}_{\geq k}}^h\}$. The only properties of that family that we will need are those stated in Claim 23. In order to define the pasted sub-measurements V, we first introduce a "pseudo-inverse" $\tilde{T}$ as follows. As usual, let $T_{\boldsymbol{x}_{>k}} = \mathrm{E}_{x_k}\sum_h T_{\boldsymbol{x}_{\geq k}}^h$ and $\eta > 0$ a small parameter to be fixed later. Define

$$\tilde{T}_{\boldsymbol{x}_{>k}} \;:=\; \Big(\sum_{r=0}^{R}\big(\mathrm{Id} - T_{\boldsymbol{x}_{>k}}\big)^r\Big)^{1/2}, \tag{21}$$

where $R := (10/\eta)\log(1/\eta)$ is chosen so that $T_{\boldsymbol{x}_{>k}}(1 - T_{\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}}^2) \leq \eta\,\mathrm{Id}$ (note that, by definition, $\tilde{T}_{\boldsymbol{x}_{>k}}$ commutes with $T_{\boldsymbol{x}_{>k}}$). Expanding out the series in the definition of $\tilde{T}_{\boldsymbol{x}_{>k}}$, Item 3 from Claim 23 implies that the following equation holds:

$$\mathrm{E}_{\boldsymbol{x}}\mathrm{Tr}_\rho\big((\tilde{T}_{\boldsymbol{x}_{>k}} - \sum_\ell B_{\boldsymbol{x}_{\neg k}}^\ell \tilde{T}_{\boldsymbol{x}_{>k}} B_{\boldsymbol{x}_{\neg k}}^\ell)^2\big) \;=\; O\big((\delta/\eta)^{c_5}\big), \tag{22}$$

where $c_5 > 0$ is a sufficiently small constant. For every $\boldsymbol{x}_{>k}$ and $g \in \mathrm{ML}(\mathbb{F}^k, \mathbb{F})$, define

$$V_{\boldsymbol{x}_{>k}}^g \;:=\; \Big(1 + \frac{R}{p}\Big)^{-1}\mathrm{E}_{y_k}\tilde{T}_{\boldsymbol{x}_{>k}}\Big(\mathrm{E}_{x_k \neq y_k} T_{x_k\boldsymbol{x}_{>k}}^{g|x_k}\Big)\tilde{T}_{\boldsymbol{x}_{>k}} T_{y_k\boldsymbol{x}_{>k}}^{g|y_k}\tilde{T}_{\boldsymbol{x}_{>k}}\Big(\mathrm{E}_{x_k \neq y_k} T_{x_k\boldsymbol{x}_{>k}}^{g|x_k}\Big)\tilde{T}_{\boldsymbol{x}_{>k}}.$$

The scaling factor $(1 + R/p)^{-1}$ is necessary to ensure that the $\{V_{\boldsymbol{x}_{>k}}^g\}$ sum to at most identity. It induces an extra error term in all our estimates; however our choice of $\eta = \delta^{c'}$ for some $c' > 0$ will ensure that this error term is of the same order as ones that already appear; for clarity in the remainder of this section we will neglect it.

Claim 28. *The $\big\{V^g_{\boldsymbol{x}_{>k}}\big\}_g$ form a family of sub-measurements of arity $k+1$.*

Proof. It is clear that $V^g_{\boldsymbol{x}_{>k}} \geq 0$ for every g. When the variable g runs over $\mathrm{ML}(\mathbb{F}^k,\mathbb{F})$, for $x_k \neq y_k \in \mathbb{F}$ the restrictions $g_{|x_k}$ and $g_{|y_k}$ independently run over $\mathrm{ML}(\mathbb{F}^{k-1},\mathbb{F})$. Hence, using convexity of the map $A \mapsto AXA^\dagger$ for any A and $X \geq 0$,

$$\begin{aligned}
\sum_g V^g_{\boldsymbol{x}_{>k}} &\leq \Big(1+\frac{R}{p}\Big)^{-1}\frac{1}{p^2}\sum_{y_k\neq x_k}\sum_{h,h'}\tilde{T}_{\boldsymbol{x}_{>k}}T^h_{x_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}}T^{h'}_{y_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}}T^h_{x_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}} \\
&= \Big(1+\frac{R}{p}\Big)^{-1}\frac{1}{p}\sum_{x_k}\sum_h \tilde{T}_{\boldsymbol{x}_{>k}}T^h_{x_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}}T_{\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}}T^h_{x_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}} \\
&\quad - \Big(1+\frac{R}{p}\Big)^{-1}\frac{1}{p^2}\sum_{x_k}\sum_h \tilde{T}_{\boldsymbol{x}_{>k}}T^h_{x_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}}T_{x_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}}T^h_{x_k\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}} \\
&\leq \Big(1+\frac{R}{p}\Big)^{-1}\Big(\mathrm{Id}+\frac{R}{p}\,\mathrm{Id}\Big) \;\leq\; \mathrm{Id},
\end{aligned}$$

where to obtain the last line we used $(T^h_{x_k\boldsymbol{x}_{>k}})^2 \leq T^h_{x_k\boldsymbol{x}_{>k}}$ as well as $\tilde{T}_{\boldsymbol{x}_{>k}} \leq R^{1/2}\,\mathrm{Id}$ and $\tilde{T}_{\boldsymbol{x}_{>k}}T_{\boldsymbol{x}_{>k}}\tilde{T}_{\boldsymbol{x}_{>k}} \leq \mathrm{Id}$. □

5.2.3 Consistency

In this section we show that the "pasted" sub-measurement V is consistent with A, proving item 1 of Lemma 18. It will be convenient to introduce the shorthand

$$W^h_{\boldsymbol{x}_{\geq k}} \;:=\; \tilde{T}_{\boldsymbol{x}_{>k}}T^h_{\boldsymbol{x}_{\geq k}}\tilde{T}_{\boldsymbol{x}_{>k}}. \tag{23}$$

We also let $\delta_W := \max(\delta, \textsc{inc}(W,A))$.

