The Computational Advantage of MIP* Vanishes in the Presence of Noise

Yangjing Dong
State Key Laboratory for Novel Software Technology, Nanjing University, China

Honghao Fu
Massachusetts Institute of Technology, Cambridge, MA, USA

Anand Natarajan
Massachusetts Institute of Technology, Cambridge, MA, USA

Minglong Qin
State Key Laboratory for Novel Software Technology, Nanjing University, China

Haochen Xu
State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China

Penghui Yao
State Key Laboratory for Novel Software Technology, Nanjing University, China
Hefei National Laboratory, China

Abstract

The class MIP* of quantum multiprover interactive proof systems with entanglement is much more powerful than its classical counterpart MIP [7, 29, 28]: while MIP = NEXP, the quantum class MIP* is equal to RE, a class including the halting problem. This is because the provers in MIP* can share unbounded quantum entanglement. However, recent works [46, 47] have shown that this advantage is significantly reduced if the provers' shared state contains noise. This paper attempts to exactly characterize the effect of noise on the computational power of quantum multiprover interactive proof systems. We investigate the quantum two-prover one-round interactive system MIP* $[\mathrm{poly}, O(1)]$, where the verifier sends polynomially many bits to the provers and the provers send back constantly many bits. We show noise completely destroys the computational advantage given by shared entanglement in this model. Specifically, we show that if the provers are allowed to share arbitrarily many EPR states, where each EPR state is affected by an arbitrarily small constant amount of noise, the resulting complexity class is equivalent to NEXP = MIP. This improves significantly on the previous best-known bound of NEEEXP (nondeterministic triply exponential time) [46]. We also show that this collapse in power is due to the noise, rather than the $O(1)$ answer size, by showing that allowing for noiseless EPR states gives the class the full power of RE = MIP* $[\mathrm{poly}, \mathrm{poly}]$. Along the way, we develop two technical tools of independent interest. First, we give a new, deterministic tester for the positivity of an exponentially large matrix, provided it has a low-degree Fourier decomposition in terms of Pauli matrices. Secondly, we develop a new invariance principle for smooth matrix functions having bounded third-order Fréchet derivatives or which are Lipschitz continuous.

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases Interactive proofs, Quantum complexity theory, Quantum entanglement, Fourier analysis, Matrix analysis, Invariance principle, Derandomization, PCP, Locally testable code, Positivity testing

Digital Object Identifier 10.4230/LIPIcs.CCC.2024.30

Related Version *Full Version*: https://arxiv.org/abs/2312.04360 [17]

Funding *Yangjing Dong*: supported by National Natural Science Foundation of China (Grant No. 62332009, 12347104, 61972191) and Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0302901).

39th Computational Complexity Conference (CCC 2024).
Editor: Rahul Santhanam; Article No. 30; pp. 30:1–30:71

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Honghao Fu: supported by the US National Science Foundation QLCI program (grant OMA-2016245).
Minglong Qin: supported by National Natural Science Foundation of China (Grant No. 62332009, 12347104, 61972191) and Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0302901).
Haochen Xu: supported by the Key Research Program of the Chinese Academy of Sciences, Grant NO. ZDRW-XX-2022-1
Penghui Yao: supported by National Natural Science Foundation of China (Grant No. 62332009, 12347104, 61972191) and Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0302901).

Acknowledgements P.Y. would like to thank the discussion with Zhengfeng Ji. Part of the work was done when H.F. and H.X. visited Nanjing University.

1 Introduction

The power of entanglement in computation has been a central topic in the theory of quantum computing. In particular, the effect of entanglement in multiprover interactive proof systems has been studied for decades [33, 32, 25, 27, 54, 53] leading to the seminal result MIP* = RE [29, 28] due to Ji, Natarajan, Vidick, Wright, and Yuen, which states that all recursively enumerable languages can be decided by multiprover interactive proof systems empowered by quantum entanglement. More precisely, the system only has two provers, one round of interaction between the provers and the verifier, and the provers share arbitrarily many copies of the EPR state.

Given the appearance of intractable complexity classes like RE in the previous result, a natural question is to what extent the body of results on MIP* are relevant to the physical world. Of course, in reality, devices do not have access to unbounded numbers of perfect EPR pairs; in a sense, what MIP* = RE means is that the power of two entangled grows unboundedly as the number of shared EPR pairs increases, even when the message size is constrained to be polynomial. In fact, using a finite number of iterations of the "compression" procedure from MIP* = RE, one can show that the class NTIME[$T(n)$] for $T(n)$ any finite tower of exponentials has an MIP* protocol, where the provers need only share a finite number of perfect EPR pairs scaling roughly with $\log T(n)$. However, the requirement that the EPR pairs be perfect seems essential to these protocols. The question naturally arises whether similar complexity results can be obtained even when the provers have access to *imperfect* entanglement only.

To isolate the role played by noise, in this work we ask the following question: what is the power of MIP* when the provers are given access to an *unbounded* number of *imperfect* EPR pairs, where each EPR pair is independently perturbed by a constant amount of depolarizing noise? (We choose this noise model because it is mathematically elegant and also physically relevant, as recent experiments suggest that the dominating noise is the localized depolarizing noise in the neutral atom platform [11].) On the one hand, known MIP* protocols all break down with states of this form. On the other hand, according to standard measures of entanglement such as distillable entanglement and entanglement of formation, such states have entanglement that grows unboundedly as the number of copies goes to infinity. Thus, it seems *a priori* reasonable that the corresponding MIP* class may also have unbounded power.

It is worth noting that this question is orthogonal to fault tolerance in quantum devices. As usual in MIP*, we assume that the provers are computationally unbounded, and may perform any quantum operation of their choice with no error. Nevertheless, this does not

mean they can use techniques from fault tolerance to simulate provers with noiseless entangled states. This is because the provers cannot jointly correct their shared entangled state, since they are not allowed to communicate in this model.