Claim 29. *The following holds*

$$\textsc{inc}(W,A) \;=\; \mathrm{E}_{\boldsymbol{x}}\sum_{h,a\neq h(\boldsymbol{x}_{<k})}\mathrm{Tr}_\rho\big(\tilde{T}_{\boldsymbol{x}_{>k}}T^h_{\boldsymbol{x}_{\geq k}}\tilde{T}_{\boldsymbol{x}_{>k}}\otimes A^a_{\boldsymbol{x}}\big) \;=\; O\big((\delta/\eta)^{c_5}\big),$$

where $c_5 > 0$ is the constant that appears in (22)*.*

Proof. We have

$$\begin{aligned}
\textsc{inc}(W,A) &= \mathrm{E}_{\boldsymbol{x}_{\geq k}}\sum_h \mathrm{Tr}_\rho\big(\tilde{T}_{\boldsymbol{x}_{>k}}T^h_{\boldsymbol{x}_{\geq k}}\tilde{T}_{\boldsymbol{x}_{>k}}\otimes(\mathrm{Id}-A^h_{\boldsymbol{x}_{\geq k}})\big) \\
&= \mathrm{E}_{\boldsymbol{x}}\sum_{h,a}\mathrm{Tr}_\rho\big(\tilde{T}_{\boldsymbol{x}_{>k}}B^h_{\boldsymbol{x}_{\geq k}}R^h_{\boldsymbol{x}_{\geq k}}B^h_{\boldsymbol{x}_{\geq k}}\tilde{T}_{\boldsymbol{x}_{>k}}\otimes(\mathrm{Id}-A^h_{\boldsymbol{x}_{\geq k}})\big) \\
&= \mathrm{E}_{\boldsymbol{x}}\sum_h \mathrm{Tr}_\rho\big(B^h_{\boldsymbol{x}_{\geq k}}\tilde{T}_{\boldsymbol{x}_{>k}}R^h_{\boldsymbol{x}_{\geq k}}\tilde{T}_{\boldsymbol{x}_{>k}}B^h_{\boldsymbol{x}_{\geq k}}\otimes(\mathrm{Id}-A^h_{\boldsymbol{x}_{\geq k}})\big) + O\big((\delta/\eta)^{c_5}\big) \\
&= \mathrm{E}_{\boldsymbol{x}}\sum_h \mathrm{Tr}_\rho\big(\tilde{T}_{\boldsymbol{x}_{>k}}R^h_{\boldsymbol{x}_{\geq k}}\tilde{T}_{\boldsymbol{x}_{>k}}\otimes(\mathrm{Id}-A^h_{\boldsymbol{x}_{\geq k}})\otimes A^h_{\boldsymbol{x}_{\geq k}}\big) + O\big((\delta/\eta)^{c_5}\big) \\
&= O\big((\delta/\eta)^{c_5}\big),
\end{aligned}$$

where the second equality follows from item 2 in Claim 23 (and some sub-measurement $\{R^h_{\boldsymbol{x}_{\geq k}}\}$), the third follows from (22), the fourth uses Lemma 40 together with consistency of B and A as in Claim 15, and the last again follows from self-consistency of A, together with $\tilde{T}_{\boldsymbol{x}_{>k}} \leq R^{1/2}\,\mathrm{Id} \leq \eta^{-1}\,\mathrm{Id}$ for small enough η. □

Claim 30. *The family of sub-measurements V is consistent with A:*

$$\textsc{inc}(V,A) \;=\; O\big(\delta_W^{1/2}\big).$$

Proof. By definition,

$$\begin{aligned}
\textsc{inc}(V,A) &= \mathrm{E}_{\boldsymbol{x},x_k',x_k''\neq y_k} \sum_{g,a\neq g(\boldsymbol{x}_{\leq k})} \mathrm{Tr}_\rho\big(W^{g_{|x_k'}}_{x_k'\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k''}}_{x_k''\boldsymbol{x}_{>k}} \otimes A^a_{\boldsymbol{x}}\big)\\
&= \mathrm{E}_{\boldsymbol{x},x_k',x_k''\neq y_k} \sum_{g,a\neq g(\boldsymbol{x}_{\leq k})} \mathrm{Tr}_\rho\big(W^{g_{|x_k'}}_{x_k'\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k''}}_{x_k''\boldsymbol{x}_{>k}} \otimes A^a_{\boldsymbol{x}} \otimes A^{g(\boldsymbol{x}_{<k},x_k')}_{\boldsymbol{x}_{\neg k}x_k'}\big) + O(\delta_W^{1/2})\\
&= \mathrm{E}_{\boldsymbol{x},x_k',x_k''\neq y_k} \sum_{g,a\neq g(\boldsymbol{x}_{\leq k})} \mathrm{Tr}_\rho\big(W^{g_{|x_k'}}_{x_k'\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k''}}_{x_k''\boldsymbol{x}_{>k}} A^{g(\boldsymbol{x}_{<k},x_k'')}_{\boldsymbol{x}_{\neg k}x_k''} \otimes A^a_{\boldsymbol{x}} \otimes A^{g(\boldsymbol{x}_{<k},x_k')}_{\boldsymbol{x}_{\neg k}x_k'}\big) + O(\delta_W^{1/2})\\
&= O(\varepsilon^{1/2}+\delta_W^{1/2}),
\end{aligned}$$

where the second and third equalities each follow from an application of Lemma 39 and the definition of δ_W, and the last follows by applying Cauchy-Schwarz and using linearity of A in the k-th direction, as in (7). □

5.2.4 Consistency with arbitrary sub-measurements

We now show that items 2 and 3 in the conclusion of Lemma 18 hold. We will make use of the bound

$$\mathrm{E}_{\boldsymbol{x}_{>k}}\mathrm{Tr}_\rho\big(\big(\mathrm{Id}-W_{\boldsymbol{x}_{>k}}\big)T_{\boldsymbol{x}_{>k}}\big(\mathrm{Id}-W_{\boldsymbol{x}_{>k}}\big)\big) \;=\; O(\eta), \tag{24}$$

which holds by definition of $\{W^h_{\boldsymbol{x}_{\geq k}}\}$ (cf. (23)) and of $\tilde{T}_{\boldsymbol{x}_{>k}}$ (cf. (21)).

Claim 31. *For any family of sub-measurements P of arity at least $k+1$,*

$$\big|\textsc{con}(P,V)-\textsc{con}(P,T)\big| \;=\; O\big(\delta_W^{1/2}+\textsc{inc}(P,A)^{1/2}+\eta^{1/2}\big).$$

Proof. Let P be an arbitrary family of sub-measurements of arity $\ell \geq k+1$. We prove the claim in case $\ell = k+1$, the other cases being exactly similar. Then $P = \{P^g_{\boldsymbol{x}_{>k}}\}_{g\in\mathrm{ML}(\mathbb{F}^k,\mathbb{F})}$, and by definition

$$\begin{aligned}
\textsc{con}(P,V) &= \mathrm{E}_{\boldsymbol{x}}\sum_g \mathrm{Tr}_\rho\big(P^g_{\boldsymbol{x}}\otimes V^g_{\boldsymbol{x}}\big)\\
&= \mathrm{E}_{\boldsymbol{x}_{\neg k},x_k,x_k'\neq y_k}\sum_g \mathrm{Tr}_\rho\big(P^g_{\boldsymbol{x}_{>k}}\otimes W^{g_{|x_k}}_{x_k\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k'}}_{x_k'\boldsymbol{x}_{>k}}\big)\\
&= \mathrm{E}_{\boldsymbol{x},x_k',x_k''\neq y_k}\sum_g \mathrm{Tr}_\rho\big(P^g_{\boldsymbol{x}_{>k}}\otimes W^{g_{|x_k'}}_{x_k'\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k''}}_{x_k''\boldsymbol{x}_{>k}} \otimes A^{g(\boldsymbol{x}_{<k}x_k')}_{\boldsymbol{x}_{\neg k}x_k'}\big) + O\big(\delta_W^{1/2}\big),
\end{aligned} \tag{25}$$

where the last equality follows from Lemma 39 and the definition of δ_W. We can then write

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{x},x_k',x_k''\neq y_k}\sum_{g,h\neq g_{|x_k'}} \mathrm{Tr}_\rho\big(P^g_{\boldsymbol{x}_{>k}}\otimes W^h_{x_k'\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k''}}_{x_k''\boldsymbol{x}_{>k}} \otimes A^{g(\boldsymbol{x}_{<k}x_k')}_{\boldsymbol{x}_{\neg k}x_k'}\big)\\
&= \mathrm{E}_{\boldsymbol{x},x_k',x_k''\neq y_k}\sum_{g,h\neq g_{|x_k'}} \mathrm{Tr}_\rho\big(A^{h(\boldsymbol{x}_{<k}x_k')}_{\boldsymbol{x}_{\neg k}x_k'} P^g_{\boldsymbol{x}_{>k}} A^{h(\boldsymbol{x}_{<k}x_k')}_{\boldsymbol{x}_{\neg k}x_k'}\otimes W^h_{x_k'\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k''}}_{x_k''\boldsymbol{x}_{>k}} \otimes A^{g(\boldsymbol{x}_{<k}x_k')}_{\boldsymbol{x}_{\neg k}x_k'}\big) + O\big(\delta_W^{1/2}\big)\\
&= \mathrm{E}_{\boldsymbol{x},x_k',x_k''\neq y_k}\sum_{\substack{h\neq g_{|x_k'}\\ h(\boldsymbol{x}_{<k})=g(\boldsymbol{x}_{<k}x_k')}} \mathrm{Tr}_\rho\big(P^g_{\boldsymbol{x}_{>k}}\otimes W^h_{x_k'\boldsymbol{x}_{>k}} T^{g_{|y_k}}_{y_k\boldsymbol{x}_{>k}} W^{g_{|x_k''}}_{x_k''\boldsymbol{x}_{>k}} \otimes A^{g(\boldsymbol{x}_{<k}x_k')}_{\boldsymbol{x}_{\neg k}x_k'}\big) + O\big(\varepsilon^{1/2}+\delta_W^{1/2}\big),
\end{aligned} \tag{26}$$

where the first equality again uses Lemma 39 and the definition of δ_W, and the last follows from an application of the Cauchy-Schwarz inequality and self-consistency of A. In the last expression, h and $g_{|x'_k}$ are two distinct $(k-1)$-linear functions over $\mathbb{F}$: by the Schwartz-Zippel lemma (see Lemma 33 for a statement) they intersect in a fraction at most $O(k/|\mathbb{F}|) = O(k/p)$ points. Hence, applying the Cauchy-Schwarz inequality to recover a non-negative expression, we can upper bound (26) by $O(\sqrt{n/p} + \varepsilon^{1/2} + \delta_W^{1/2}) = O(\delta_W^{1/2})$ since $\delta_W \geq \delta \geq np^{-1}$. Together with (25), this shows that

$$\begin{aligned}\mathrm{CON}(P,V) &= \mathrm{E}_{\boldsymbol{x},x'_k,x''_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(P^g_{\boldsymbol{x}_{>k}} \otimes W_{x'_k\boldsymbol{x}_{>k}} T^{g|y_k}_{y_k\boldsymbol{x}_{>k}} W^{g|x''_k}_{x''_k\boldsymbol{x}_{>k}} \otimes A^{g(\boldsymbol{x}_{<k}x'_k)}_{\boldsymbol{x}_{\neg k}x'_k}\big) + O\big(\delta_W^{1/2}\big)\\ &= \mathrm{E}_{\boldsymbol{x},x''_k \neq y_k} \sum_g \mathrm{Tr}_\rho\big(P^g_{\boldsymbol{x}_{>k}} \otimes T^{g|y_k}_{y_k\boldsymbol{x}_{>k}} W^{g|x''_k}_{x''_k\boldsymbol{x}_{>k}} \otimes A^{g(\boldsymbol{x}_{<k}x'_k)}_{\boldsymbol{x}_{\neg k}x'_k}\big) + O\big(\delta_W^{1/2} + \eta^{1/2}\big),\end{aligned}$$

where the second equality follows from the Cauchy-Schwarz inequality and (24). Repeating the same steps for the remaining term $W^{g|x''_k}_{x''_k\boldsymbol{x}_{>k}}$, and using consistency of P and A to conclude, proves the claim. □

Claim 32. *For any sub-measurement P, of arbitrary arity,*

$$\big|\mathrm{CON}(P,V) - \mathrm{CON}(P,T)\big| \;=\; O\big(\big|\mathrm{CON}(T,T) - \mathrm{Tr}_\rho(T)\big|^{1/2} + \delta_W^{1/2} + \eta^{1/2}\big).$$

Proof. The proof closely follows that of Claim 31, and we omit the details. □

This concludes the proof of Lemma 18 provided c_2 is chosen to be a sufficiently small constant.

A Auxiliary lemmas

We first recall a key lemma in the analysis of low-degree polynomials over a finite field, the Schwartz-Zippel lemma [Sch80, Zip79], which we state in a form that will be useful to us.

Lemma 33 (Schwartz-Zippel). *Let $\mathbb{F}$ be a finite field, n an integer, and $f : \mathbb{F}^n \to \mathbb{F}$ a non-zero multilinear function. Then f has at most $sn|\mathbb{F}|^{n-1}$ zeros.*

The next series of claims are all based on variants of the Cauchy-Schwarz inequality. The first follows from Eq. (3) of Bhatia and Davis [BD95] (see also [Bha88]), substituting the norm $|||\cdot|||$ by $\|\cdot\|_1$.

Theorem 34. *Let A and B be arbitrary matrices such that the product $A^\dagger B$ is well-defined. Then,*

$$\big\|A^\dagger B\big\|_1 \leq \big\|A\big\|_F \big\|B\big\|_F.$$

Winter's gentle measurement lemma [Win99, Lemma 9] (see also Aaronson's "almost as good as new" lemma [Aar05, Lemma 2.2]) is a key lemma formalizing the intuitive fact that if a measurement produces a certain outcome with near-certainty when performed on a specific state, then the post-measurement state is close to the original state. The following is a variant of that lemma, and we give a proof following Ogawa and Nagaoka [ON07, Appendix C].