This question is closely related to the quantum information primitive of self-testing. Self-tests are essentially MIP* protocols that certify physical properties of quantum states, rather than computational statements. The protocols in MIP* = RE all rely on highly efficient self-tests for EPR pairs, but these tests are not at all tolerant of noise. Designing self-tests that *are* tolerant to noise, and certify some useful measure of entanglement, is a current research question [5, 3], and studying the power of MIP* in the presence of noise gives us insight on this question from a different angle. In particular, for an entangled state ρ, one can think of the power of the complexity class MIP$^*[\rho]$ where the provers are restricted to sharing copies of ρ, as a particular operational measure of the amount of useful entanglement in ρ. In passing, we remark that recent work of Vidick, Arnon-Friedman and Brakerski has studied "computationally efficient" measures of entanglement from somewhat different perspective [4].

The first partial answer to this question was given by Qin and Yao [46]. They investigated two-player nonlocal games[1] when the states shared between the players are arbitrarily many copies of a maximally entangled state (MES) with an arbitrarily small but *constant* amount of noise on each copy, which is termed as *noisy MES* in their paper. They showed that the supremum winning probability over all strategies using these states can be computably approximated to any finite precision. In fact, they showed that for any ε, there is a number of copies of the noisy MES, which is a computable function of only ε and the size of the nonlocal game, that is sufficient to achieve winning probability within ε of this supremum. This implies that any language in MIP* restricted to such states is decidable, meaning that this class is strictly smaller than RE.

This result was later generalized to nonlocal games that allow quantum questions and quantum answers [47]. To put these results in the language of complexity classes, let MIP$^*[q, a, \psi]$ be the set of languages that are decidable in the model of two-prover, one-round quantum multiprover interactive proof systems, where the provers share arbitrarily many copies of ψ, the messages from the verifier are classical and q-bits long, and the messages from the provers are also classical and a-bits long. If the messages are quantum, the complexity class is denoted by QMIP$[q, a, \psi]$. Thus, the prior work implies that both the class QMIP$[\text{poly}, \text{poly}, |\text{EPR}\rangle]$ and the class MIP$^*[\text{poly}, \text{poly}, |\text{EPR}\rangle]$ are equal to RE [49, 29, 28], while both the complexity classes MIP$^*[\text{poly}, \text{poly}, \psi]$ and QMIP$[\text{poly}, \text{poly}, \psi]$ are computable if ψ is a noisy MES state [46, 47]. Moreover, [46, 47] showed explicit, though very large, time bounds for computing approximations to the game value for noisy states.

Although these results show that the full power of MIP* is not robust against noise in the shared entanglement, it is still possible that multiprover interactive proof systems gain a finite but very large computational advantage by sharing noisy maximally entangled states, since the time bounds from the previous work are much larger than for the classes with no entanglement. Thus, it was consistent with prior work that MIP$^*[\text{poly}, \text{poly}, \psi]$ is contained in nondeterministic quadruply exponential time complexity class for noisy ψ [46], which is much more powerful than MIP$[\text{poly}, \text{poly}]$ = NEXP. This paper attempts to answer this question by investigating the complexity classes MIP$^*[\text{poly}, O(1), \psi]$ (i.e. protocols with constant-size answers) when ψ is a noisy MES, whose local dimension is a constant. Classically, it is known that MIP$[\text{poly}, \text{poly}]$ = MIP$[\text{poly}, O(1)]$ = NEXP [7]. Our main result, stated in the language of nonlocal games, is the following.

[1] An MIP* protocol is essentially a uniform family of two-player nonlocal games, with efficent algorithms for sampling pairs of questions and for evaluating the game decision predicate.

▶ **Theorem 1** (Informal). *Given a nonlocal game in which the players share arbitrarily many copies of a noisy MES ψ, and the size of the answer sets is constant, then approximating the value of the game up to any sufficiently small constant precision is* NP*-complete.*

The runtime in Theorem 1 is measured in terms of the size of a description of the nonlocal game as a table containing the distribution over question pairs and the verifier's predicate for every tuple of questions and answers. Translating this result to the MIP* world requires parametrizing the runtime in terms of the *number of bits* in the questions and answers. Thus, Theorem 1 shows that MIP* with $O(\log(n))$-bit questions and $O(1)$-bit answers is NP-complete. Scaling our result up to MIP* protocols with $O(\text{poly}(n))$-bit questions and $O(1)$-bit answers, we get the following.

▶ **Corollary 2.** $\text{MIP}^*[\text{poly}, O(1), \psi] = \text{NEXP}$*, where ψ is a noisy MES.*

Intuitively, Theorem 1 says that for any nonlocal game, if the shared MES has constant noise, the players' optimal strategy has a concise classical description which is also easy to verify. It is interesting to compare such nonlocal games with their classical counterparts. Håstad in his seminal work [23] proved that it is NP-hard to approximate the value of a classical nonlocal game to a constant precision even if the size of the answer set is a constant. It is also worth noting that sharing entanglement does not always strengthen the hardness of nonlocal games. It may weaken the hardness of certain games as well. For example, the quantum XOR games and quantum unique games are easy [15, 33], while the classical XOR games are NP-hard, and the classical unique games are conjectured to be NP-hard as well [34]. Thus introducing noisy quantum states doesn't introduce any quantum effect to the hardness at all.

One may wonder whether this surprising collapse in complexity is caused by the restriction to noisy states or the restriction to $O(1)$-size answers. We give strong evidence that it is the former, by showing that MIP* with *noiseless* states and $O(1)$-sized answers is still equal to RE.