Lemma 35. *Let ρ be a density operator on a Hilbert space $\mathcal{H}$, and X and Y be linear operators from $\mathcal{H}$ to a Hilbert space $\mathcal{K}$ such that $X^*X \preceq I$ and $Y^*Y \preceq I$. Then,*

$$\|X\rho X^* - Y\rho Y^*\|_1 \leq 2\sqrt{\mathrm{Tr}(X-Y)\rho(X-Y)^*}.$$

Proof. By the triangle inequality,

$$\|X\rho X^* - Y\rho Y^*\|_1 \leq \|(X-Y)\rho X^*\|_1 + \|Y\rho(X-Y)^*\|_1.$$

By Theorem 34,

$$\begin{aligned}\|(X-Y)\rho X^*\|_1 &\leq \|(X-Y)\sqrt{\rho}\|_2 \|\sqrt{\rho}\, X^*\|_2\\ &= \sqrt{\mathrm{Tr}(X-Y)\rho(X-Y)^*}\sqrt{\mathrm{Tr} X\rho X^*}\\ &\leq \sqrt{\mathrm{Tr}(X-Y)\rho(X-Y)^*}.\end{aligned}$$

Similarly, $\|Y\rho(X-Y)^*\|_1 \leq \sqrt{\mathrm{Tr}(X-Y)\rho(X-Y)^*}$, and the lemma follows. □

We state the following two corollaries of Lemma 35.

Claim 36. *Let $\{A_i\}$ and $\{B_i\}$ be two sets of positive matrices of the same dimension, and $\rho \geq 0$. Then*

$$\Big\| \sum_i \sqrt{A_i}\,\rho\,\sqrt{A_i} - \sqrt{B_i}\,\rho\,\sqrt{B_i}\Big\|_1 \;\leq\; 2\Big(\sum_i \mathrm{Tr}\big(\big(\sqrt{A_i}-\sqrt{B_i}\big)^2\rho\big)\Big)^{1/2}.$$

Proof. Let X be a block-column matrix with blocks the $\sqrt{A_i}$, and similarly for Y and the $\sqrt{B_i}$. Then

$$\Big\| \sum_i \sqrt{A_i}\,\rho\,\sqrt{A_i} - \sqrt{B_i}\,\rho\,\sqrt{B_i}\Big\|_1 \;\leq\; \sum_i \Big\|\sqrt{A_i}\,\rho\,\sqrt{A_i} - \sqrt{B_i}\,\rho\,\sqrt{B_i}\Big\|_1 \;\leq\; \big\|X\rho X^\dagger - Y\rho Y^\dagger\big\|_1,$$

and

$$\mathrm{Tr}\big((X-Y)\rho(X-Y)^\dagger\big) \;=\; \sum_i \mathrm{Tr}\big(\big(\sqrt{A_i}-\sqrt{B_i}\big)^2\rho\big),$$

so that the claim follows from Lemma 35. □

Claim 37. *Let $\sigma \geq 0$ be a (possibly un-normalized) density matrix on 3 registers, and suppose that σ is invariant with respect to permutation of the first two registers. Let $\{A_i\}_i$ be a POVM on either of the first two registers, and let*

$$\delta := \sum_{i\neq j} \mathrm{Tr}\big((A_i \otimes A_j \otimes \mathrm{Id})\sigma\big).$$

Then

$$\|\sum_i \big(\sqrt{A_i}\otimes \mathrm{Id}\big)\,\mathrm{Tr}_2(\sigma)\,\big(\sqrt{A_i}\otimes\mathrm{Id}\big) - \mathrm{Tr}_2(\sigma)\big\|_1 \;=\; O(\sqrt{\delta}),$$

where here $\sqrt{A_i}$ acts on the first register of σ, and the identity on the third.

Proof. First note that, $\{A_i\}_i$ being a POVM,

$$\mathrm{Tr}_2\big(\sum_i \big(\mathrm{Id}\otimes\sqrt{A_i}\otimes\mathrm{Id}\big)\,\sigma\,\big(\mathrm{Id}\otimes\sqrt{A_i}\otimes\mathrm{Id}\big)\big) \;=\; \mathrm{Tr}_2(\sigma).$$

Hence by monotonicity of the trace norm

$$\begin{aligned}
&\Big\| \sum_i \sqrt{A_i} \otimes \mathrm{Id}(\mathrm{Tr}_2(\sigma))\sqrt{A_i} \otimes \mathrm{Id} - \mathrm{Tr}_2(\sigma)\Big\|_1 \\
&\qquad\leq \Big\| \sum_{i,j} \sqrt{A_i} \otimes \sqrt{A_j} \otimes \mathrm{Id}\,\sigma \sqrt{A_i} \otimes \sqrt{A_j} \otimes \mathrm{Id} - \sum_j \mathrm{Id} \otimes \sqrt{A_j} \otimes \mathrm{Id}\,\sigma \mathrm{Id} \otimes \sqrt{A_j} \otimes \mathrm{Id}\, \Big\|_1 \\
&\qquad\leq \Big\| \sum_{i} \sqrt{A_i} \otimes \sqrt{A_i} \otimes \mathrm{Id}\,\sigma \sqrt{A_i} \otimes \sqrt{A_i} \otimes \mathrm{Id} - \sum_i \mathrm{Id} \otimes \sqrt{A_i} \otimes \mathrm{Id}\,\sigma \mathrm{Id} \otimes \sqrt{A_i} \otimes \mathrm{Id}\, \Big\|_1 \\
&\qquad\qquad + \sum_{i \neq j} \mathrm{Tr}\big(A_i \otimes A_j \otimes \mathrm{Id}\,\sigma\big) \\
&\qquad\leq 2\sqrt{\sum_i \mathrm{Tr}\big((\sqrt{A_i} \otimes \sqrt{A_i} \otimes \mathrm{Id} - \mathrm{Id} \otimes \sqrt{A_i} \otimes \mathrm{Id})^2\sigma\big)} + \delta \\
&\qquad\leq 2\sqrt{\delta} + \delta
\end{aligned}$$

where the second inequality is the triangle inequality, the third is by Claim 36, and for the last we expanded

$$\begin{aligned}
&\sum_i \mathrm{Tr}\big((\sqrt{A_i} \otimes \sqrt{A_i} \otimes \mathrm{Id} - \mathrm{Id} \otimes \sqrt{A_i} \otimes \mathrm{Id})^2\sigma\big) \\
&\qquad= \sum_i \Big(\mathrm{Tr}\big(A_i \otimes A_i \otimes \mathrm{Id}\,\sigma\big) + \mathrm{Tr}\big(\,\mathrm{Id} \otimes A_i \otimes \mathrm{Id}\,\sigma\big) - 2\mathrm{Tr}\big(\sqrt{A_i} \otimes A_i \otimes \mathrm{Id}\,\sigma\big)\Big) \\
&\qquad\leq \sum_i \Big(\mathrm{Tr}\big(A_i \otimes A_i \otimes \mathrm{Id}\,\sigma\big) + \mathrm{Tr}\big(\,\mathrm{Id} \otimes A_i \otimes \mathrm{Id}\,\sigma\big) - 2\mathrm{Tr}\big(A_i \otimes A_i \otimes \mathrm{Id}\,\sigma\big)\Big) \\
&\qquad= \delta,
\end{aligned}$$

where for the inequality $\sqrt{A_i} \geq A_i$ follows from $0 \leq A_i \leq \mathrm{Id}$ for every i, and the last equality uses the definition of δ and $\sum_i A_i = \mathrm{Id}$. □

The following lemma follows from the standard expansion properties of the hypercube. Recall that for $\rho \geq 0$ and any A, $\|A\|_\rho^2 = \mathrm{Tr}\big(AA^\dagger \rho\big)$.