▶ **Theorem 3** (Theorem 36). RE *is equal to* $\text{MIP}^*[\text{poly}, O(1), |\text{EPR}\rangle]$ *with completeness* 1 *and constant soundness.*

To put this in context, the original work [29, 28] proves that nonlocal games with noiseless EPR states are RE-complete to approximate if both the question set and answer set are of *polynomial* size. Very recently, Natarajan and Zhang [40] proved, by repeatedly applying the "question reduction" technique from [28], that it is still RE-complete if the question length is $O(1)$ and the answer length is $\text{polylog}(n)$, respectively. Here, we achieve constant *answer* length by one application of an "answer reduction" transformation: the error-correcting code-based scheme of [39], instantiated with the Hadamard code.

Theorems 1 and 3 give us strong evidence that the computational power of MIP* will vanish in the presence of noise. So for any complexity class slightly larger than NEXP, we cannot hope for an MIP* protocol robust against noise. They also suggest that the key resource behind the computational power of MIP* is specifically copies of the MES state, not just entanglement. This is because as we remarked above, as n tends to infinity, n copies of a noisy MES contain an amount of entanglement going to infinity under standard entanglement measures. Alternately, using the power of $\text{MIP}^*[\psi]$ as a measure of entanglement for ψ, we show that an MES and an ε-noisy MES are sharply separated by this measure for any constant ε.

Since efficient self-tests for large entangled states are the key technique behind the proof of $\text{MIP}^* = \text{RE}$, our result puts some limitations on the design of self-tests robust against noise. More specifically, our results suggests that to noise-robustly self-test larger entangled states, the numbers of questions and answers must grow with the dimension of the tested state. For comparison, if we don't need a self-test to be noise-robust, this is not necessary [19].

1.1 Proof Overview

1.1.1 Approximating the Values of Noisy Games is NP-Complete

The harder part is to show that there is an NP-algorithm for this problem, so we give an overview of this algorithm first.

Given a nonlocal game sharing arbitrary copies of a noisy MES ψ, Qin and Yao [46] showed that it suffices for the players to share D copies of ψ to achieve the value of the game to an arbitrarily small precision, where D only depends on the size of the game and the precision.

We first improve the upper bound D to make it only depend exponentially on the length of the questions instead of doubly exponentially as in [46]. To prove this upper bound, we use ideas from Fourier analysis. For illustration, let's assume $\psi = \rho|\mathrm{EPR}\rangle\langle\mathrm{EPR}| + (1-\rho)\mathbb{1}_2/2 \otimes \mathbb{1}_2/2$ is a depolarized noisy EPR state for simplicity. Given a strategy S, let P be a POVM element from the strategy, which acts on n qubits. We are going to show the upper bound is independent of n, so in the rest of the section by "constant" we mean independent of n. Let the Pauli expansion of P be

$$P = \sum_{\sigma\in\{0,1,2,3\}^n} \widehat{P}(\sigma)\,\mathcal{P}_\sigma,$$

where $\mathcal{P}_\sigma = \otimes_{i=1}^n \mathcal{P}_{\sigma_i}$ and $\mathcal{P}_0 = I, \mathcal{P}_1 = X, \mathcal{P}_2 = Y, \mathcal{P}_3 = Z$ are the single-qubit Pauli operators. The degree of a term $\widehat{P}(\sigma)\,\mathcal{P}_\sigma$ is the number of nontrivial Pauli's in it, denoted by $|\sigma|$. First, we adapt the smoothing technique in [46], which applies a depolarizing channel with small noise to P and removes the high-degree part of P, i.e. terms with $|\sigma| > d$ where d is a constant. After smoothing, S only contains degree-d operators

$$P^{(\mathrm{Smooth})} = \sum_{\sigma:|\sigma|\le d} \widehat{P^{(\mathrm{Smooth})}}(\sigma)\,\mathcal{P}_\sigma,$$

so we denote the new strategy by $S^{(\mathrm{Smooth})}$. Using the argument in [46], the probability of winning the game with this new strategy changes at most slightly, i.e.

$$\mathrm{val}^*(G, S^{(\mathrm{Smooth})}) \approx \mathrm{val}^*(G, S).$$

Let τ be a small constant independent of n. Since the degree of $P^{(\mathrm{Smooth})}$ is d, using a standard argument in the analysis of Boolean functions, the number of registers having influence that exceeds a given small τ is at most d/τ. Notice that d is independent of n, so is d/τ. Assume without loss of generality that $H = \{1, \ldots, |H|\}$ is the set of all registers whose influence exceeds τ. We apply the invariance principle from [46] to replace all the non-identity Pauli bases in the registers with low influence by Gaussian variables while maintaining the strategy value. Let

$$\mathbf{P}^{(\mathrm{Apprx})} = \sum_{\sigma:|\sigma|\le d} \widehat{P^{(\mathrm{Smooth})}}(\sigma)\,\mathcal{P}_{\sigma_1}\otimes\mathcal{P}_{\sigma_2}\otimes\ldots\mathcal{P}_{\sigma_{|H|}}\otimes\mathbf{z}_{\sigma_{|H|+1}}^{(|H|+1)}\mathbb{1}_2\otimes\mathbf{z}_{\sigma_{|H|+2}}^{(|H|+2)}\mathbb{1}_2\otimes\ldots\otimes\mathbf{z}_{\sigma_n}^{(n)}\mathbb{1}_2,$$

where $\mathbb{1}_2$ is a 2×2 identity matrix; $\left\{\mathbf{z}_j^{(i)}\right\}_{|H|+1\le i\le n, 1\le j\le 3}$ are independent Gaussian variables and $\mathbf{z}_0^{(|H|+1)} = \ldots \mathbf{z}_0^{(n)} = 1$. Denote the new strategy by $S^{(\mathrm{Apprx})}$, then

$$\mathrm{val}^*(G, S^{(\mathrm{Apprx})}) \approx \mathrm{val}^*(G, S^{(\mathrm{Smooth})}).$$

Notice that this process significantly reduces the dimension of $\mathbf{P}^{(\text{Apprx})}$ to a constant. To round such a randomized strategy back to a valid POVM strategy, we first need to reduce the number of Gaussian variables from $O(n)$ to a constant, which is the most difficult step. In this paper, we avoid the use of a crude union bound as in [46], by taking the distribution of the questions into account. Furthermore, we manage to ensure that the expectation of the distance from a random operator in the intermediate step to positive matrices after the dimension reduction step is independent of the question size. Then the inverse of the invariance principle allows us to round the randomized strategy back to a valid POVM strategy only acting on constantly many qubits. The improvements in the Gaussian dimension reduction step give us the improved bound.