Claim 38 (Expansion lemma). *Let $\varepsilon > 0$, S a finite set of size $|S| = p$, n, d integers and $A : S^n \to \mathbb{C}^{d\times d}$ such that for every $\boldsymbol{x} \in S^n$, $0 \leq A_{\boldsymbol{x}} \leq \mathrm{Id}$, and*

$$\mathrm{E}_{i,\boldsymbol{x}_{\neg i},x_i,x_i'} \big\|A_{\boldsymbol{x}} - A_{\boldsymbol{x}'}\big\|_\rho^2 \;\leq\; \varepsilon,$$

where the expectation is taken with respect to the uniform distribution on $[n] \times S^{n-1} \times S \times S$. Then

$$\mathrm{E}_{\boldsymbol{x}}\Big\|A_{\boldsymbol{x}} - \mathrm{E}_{\boldsymbol{x}} A_{\boldsymbol{x}}\Big\|_\rho^2 \;\leq\; 2n\varepsilon,$$

where both expectations are taken under the uniform distribution over S^n.

Proof. Let $M := \sum_{\boldsymbol{x},i,x_i'} |\boldsymbol{x}\rangle\langle \boldsymbol{x}'|$ be the adjacency matrix of the hypercube S^n, $L := np\,\mathrm{Id} - M$ the Laplacian, and $\tilde{L} = L \otimes \rho$. Let $A = \sum_{\boldsymbol{x}} |\boldsymbol{x}\rangle \otimes A_x$. Then

$$A^\dagger \tilde{L} \cdot A = \frac{1}{2} \sum_{\boldsymbol{x},i,x_i'} (A_{\boldsymbol{x}} - A_{\boldsymbol{x}'})^\dagger \rho (A_{\boldsymbol{x}} - A_{\boldsymbol{x}'}). \tag{27}$$

The normalized Laplacian $L/(np)$ has smallest eigenvalue 0, and second smallest $\lambda_1 \geq 1/(2n)$. Let the smallest eigenvector of L be $|v_0\rangle = p^{-n/2}\sum_{\boldsymbol{x}} |\boldsymbol{x}\rangle$, and write $A = |v_0\rangle \otimes A_0 + |v_1\rangle \otimes A_1$, where $|v_1\rangle$ is orthogonal to $|v_0\rangle$, and $A_0 = p^{-n/2}\sum_{\boldsymbol{x}} A_{\boldsymbol{x}}$. Then

$$A^\dagger \tilde{L} A \;=\; \lambda_1 A_1^\dagger \rho A_1 \;\geq\; \frac{1}{2n} A_1^\dagger \rho A_1.$$

Taking the trace and using the assumption made in the claim's statement together with (27), we get $\|A_1\|_\rho^2 \leq 2n\varepsilon p^n$, and hence by definition of A,

$$\mathrm{Tr}\big((A - |v_0\rangle \otimes A_0)^\dagger (\mathrm{Id} \otimes \rho)(A - |v_0\rangle \otimes A_0)\big) \;=\; \|A_1\|_\rho^2 \;\leq\; 2n\varepsilon p^n,$$

which proves the claim. □

B Lemmas about consistency

The following useful lemma relates the consistency of a measurement when performed on two separate subsystems of a permutation-invariant state with the possibility of exchanging the subsystem on which the measurement is performed. Here ρ is the reduced density of a permutation-invariant state.

Lemma 39. *Let $k \geq \ell \geq 1$ be two integers, T a family of sub-measurements of arity k, and V a family of sub-measurements of arity ℓ. Let $\{Z^h_{\boldsymbol{x}_{\geq k}}\}$ be such that $\mathrm{E}_{\boldsymbol{x}} \sum_h Z^h_{\boldsymbol{x}_{\geq k}} \big(Z^h_{\boldsymbol{x}_{\geq k}}\big)^\dagger \leq \mathrm{Id}$. Then it holds that*

$$\Big|\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho\big(Z^h_{\boldsymbol{x}_{\geq k}} T^h_{\boldsymbol{x}_{\geq k}} \otimes V_{\boldsymbol{x}_{\geq \ell}}\big) - \mathrm{E}_{\boldsymbol{x}} \sum_{g,h:\, h_{|x_\ell,\dots,x_{k-1}} = g} \mathrm{Tr}_\rho\big(Z^h_{\boldsymbol{x}_{\geq k}} T^h_{\boldsymbol{x}_{\geq k}} \otimes V^g_{\boldsymbol{x}_{\geq \ell}}\big)\Big| \;\leq\; \sqrt{\textsc{inc}(T,V)}.$$

Proof. The proof is a direct consequence of the Cauchy-Schwarz inequality: write

$$\begin{aligned}
&\Big|\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho\big(Z^h_{\boldsymbol{x}_{\geq k}} T^h_{\boldsymbol{x}_{\geq k}} \otimes V_{\boldsymbol{x}_{\geq \ell}}\big) - \mathrm{E}_{\boldsymbol{x}} \sum_{g,h:\, h_{|x_\ell,\dots,x_{k-1}} = g} \mathrm{Tr}_\rho\big(Z^h_{\boldsymbol{x}_{\geq k}} T^h_{\boldsymbol{x}_{\geq k}} \otimes V^g_{\boldsymbol{x}_{\geq \ell}}\big)\Big| \\
&\quad = \Big|\mathrm{E}_{\boldsymbol{x}} \sum_{g,h:\, h_{|x_\ell,\dots,x_{k-1}} \neq g} \mathrm{Tr}_\rho\big(Z^h_{\boldsymbol{x}_{\geq k}} T^h_{\boldsymbol{x}_{\geq k}} \otimes V^g_{\boldsymbol{x}_{\geq \ell}}\big)\Big| \\
&\quad \leq \Big(\mathrm{E}_{\boldsymbol{x}} \sum_{g,h:\, h_{|x_\ell,\dots,x_{k-1}} \neq g} \mathrm{Tr}_\rho\big(T^h_{\boldsymbol{x}_{\geq k}} \otimes V^g_{\boldsymbol{x}_{\geq \ell}}\big)\Big)^{1/2} \Big(\mathrm{E}_{\boldsymbol{x}} \sum_{g,h} \mathrm{Tr}_\rho\big(Z^h_{\boldsymbol{x}_{\geq k}} (Z^h_{\boldsymbol{x}_{\geq k}})^\dagger \otimes V^g_{\boldsymbol{x}_{\geq \ell}}\big)\Big)^{1/2} \\
&\quad \leq \sqrt{\textsc{inc}(T,V)},
\end{aligned}$$

where the last inequality follows from the definition of $\textsc{inc}(T,V)$ and our assumption on $Z^h_{\boldsymbol{x}_{\geq k}}$. □

Lemma 40. *Let T be a family of sub-measurements of arity k, X such that $X^\dagger X \leq \mathrm{Id}$, and $\{Z^h_{\boldsymbol{x}_{\geq k}}\}$ such that $\mathrm{E}_{\boldsymbol{x}} \sum_h Z^h_{\boldsymbol{x}_{\geq k}} \big(Z^h_{\boldsymbol{x}_{\geq k}}\big)^\dagger \leq \mathrm{Id}$ (for instance, a family of sub-measurements of arity ℓ, for any ℓ). Then*[20]

$$\Big|\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho(Z^h_{\boldsymbol{x}_{\geq k}} T^h_{\boldsymbol{x}_{\geq k}} \otimes T_{\boldsymbol{x}_{\geq k}}) - \mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho(Z^h_{\boldsymbol{x}_{\geq k}} T_{\boldsymbol{x}_{\geq k}} \otimes T^h_{\boldsymbol{x}_{\geq k}})\Big| \;\leq\; \sqrt{\textsc{inc}(T,T)} \tag{28}$$

$$\Big|\mathrm{E}_{\boldsymbol{x}} \sum_{h \neq h'} \mathrm{Tr}_\rho(T^h_{\boldsymbol{x}_{\geq k}} X T^{h'}_{\boldsymbol{x}_{\geq k}} \otimes T_{\boldsymbol{x}_{\geq k}})\Big| \;\leq\; 2\sqrt{\textsc{inc}(T,T)} \tag{29}$$

[20] A special case of interest is when the measurements are *complete*, in which case the statements simplify.

Proof. We first prove (28). We have

$$
\begin{aligned}
&\Big|\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho (\mathrm{Z}^h_{\boldsymbol{x}_{\geq \ell}} T^h_{\boldsymbol{x}_{\geq k}} \otimes T_{\boldsymbol{x}_{\geq k}}) - \mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho (\mathrm{Z}^h_{\boldsymbol{x}_\ell} \otimes T^h_{\boldsymbol{x}_{\geq k}})\Big| \\
&\qquad\qquad = \Big|\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho \big(\mathrm{Z}^h_{\boldsymbol{x}_{\geq \ell}} (T^h_{\boldsymbol{x}_{\geq k}} \otimes T_{\boldsymbol{x}_{\geq k}} - T_{\boldsymbol{x}_{\geq k}} \otimes T^h_{\boldsymbol{x}_{\geq k}})\big)\Big| \\
&\qquad\qquad \leq \Big(\sum_i \mathrm{Tr}_\rho \big(\mathrm{Z}^h_{\boldsymbol{x}_k} (\mathrm{Z}^h_{\boldsymbol{x}_k})^\dagger\big)\Big)^{1/2} \Big(\sum_{h \neq h'} \mathrm{Tr}_\rho \big((T^h_{\boldsymbol{x}_{\geq k}} \otimes T^{h'}_{\boldsymbol{x}_{\geq k}})^2\big)\Big)^{1/2} \\
&\qquad\qquad \leq \sqrt{\mathrm{INC}(T, T)},
\end{aligned}
$$

where the second inequality follows from Cauchy-Schwarz. Regarding (29), we have

$$
\Big|\mathrm{E}_{\boldsymbol{x}} \sum_{h \neq h'} \mathrm{Tr}_\rho (T^h_{\boldsymbol{x}_{\geq k}} X T^{h'}_{\boldsymbol{x}_{\geq k}} \otimes T_{\boldsymbol{x}_{\geq k}})\Big| = \Big|\mathrm{E}_{\boldsymbol{x}} \mathrm{Tr}_\rho \big(T_{\boldsymbol{x}_{\geq k}} X T_{\boldsymbol{x}_{\geq k}} \otimes T_{\boldsymbol{x}_{\geq k}}\big) - \mathrm{E}_x \sum_h \mathrm{Tr}_\rho \big(T^h_{\boldsymbol{x}_{\geq k}} X T^h_{\boldsymbol{x}_{\geq k}} \otimes T_{\boldsymbol{x}_{\geq k}}\big)\Big|
$$

From (28) we know that

$$
\Big|\mathrm{E}_x \sum_h \mathrm{Tr}_\rho \big(T^h_{\boldsymbol{x}_{\geq k}} X T^h_{\boldsymbol{x}_{\geq k}}\big) \otimes T_{\boldsymbol{x}_{\geq k}} - \mathrm{E}_x \sum_h \mathrm{Tr}_\rho \big(T^h_{\boldsymbol{x}_{\geq k}} X T_{\boldsymbol{x}_{\geq k}} \otimes T^h_{\boldsymbol{x}_{\geq k}}\big)\Big| \leq \sqrt{\mathrm{INC}(T, T)}.
$$

The second term on the left-hand side satisfies

$$
\begin{aligned}
&\Big|\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho \big(T^h_{\boldsymbol{x}_{\geq k}} X T_{\boldsymbol{x}_{\geq k}} \otimes T^h_{\boldsymbol{x}_{\geq k}}\big) - \mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho \big(T_{\boldsymbol{x}_{\geq k}} X T_{\boldsymbol{x}_{\geq k}} \otimes T^h_{\boldsymbol{x}_{\geq k}}\big)\Big| \\
&\qquad \leq \Big(\mathrm{E}_{\boldsymbol{x}} \sum_h \mathrm{Tr}_\rho \big(T_{\boldsymbol{x}_{\geq k}} X^\dagger X T_{\boldsymbol{x}_{\geq k}} \otimes T^h_{\boldsymbol{x}_{\geq k}}\big)\Big)^{1/2} \Big(\mathrm{E}_x \sum_h \mathrm{Tr}_\rho \big((T_{\boldsymbol{x}_{\geq k}} - T^h_{\boldsymbol{x}_{\geq k}})^2 \otimes T^h_{\boldsymbol{x}_{\geq k}}\big)\Big)^{1/2} \\
&\qquad \leq \sqrt{\mathrm{INC}(T, T)},
\end{aligned}
$$

and this concludes the proof. □

C Proof of Corollary 10

In this section we give the proof of Corollary 10. A standard method to convert multiple constraints to a single constraint involving an exponential sum is by using small-bias probability spaces.