This upper bound has already yielded an NEXP algorithm, where the certificate is an exponential-sized description of the strategy. To design a more efficient nondeterministic algorithm, we need to further compress the certificate to polynomial length. To compress the certificate, we first smoothen again the strategy by introducing additional noise as in the proof of the upper bound of D to remove all the high-degree terms. Such a transformation exponentially reduces the length of the certificate. The smoothed strategy only contains a polynomial number of coefficients since the maximal degree is a constant. Nonetheless, the smoothed strategy is only a *pseudo-strategy*, probably not a valid strategy because these smoothed operators may not form valid POVMs. The prover sends the description of a pseudo-strategy to the verifier, which is of polynomial length. The verifier performs a test on the given certificate to see if it is close to a valid strategy that gives a high winning probability with the following steps:

1. Check that the pseudo-POVM elements contained in the pseudo-strategy still sum up to the identity.
2. Compute and check the winning probability of the pseudo-strategy.
3. Check that all the operators in the pseudo-strategy are close to being positive semidefinite.

Item 1 is straightforward. For item 2, notice that $\operatorname{Tr}\left(\mathcal{P}_i \otimes \mathcal{P}_j\right)\psi = \delta_{i,j}c_i$, where $c_0 = 1$ and $c_1 = c_2 = c_3 = \rho$. Thus for any degree-d operators A, B, we have

$$\operatorname{Tr}\left(A \otimes B\right)\psi^{\otimes D} = \sum_{\sigma:|\sigma|\leq d} \widehat{A}\left(\sigma\right)\widehat{B}\left(\sigma\right)c_\sigma, \tag{1}$$

where $c_\sigma = c_{\sigma_1}\cdots c_{\sigma_n}$. This computation can be done in polynomial time. The winning probability is simply a linear combination of a polynomial number of the terms in the form of Eq.(1), which, therefore, can also be computed in polynomial time. Item 3 is the most challenging. Notice that the dimension of each operator in the pseudo-strategy is still exponential. Thus, the verifier cannot directly compute its eigenvalues and check its positivity. Instead, we need an efficient positivity tester for large matrices.

The key component of our efficient positivity tester is a *derandomized invariance principle*, which enables us to further reduce the dimension of the operators to a constant and maintain the distance between the operator and the set of positive operators. To be more specific, let us define the real function ζ to be

$$\zeta\left(x\right) = \begin{cases} x^2 & \text{if } x \leq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Then $\operatorname{Tr}\,\zeta(P)$ is the distance from P to its positive part. As before, when the degree of an operator is bounded by a constant d, the number of quantum registers having influence that exceeds a given small constant τ is at most d/τ, which is also a constant. To further reduce the dimension of the operators, we prove a more general invariance principle for all

smooth functions compared with the one in [46]. It states that if all non-identity Pauli bases in the registers with low influence are substituted by Rademacher variables or Gaussian variables, the expectation of the distance to the set of positive semidefinite matrices is almost unchanged. We replace all such registers with Rademacher variables, which significantly reduces the dimension of a constant-degree operator to a constant, making it possible to compute its expected ζ function value efficiently. However, the invariance principle introduces $\mathrm{poly}(s)$-many random variables, where s is the size of the question sets. This only leads to a randomized positivity tester. To reduce the randomness, we further apply the well-known Meka-Zuckerman pseudorandom generator [36] to obtain a derandomized invariance principle, which only uses a logarithmic number of independent bits to simulate these variables[2]. This gives a deterministic algorithm to approximately compute the expected ζ function values of all the measurement operators .

To prove the approximation problem is NP-hard, we can compile any $\mathrm{MIP}[\log, O(1)]$ protocol for 3-SAT into a family of noisy nonlocal games one for each 3-SAT instance such that if a 3-SAT instance is satisfiable, the corresponding game has value 1 and if not, the value of the corresponding game is below some constant. In the compiled nonlocal game, the verifier checks with equal probability, if the provers can give consistent answers for the same question or if the provers can give valid answers for queries of their assignment of the instance. Using Fourier analysis, we show that when the provers share noisy MESs, winning the consistency checks with high probability implies that their strategy is essentially deterministic. Then we can relate the classical completeness and soundness of the MIP protocol to the values of the noisy nonlocal games.

1.1.2 Hardness of Noiseless $\mathrm{MIP}^*[\mathrm{poly}, O(1)]$

To show hardness of $\mathrm{MIP}^*[\mathrm{poly}, O(1)]$, we start from the known result $\mathrm{MIP}^*[\mathrm{poly}, \mathrm{poly}] = \mathrm{RE}$ [28], and apply an *answer reduction* transformation to the protocol to get answer length $O(1)$. Answer reduction is essentially PCP composition adapted to the MIP^* setting, and was already an essential component in [39] and [28]. Intuitively, the idea of answer reduction is to ask the two provers in an MIP^* protocol to compute a PCP proof that their answers satisfy the verifier's predicate. The verifier will check this proof rather than checking the answers directly. In order to instantiate this, one requires a PCP of proximity that remains sound when implemented as a two-player quantum game. Showing this soundness condition is technically challenging and usually involves showing that the local tester for a locally testable code, when converted to a two-prover game, is sound against entangled provers. In [28], the code that was used was the Reed-Muller code, which has superconstant alphabet size, ultimately yielding poly-sized answers.