Definition 41 (Small-bias probability space)**.** *Let $n \in \mathbb{N}$. A set $S \subseteq \mathbb{F}_2^n$ is called an ε-*bias probability space *if for every $\boldsymbol{c} \in \mathbb{F}_2^n \setminus \{0\}$, it holds that*

$$
\big|\Pr_{\boldsymbol{\zeta} \in S}[\boldsymbol{c} \cdot \boldsymbol{\zeta} = 0] - \Pr_{\boldsymbol{\zeta} \in S}[\boldsymbol{c} \cdot \boldsymbol{\zeta} = 1]\big| \leq \varepsilon.
$$

Proposition 42. *Let $n \in \mathbb{N}$, and let $S \subset \mathbb{F}_2^n$ be an ε-bias probability space. Let $\mathbb{F}$ be a finite field of characteristic two. If $\boldsymbol{c} \in \mathbb{F}^n \setminus \{0\}$, then*

$$
\Pr_{\zeta \in S}\left[\sum_{i=1}^n \zeta_i c_i = 0\right] \leq \frac{1 + \varepsilon}{2}.
$$

Proof. If $\mathbb{F} = \mathbb{F}_2$, then the proposition holds because

$$\Pr_{\zeta \in S}\left[\sum_{i=1}^{n} c_i \zeta_i = 0\right] = \frac{1}{2} + \frac{1}{2}\left(\Pr_{\zeta \in S}\left[\sum_{i=1}^{n} c_i \zeta_i = 0\right] - \Pr_{\zeta \in S}\left[\sum_{i=1}^{n} c_i \zeta_i = 1\right]\right)$$
$$\leq \frac{1+\varepsilon}{2}.$$

For general $\mathbb{F}$, regard $\mathbb{F}$ as a vector space over $\mathbb{F}_2$, and let $\{\alpha_1, \dots, \alpha_k\}$ be a basis of $\mathbb{F}$ over $\mathbb{F}_2$. Write $\boldsymbol{c}$ as $\boldsymbol{c} = \alpha_1 \boldsymbol{c}^{(1)} + \cdots + \alpha_k \boldsymbol{c}^{(k)}$, where $\boldsymbol{c}^{(1)}, \dots, \boldsymbol{c}^{(k)} \in \mathbb{F}_2^n$. Because $\boldsymbol{c} \neq 0$, we have that $\boldsymbol{c}^{(j^*)} \neq 0$ for some j^*. By using the case of $\mathbb{F}_2$, it holds that

$$\Pr_{\zeta \in S}\left[\sum_{i=1}^{n} c_i^{(j^*)} \zeta_i = 0\right] \leq \frac{1+\varepsilon}{2}.$$

Since $\alpha_1, \dots, \alpha_k$ are linearly independent over $\mathbb{F}_2$, $\sum_{i=1}^{n} c_i \zeta_i = 0$ implies $\sum_{i=1}^{n} c_i^{(j)} \zeta_i = 0$ for all j, and therefore in particular $\sum_{i=1}^{n} c_i^{(j^*)} \zeta_i = 0$. Therefore,

$$\Pr_{\zeta \in S}\left[\sum_{i=1}^{n} c_i \zeta_i = 0\right] \leq \Pr_{\zeta \in S}\left[\sum_{i=1}^{n} c_i^{(j^*)} \zeta_i = 0\right] \leq \frac{1+\varepsilon}{2}. \qquad \square$$

Theorem 43 (Alon, Goldreich, Håstad, and Peralta [AGHP92]). *There exist a constant $c > 0$ and a polynomial-time algorithm C which, given $K, M \in \mathbb{N}$, $i \in \{1, \dots, K\}$ and $j \in \{1, \dots, M\}$, outputs a $C(K, M, i, j) \in \mathbb{F}_2$ such that the set $\{\boldsymbol{\zeta}^{(j)} \colon 1 \leq j \leq M\}$ defined by $\boldsymbol{\zeta}^{(j)} = (C(K, M, 1, j), \dots, C(K, M, K, j))$ is an (K/M^c)-bias probability space in $\mathbb{F}_2^K$.*

By arithmetizing the Boolean circuit for C by using a similar idea to the proof of Proposition 4.2 of Ref. [BFL91], we obtain the following corollary.

Corollary 44. *There exist a constant $c > 0$ and a polynomial-time algorithm A which, given 1^k and 1^m, outputs 1^t and an arithmetic expression $f(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{l})$ in $k + m + t$ variables such that the set $\{\boldsymbol{\zeta}^{(j)} \colon \boldsymbol{j} \in \{0,1\}^m\}$ defined by $\boldsymbol{\zeta}^{(j)} = (\sum_{\boldsymbol{l} \in \{0,1\}^t} f(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{l}))_{\boldsymbol{i} \in \{0,1\}^k}$ is an 2^{k-cm}-bias probability space in $\mathbb{F}_2^{2^k}$.*

Proof of Corollary 10. The protocol works as follows. The verifier first computes $m = \lceil (k+2)/c \rceil$, where c is the constant in Corollary 44. He runs the algorithm of Corollary 44 with parameters k and m to obtain $t \in \mathbb{N}$ and an arithmetic expression $f(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{l})$ in $k + m + t$ variables. Let d' be the maximum degree of f in single variables. He chooses $\boldsymbol{j} \in \{0,1\}^m$ uniformly at random, and sends $\boldsymbol{j}$ to the prover. Then he simulates the protocol in Lemma 9 with explicit inputs $k + t$ and $d + d'$ and implicit input $h_j(\boldsymbol{i}, \boldsymbol{l}) := f(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{l}) h(\boldsymbol{i})$.

For $\boldsymbol{i} \in \mathbb{F}^k$, $\boldsymbol{j} \in \mathbb{F}^m$, and $\boldsymbol{l} \in \mathbb{F}^t$, let $\zeta_{\boldsymbol{i}}^{(j)} = \sum_{\boldsymbol{l}} f(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{l}) \in \mathbb{F}$ and $\boldsymbol{\zeta}^{(j)} = (\zeta_{\boldsymbol{i}}^{(j)})_{\boldsymbol{i} \in \{0,1\}^k} \in \mathbb{F}^{2^k}$. Because $m \geq (k+2)/c$, Corollary 44 guarantees that $\{\boldsymbol{\zeta}^{(j)} \colon \boldsymbol{j} \in \{0,1\}^m\}$ is a $1/4$-bias probability space.

Let $c_{\boldsymbol{i}} = h(\boldsymbol{i})$. Then for all $\boldsymbol{j} \in \{0,1\}^m$, it holds that

$$\sum_{\boldsymbol{i} \in \{0,1\}^k, \boldsymbol{l} \in \{0,1\}^t} h_j(\boldsymbol{i}, \boldsymbol{l}) = \sum_{\boldsymbol{i} \in \{0,1\}^k} \zeta_{\boldsymbol{i}}^{(j)} c_{\boldsymbol{i}}. \tag{30}$$

Completeness: Suppose that $c_{\boldsymbol{i}} = 0$ for all $\boldsymbol{i} \in \{0,1\}^k$. Then, by Eq. (30), it holds that

$$\sum_{\boldsymbol{i} \in \{0,1\}^k, \boldsymbol{l} \in \{0,1\}^t} h_j(\boldsymbol{i}, \boldsymbol{l}) = 0$$

for all $\boldsymbol{j} \in \{0,1\}^m$. Therefore, the completeness of the protocol in Lemma 9 implies that the protocol constructed above also has perfect completeness.

Soundness: Suppose that $\boldsymbol{c} \neq 0$. By Proposition 42, it holds that

$$\Pr_{\boldsymbol{j}\in\{0,1\}^m}\Big[\sum_{\boldsymbol{i}\in\{0,1\}^k} \zeta_{\boldsymbol{i}}^{(\boldsymbol{j})} c_{\boldsymbol{i}} = 0\Big] \leq \frac{1+1/4}{2} = \frac{5}{8}.$$

Eq. (30) and the soundness in Lemma 9 imply that for any $\boldsymbol{j} \in \{0,1\}^m$ such that $\sum_{\boldsymbol{i}\in\{0,1\}^k} \zeta_{\boldsymbol{i}}^{(\boldsymbol{j})} c_{\boldsymbol{i}} \neq 0$, the acceptance probability conditioned on the choice of $\boldsymbol{j}$ is at most $(d+d')(k+t)/|\mathbb{F}|$. Therefore, the overall acceptance probability is at most $5/8 + (d+d')(k+t)/|\mathbb{F}|$. The corollary follows because d' and t are polynomially bounded in k. □

References

[Aar05] Scott Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1(1):1–28, 2005.

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992.

[ALMSS98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

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

[BD95] Rajendra Bhatia and Chandler Davis. A Cauchy–Schwarz inequality for operators with applications. *Linear Algebra and its Applications*, 223–224:119–129, 1995.

[Bel64] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[BFK10] Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. QMIP = MIP*. Technical report, arXiv:1004.1130, 2010.

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3–40, 1991.

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

[Bha88] Rajendra Bhatia. Perturbation inequalities for the absolute value map in norm ideals of operators. *Journal of Operator Theory*, 19(1):129–136, 1988.

[BHP08] Michael Ben-Or, Avinatan Hassidim, and Haran Pilpel. Quantum multi prover interactive proofs with communicating provers. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 467–476, 2008.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.

[CGJ09] Richard Cleve, Dmitry Gavinsky, and Rahul Jain. Entanglement-resistant two-prover interactive proof systems and non-adaptive PIR. *Quantum Information and Computation*, 2009.

[CHTW04] Richard Cleve, Peter Høyer, Benjamin Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *Proceedings of the 19th IEEE Annual Conference on Computational Complexity*, pages 236–249, 2004.

[DLTW08] Andrew C. Doherty, Yeong-Cherng Liang, Benjamin Toner, and Stephanie Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *Proceedings of the 23rd IEEE Annual Conference on Computational Complexity*, pages 199–210, 2008.

[FL92] Uriel Feige and László Lovász. 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.

[Gol08] Oded Goldreich. *Computational Complexity: A Conceptual Perspective*. Cambridge University Press, 2008.

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48:798–859, 2001.

[IKM09] Tsuyoshi Ito, Hirotada Kobayashi, and Keiji Matsumoto. Oracularization and two-prover one-round interactive proofs against nonlocal strategies. In *Proceedings of the 24th IEEE Annual Conference on Computational Complexity*, pages 217–228, 2009.

[IKPSY08] Tsuyoshi Ito, Hirotada Kobayashi, Daniel Preda, Xiaoming Sun, and Andrew C.-C. Yao. Generalized Tsirelson inequalities, commuting-operator provers, and multi-prover interactive proof systems. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 187–198, 2008.

[Ito11] Tsuyoshi Ito. Parallelization of entanglement-resistant multi-prover interactive proofs, 2011. Submitted.

[JJUW11] Rahul Jain, Zhengfeng Ji, Sarvaghya Upadhyay, and John Watrous. QIP = PSPACE. *Journal of the ACM*, 58(6):30:1–30:27, 2011.

[JPPVW10] M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf. Operator space theory: A natural framework for bell inequalities. *Phys. Rev. Lett.*, 104:170405, Apr 2010.

[JUW09] Rahul Jain, Sarvagya Upadhyay, and John Watrous. Two-message quantum interactive proofs are in PSPACE. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 534–543, 2009.

[KKMTV11] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, Ben Toner, and Thomas Vidick. Entangled games are hard to approximate. *SIAM Journal on Computing*, 40(3):848–877, 2011.

[KM03] Hirotada Kobayashi and Keiji Matsumoto. Quantum multi-prover interactive proof systems with limited prior entanglement. *Journal of Computer and System Sciences*, 66(3):429–450, 2003.

[KRT10] Julia Kempe, Oded Regev, and Ben Toner. Unique games with entangled provers are easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010.

[KSV02] Alexei Yu. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.

[KV11] Julia Kempe and Thomas Vidick. Parallel repetition of entangled games. In *Proceedings of the 43rd Annual ACM Symposium on the Theory of Computing,* San Jose CA, pages 353–362, 2011.

[LFKN92] Carsten Lund, Lance Fortnow, Howard Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM*, 39:859–868, 1992.

[Mer90] N. David Mermin. Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.*, 65:3373–3376, 1990.

[NC01] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2001.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(073013), 2008.

[ON07] Tomohiro Ogawa and Hiroshi Nagaoka. Making good codes for classical-quantum channel coding via quantum hypothesis testing. *IEEE Transactions on Information Theory*, 53(6):2261–2266, 2007.

[Per90] Asher Peres. Incompatible results of quantum measurements. *Physics Letters A*, 151(3-4):107–108, 1990.

[RT07] Alex Rapaport and Amnon Ta-Shma. On the power of quantum, one round, two prover interactive proof systems. *Quantum Information Processing*, 6:445–459, 2007. 10.1007/s11128-007-0068-z.

[Sch80] Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):707–717, 1980.

[Sha92] Adi Shamir. IP = PSPACE. *Journal of the ACM*, 39(4):869–877, 1992.

[SW08] Volkher B. Scholz and Reinhard F. Werner. Tsirelson's Problem. Technical report, arXiv:0812.4305, 2008.

[Tsi80] Boris S. Tsirelson. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[Wat09] John Watrous. Quantum computational complexity. In Robert A. Meyers, editor, *Encyclopedia of Complexity and System Science*. Springer, 2009.

[Weh06] Stephanie Wehner. Entanglement in interactive proof systems with binary answers. In *Proceedings of the 23rd Annual Symposium on Theoretical Aspects of Computer Science*, volume 3884 of *Lecture Notes in Computer Science*, pages 162–171, 2006.

[Win99] Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999.

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposiumon on Symbolic and Algebraic Computation*, pages 216–226, 1979.