In order to obtain $O(1)$-sized answers, we use the Hadamard code, which is a locally testable code over the binary alphabet. Fortunately for us, it is known that the local tester for this code is "quantum sound" [26, 38]. Moreover, the answer-reduction protocol in [39] is *modular*: it was shown in that work that *any* code with sufficiently good parameters and a quantum-sound tester can be combined with an off-the-shelf PCP of proximity to achieve answer reduction. Our main challenge is to show that the Hadamard code (or a slight variant of it) has a tester meeting the conditions of this theorem. Our new tester for the Hadamard code allows us to reduce the answer length from poly to $O(1)$ directly.

[2] An alternate approach is using Gaussian variables and derandomizing the Gaussian variables as in [30], which discretizes the Gaussian variables via the Box-Muller transformation and further derandomizes the discrete random variables.

1.2 Technical Contributions

1.2.1 Invariance Principle and Derandomized Invariance Principle for Matrix Functions

The invariance principle [37] is a generalization of the Berry-Esseen Theorem, which is a quantitative version of the Central Limit Theorem, to multilinear low-degree polynomials. Before illustrating the invariance principle, we need to introduce the notion of *influence*, a fundamental notion in the analysis of Boolean functions. Given a real function $f : \mathbb{R}^n \to \mathbb{R}$ and i.i.d. random variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the influence of i-th coordinate is

$$\mathrm{Inf}_i(f) = \mathbb{E}\left[\left|f(\mathbf{x}) - f\left(\mathbf{x}^{(i)}\right)\right|^2\right],$$

where $\mathbf{x}^{(i)}$ is obtained from $\mathbf{x}$ by resampling the i-th variable. Hence, it captures the effect of the i-th variable on the function in average. Given a multilinear low-degree polynomial f in which all variables have low influence, the invariance principle states that the distributions of $f(X_1, \ldots, X_n)$ and $f(Y_1, \ldots, Y_n)$ are similar as long as the first and second moments of the random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ match, and the variables X_i, Y_i behave nicely[3]. The invariance principle is a versatile tool that allows us to connect the distribution of a function on complicated random variables to the distribution obtained by replacing these random variables with simpler ones, such as Gaussian variables or Rademacher random variables. The proof of the classical invariance principle in [37] is via Lindeberg's hybrid argument, which is also a classic method to prove the Central Limit Theorem.

In [46], Qin and Yao started investigating the invariance principle on matrix spaces. Suppose that P is a $m^n \times m^n$ matrix, viewed as an operator acting on n registers, each of dimension m. Let $\xi : \mathbb{R} \to \mathbb{R}$ be a smooth real function. Suppose all registers have low influence in P, where the influence is a generalization of the influence for functions. When substituting all registers with independent standard Gaussians or Rademacher variables multiplied by an identity matrix, we expect that the change of $\mathrm{Tr}\, \xi(P)$ is small in expectation. The most challenging part of extending Lindeberg's argument to matrix functions is computing the high-order Fréchet derivatives, which are complicated and difficult to analyze in general [50]. Qin and Yao [46] established an invariance principle for a specific spectral function by directly computing the Fréchet derivatives and applying many complicated matrix-analytic techniques. Hence, the first obstacle we face is to prove an invariance principle for more general functions.

To overcome it, we adapt the theory of multilinear operator integrals [52], which provides a unified way to compute and bound the Fréchet derivatives. With such a tool, we establish an invariance principle applicable to a broader class of functions, including those that are smooth with a bounded third derivative and those that are Lipschitz continuous.

The invariance principle reduces the dimension from poly to constant but introduces a poly number of independent random variables. Thus, the second obstacle is that the size of the overall probability space is exponential. To improve the computational efficiency of our invariance principle, we use the ideas of [36, 22, 44] to use a Pseudorandom generator (PRG) to reduce the number of independent random variables. We apply this derandomized invariance principle to our positivity tester introduced below. Derandomized invariance principles build

[3] To be more specific, $\mathbf{x}_i, \mathbf{y}_i$ need to be hypercontractive. Informally speaking, the p-norms $\|\mathbf{x}_i\|_p = \mathbb{E}\left[|\mathbf{x}_i|^p\right]^{1/p}$ $\|\mathbf{y}_i\|_p = \mathbb{E}\left[|\mathbf{y}_i|^p\right]^{1/p}$ do not increase drastically with respect to p. Many basic random variables, such as uniformly random variables, and Gaussian variables, are hypercontractive.

upon the crucial observation that the highest moment of variables involved in the proof is at most $2d$, where d is the degree of the operator, which is a constant. Thus, it suffices to use $4d$-wise uniform random variables instead of polynomially many independent random variables when we replace the Pauli basis elements in the low-influence registers, which saves the randomness exponentially. To this end, we employ the well-known Meka-Zuckerman pseudorandom generator [36] to construct $4d$-wise uniform random variables.

As the invariance principle has found numerous applications, we anticipate that the invariance principle for spectral functions is interesting in its own right. The positivity testing for low-degree matrices introduced below is an example of its applications.

1.2.2 Positivity Tester for Low-degree Matrices

A Hermitian matrix A is said to be positive semidefinite (PSD) if all the eigenvalues of A are non-negative. This testing problem has received increasing attention in the past couple of years [35, 21, 8, 41]. In this work, we present an efficient PSD tester for low-degree matrices, where the input matrix is given in terms of its Fourier coefficients. Given an $m^n \times m^n$ matrix, viewed as an operator acting on n-qudits, each of which has dimension m, if the degree of the operator is d, then the number of Fourier coefficients is bounded by $\sum_{i\le d}\binom{n}{i}(m^2-1)^i = O(dn^d m^{2d})$. Hence, this allows for a compact description of a low-degree, exponential-dimension operator. If m, d are constants, the input is of size $\mathrm{poly}(n)$, and we work in this setting when we explain how the tester works below.

Given the Fourier coefficients of a matrix P, our tester estimates the distance between P and the set of positive semidefinite matrices measured by $\mathrm{Tr}\zeta(P)$, where $\zeta(\cdot)$ is defined in Eq. (2). Estimating $\mathrm{Tr}\zeta(P)$ involves applying the derandomized invariance principle introduced above. More specifically, our tester enumerates all the possible seeds of the Meka-Zuckerman PRG to estimate this distance. For each seed, the computation time is $O(1)$ because the derandomized invariance principle has effectively reduced the dimension of P to a constant. Hence, our tester runs in time $\mathrm{poly}(n)$, because there are only $\mathrm{poly}(n)$ seeds. Its guarantees are summarized below.

▶ **Theorem** (informal). *Given as input the Fourier coefficients of a degree-d operator P acting on n qudits, each of dimension m, and error parameters $\beta \ge \delta \ge 0$, there exists an algorithm that runs in time $\exp(m^d/\delta)\cdot\mathrm{poly}(n)$ such that*

- *the algorithm accepts if there exists a PSD operator Q such that $\|P-Q\|_F^2 < (\beta-\delta)\,m^n$;*
- *the algorithm rejects if $\|P-Q\|_F^2 > (\beta+\delta)m^n$ for any PSD operator Q.*

This approach completely differs from all previous works on positivity testing [41, 21, 8], where they only consider polynomial-sized matrices and the testers are randomized. In contrast, our tester is deterministic, and the dimension of the testing matrix can be exponential in input size if the degree is constant.

1.2.3 Answer Reduction with the Hadamard Code

As mentioned above, we obtain $O(1)$-sized answers in the noiseless setting by applying the code-based answer reduction of [39], with the code chosen to be the Hadamard code. To implement this required two new technical components. First, we showed a *quantum-sound subset tester* for the Hadamard code: essentially, an interactive protocol that forces the provers to respond with the values of a subset F of the coordinates of a Hadamard codeword, where F is sampled from some (not necessarily uniform) distribution. Our proof of this result is essentially a generalization of the Fourier-analytic proof of the quantum soundness

of the BLR test [10, 38]. Secondly, the answer reduction procedure in [39] only works if the code has a relative distance close to 1 (i.e., distinct codewords differ on almost all locations), whereas the Hadamard code has a distance 1/2. To overcome this, we slightly modified the answer-reduced verifier's protocol of [39] by querying a large constant number of "dummy coordinates" from the provers. It is worth mentioning that the answer reduction procedure from [39] is different from the procedure used in [28]; the former works for any error-correcting code satisfying certain properties but does not yield protocols that can be recursively compressed, whereas the latter is specialized to the low-degree code but is compatible with recursive compression. In the end, we are in effect using both versions of answer reduction: the [28] version inside the recursive compression to obtain a protocol for RE, and then one layer of the [39] version to bring the answer size successively down from polynomial to constant, using the Hadamard code.

We remark that it might be possible to achieve constant answer size by repeatedly applying the answer reduction technique of [28], but we decide to proceed with the current approach for a one-shot solution, which is easier to analyze and gives better soundness.

1.3 Discussions and Open Problems

Our result characterizes the effect of noise on the computational complexity class MIP*. To our knowledge, this is the first example of a quantum computational complexity class whose quantum advantage over its classical counterpart completely vanishes in the presence of noise. For comparison, noise causes *no* collapse in the BQP model, or in general, for BQTIME because the algorithms in these classes can be implemented fault-tolerantly. Even for algorithms with bounded space, it seems that the same reasoning still applies because all the intermediate measurements to achieve fault tolerance can be eliminated without a large space overhead [18]. Hence, our work raises the natural question of which quantum complexity classes are truly fault tolerant. In contrast, for complexiy classes like MIP*, the fault-tolerance theorem [1] cannot be applied as the model of computation disallows the operations needed to perform error correction. For the specific case of MIP*, our result further shows that no form of fault tolerance is possible.

More broadly, we know other examples where constant noise destroys the quantum advantage. Random circuit sampling has been proposed to demonstrate the quantum advantage offered by near-term quantum devices [12]. However, when the random circuits are subject to constant noise, this sampling task becomes classically easy [2]. We have more of such examples in quantum query algorithms. For example, if the oracle is noisy or faulty, no quantum algorithm can achieve any speed-up in the unstructured search problem [48]. In a setting closer to the near-term devices, where each gate in the circuit is subject to independent noise but the oracle is perfect, the authors of [14] showed that no quantum algorithm could achieve any speed-up in the unstructured search problem either. For a more detailed survey about the effect of noise on quantum query algorithms, we refer to [14, Section 3].

Our result also raises some natural but intriguing questions. We list some of them below.

1. For MIP* protocols with more rounds of interactions and larger answer sets, it is unclear how big the effect of noise is. Hence, we ask: Does the vanishing phenomenon for computational advantages occur for general MIP* protocols?
2. What is the computational power of MIP* with unbounded copies of a pure (noiseless) non-EPR state? Will MIP* = RE still hold for any noiseless non-EPR state? The MIP* protocol for RE of [28] requires EPR states for the provers to succeed, and in general, it is known that any protocol which is symmetric and synchronous requires the provers

to use an MES [56, 45]. Moreover, the question reduction technique of [39, 28] first certifies that the provers share many copies of the EPR state, on which they sample their own questions. Hence, to accommodate any non-EPR state, we need to redesign MIP* protocols.

3. What non-computational capabilities of the MIP* model remain in the noisy setting? Specifically, it is known that nonlocal games and correlations can be used to self-test entangled states. In the noisy setting, can we certify any properties of the provers' shared entanglement? Previous work on this question has studied entanglement of formation [5] and one-shot distillable entanglement [3], but the general picture remains unclear.
4. Invariance principle has found applications in designing various pseudorandom generators and counting algorithms [22, 44, 43, 6, 31]. Will our invariance principle lead to new pseudorandom generators?
5. Testing whether a matrix is positive has played an important role in the study of algorithm designs for linear algebra problems, community structure detection, differential equations, etc (see [8] and references therein). Multiple studies have been devoted to designing efficient algorithms for positivity testing [41, 21, 8]. Will our algorithm of positivity testing find new applications?

2 Nonlocal Games and MIP* Protocols

In this paper we use the standard notations for matrix spaces, random variables etc. For a detailed description see Appendix A. Two-player one-round MIP* protocols are also nonlocal games. We follow the notations of [28] for nonlocal games.

▶ **Definition 4** (Two-player one-round games). *A two-player one-round game G is specified by a tuple $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \mu, V)$ where*

- *$\mathcal{X}$ and $\mathcal{Y}$ are finite sets, called the* question sets,
- *$\mathcal{A}$ and $\mathcal{B}$ are finite sets, called the* answer sets,
- *μ is a probability distribution over $\mathcal{X} \times \mathcal{Y}$, called the* question distribution, *and*
- *$V : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B} \to \{0, 1\}$ is a function, called the* decision predicate.

▶ **Definition 5** (Tensor-product strategies). *A tensor-product strategy S of a nonlocal game $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \mu, V)$ is a tuple (ψ, A, B) where*

- *a bipartite quantum state $\psi \in \mathcal{H}_A \otimes \mathcal{H}_B$ for finite dimensional complex Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$,*
- *A is a set $\{A^x\}$ such that for every $x \in \mathcal{X}$, $A^x = \{A^x_a \mid a \in \mathcal{A}\}$ is a POVM over $\mathcal{H}_A$, and*
- *B is a set $\{B^y\}$ such that for every $y \in \mathcal{Y}$, $B^y = \{B^y_b \mid b \in \mathcal{B}\}$ is a POVM over $\mathcal{H}_B$.*

▶ **Definition 6** (Tensor product value). *The tensor product value of a tensor product strategy $S = (\psi, A, B)$ for a nonlocal game $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \mu, V)$ is defined as*

$$\mathrm{val}^*(G, S) = \sum_{x,y,a,b} \mu(x, y) V(x, y, a, b) \mathrm{Tr}\left(A^x_a \otimes B^y_b\right) \psi.$$

For $v \in [0, 1]$ we say that the strategy passes or wins G with probability v if $\mathrm{val}^*(G, S) \geq v$. The quantum value or tensor product value of G is defined as

$$\mathrm{val}^*(G) = \sup_S \mathrm{val}^*(G, S)$$

where the supremum is taken over all tensor product strategies S for G.

When we prove the quantum soundness of an MIP* protocol, we focus on projective strategies, where the measurements A^x and B^y are all projective, following Naimark's Dilation theorem [29, Theorem 5.1].

▶ **Definition 7.** *A game $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \mu, V)$ is symmetric if $\mathcal{X} = \mathcal{Y}$ and $\mathcal{A} = \mathcal{B}$, the distribution μ is symmetric (i.e. $\mu(x, y) = \mu(y, x)$ for all x and y), and the predicate V treats both players symmetrically (i.e. $V(x, y, a, b) = V(y, x, b, a)$ for all x, y, a, b).*

We call a strategy $S = (|\psi\rangle, A, B)$ symmetric if $|\psi\rangle$ is a pure state in $\mathcal{H} \otimes \mathcal{H}$, for some Hilbert space $\mathcal{H}$, that is invariant under permutation of the two factors, and the measurement operators of both players are identical.

A symmetric game is denoted by $(\mathcal{X}, \mathcal{A}, \mu, V)$, and a symmetric strategy is denoted by $(|\psi\rangle, M)$ where M denotes the set of measurement operators for both players.

▶ **Lemma 8** (Lemma 5.7 in [28])**.** *Let $G = (\mathcal{X}, \mathcal{A}, \mu, V)$ be a symmetric game with value $1 - \varepsilon$ for some $\varepsilon \geq 0$. Then there exists a symmetric and projective strategy $S = (|\psi\rangle, M)$ such that the $\mathrm{val}^*(G, S) \geq 1 - \varepsilon$.*

Hence, for symmetric nonlocal games, it suffices to only consider symmetric strategies.

3 Invariance Principle for Matrix Spaces

This section will present an invariance principle for general functions on matrix spaces. Hypercontractivity is crucial in the proofs of all previous invariance principles [37]. We also need to establish a new hypercontractive inequality before proving the invariance principle. The proofs of the results in this section can be found in Appendix B.1.

3.1 Hypercontractivity

In this subsection, we adopt the concept of orthonormal ensembles as introduced in [37].

▶ **Definition 9.** *Given $m, n \in \mathbb{Z}_{>0}$, a collection of n real random variables $\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ are orthonormal if $\mathbb{E}[\mathbf{z}_i \mathbf{z}_j] = \delta_{i,j}$. We call a collection of m orthonormal real random variables, the first of which is constant 1, an m-orthonormal ensemble. We call $\mathbf{x}$ an (m, n) ensemble if $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, where for all $i \in [n]$, $\mathbf{x}_i = \{\mathbf{x}_{i,0} = 1, \mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,m-1}\}$ is an m-orthonormal ensemble.*

▶ **Definition 10.** *Given $m, n \in \mathbb{Z}_{>0}$, $\tau \in [m]^n_{\geq 0}$ and an (m, n) ensemble $\mathbf{x}$, denote $\mathbf{x}_\tau = \prod_{i=1}^n \mathbf{x}_{i,\tau_i}$. Define a multilinear polynomial over $\mathbf{x}$ to be $Q(\mathbf{x}) = \sum_{\tau \in [m]^n_{\geq 0}} \widehat{Q}(\tau) \mathbf{x}_\tau$, where the $\widehat{Q}(\tau)$'s are real constants.*

For $\gamma \in [0, 1]$, we define the operator T_γ acting on multilinear polynomial $Q(\mathbf{x})$ by

$$T_\gamma Q(\mathbf{x}) = \sum_{\tau \in [m]^n_{\geq 0}} \gamma^{|\tau|} \widehat{Q}(\tau) \mathbf{x}_\tau.$$

▶ **Definition 11.** *For $1 \leq r < \infty$, let $\mathbf{y}$ be a random variable with $\mathbb{E}[|\mathbf{y}|^r] < \infty$. Define $\|\mathbf{y}\|_r = (\mathbb{E}[|\mathbf{y}|^r])^{1/r}$. Given $1 \leq p \leq q < \infty$, $0 < \eta < 1$, $m, n \in \mathbb{Z}_{>0}$ and an (m, n) ensemble $\mathbf{x}$, we say that $\mathbf{x}$ is (p, q, η) -hypercontractive if for any multilinear polynomial Q, it holds that $\|(T_\eta Q)(\mathbf{x})\|_q \leq \|Q(\mathbf{x})\|_p$.*

Consider an (m,n) ensemble $\mathbf{x}$. If for all $i \in [n]$, $j \in [m-1]$, $\mathbf{x}_{i,j}$ are either independent standard Gaussians or independent Rademacher variables, then $\mathbf{x}$ is $(2, q, (q-1)^{-1/2})$-hypercontractive. These two types are represented as significant examples of hypercontractive ensembles. Readers can refer to [37] for an extensive treatment on hypercontractive ensembles.

We then introduce the noise operator Γ_γ for random matrices, which is a hybrid of T_γ in Definition 10 and Δ_γ in Definition 47.

▶ **Definition 12.** *Given $0 \le \gamma \le 1$, $h, n, m \in \mathbb{Z}_{>0}$, $m \ge 2$, an (m^2, n) ensemble $\mathbf{x}$, and a random matrix $P(\mathbf{x}) = \sum_{\sigma \in [m^2]^h_{\ge 0}} p_\sigma(\mathbf{x}) \mathcal{B}_\sigma$, where $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$ is a standard orthonormal basis and p_σ is a real multilinear polynomial for all $\sigma \in \left[m^2\right]^h_{\ge 0}$, the noise operator Γ_γ is defined to be*

$$\Gamma_\gamma(P(\mathbf{x})) = \sum_{\sigma \in [m^2]^h_{\ge 0}} (T_\gamma p_\sigma)(\mathbf{x}) \Delta_\gamma(\mathcal{B}_\sigma).$$

The main result in this subsection is stated below.

▶ **Theorem 13** (Hypercontractivity for random matrices)**.** *Given $h, n, m \in \mathbb{Z}_{>0}$, $m \ge 2$, $0 < \eta < 1$, $0 \le \gamma \le \min\left\{\eta, (9m)^{-1/4}\right\}$, a $(2,4,\eta)$-hypercontractive (m^2, n) ensemble $\mathbf{x}$ and a random matrix $P(\mathbf{x}) = \sum_{\sigma \in [m^2]^h_{\ge 0}} p_\sigma(\mathbf{x}) \mathcal{B}_\sigma$, where $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$ is a standard orthonormal basis, and p_σ is a real multilinear polynomial for all $\sigma \in [m^2]^h_{\ge 0}$, it holds that*

$$\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|\left|\Gamma_\gamma(P(\mathbf{x}))\right|\right\|_4^4\right] \le \left(\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|\left|P(\mathbf{x})\right|\right\|_2^2\right]\right)^2,$$

where Γ_γ is defined in Definition 12.

The following is an application of Theorem 13.

▶ **Theorem 14.** *Given $h, n, m, d \in \mathbb{Z}_{>0}$, $m \ge 2$, $0 < \eta < 1$, a $(2,4,\eta)$-hypercontractive (m^2, n) ensemble $\mathbf{x}$, and a random matrix $P(\mathbf{x}) = \sum_{\sigma \in [m^2]^h_{\ge 0}} p_\sigma(\mathbf{x}) \mathcal{B}_\sigma$, where $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$ is a standard orthonormal basis and for all $\sigma \in \left[m^2\right]^h_{\ge 0}$ and p_σ is a real multilinear polynomial satisfying $\deg(p_\sigma) + |\sigma| \le d$, it holds that*

$$\mathbb{E}\left[\left\|\left|P(\mathbf{x})\right|\right\|_4^4\right] \le \max\left\{9m, 1/\eta^4\right\}^d \left(\mathbb{E}\left[\left\|\left|P(\mathbf{x})\right|\right\|_2^2\right]\right)^2.$$

3.2 Invariance Principle

We are now prepared to introduce an invariance principle on matrix space applicable to general functions. Initially, we establish the proof for functions in $\mathcal{C}^4$.

▶ **Theorem 15.** *Given $0 < \tau, \eta < 1$, $d, h, m, n \in \mathbb{Z}_{>0}$, $H \subseteq [n]$ of size $|H| = h$, $\xi \in \mathcal{C}^3$ satisfying $\|\xi^{(3)}\|_\infty \le B$ where B is a constant, and a $(2,4,\eta)$-hypercontractive (m^2, n) ensemble $\mathbf{x}$, let $P \in \mathcal{H}_m^{\otimes n}$ be a degree-d operator satisfying $\mathrm{Inf}_i(P) \le \tau$ for all $i \notin H$. Suppose that P has a Fourier expansion $P = \sum_{\sigma \in [m^2]^n_{\ge 0}} \widehat{P}(\sigma) \mathcal{B}_\sigma$. Let $P^H(\mathbf{x}) = \sum_{\sigma \in [m^2]^n_{\ge 0}} \widehat{P}(\sigma) \mathbf{x}_{\sigma_{\overline{H}}} \mathcal{B}_{\sigma_H}$. If $\sum_{\sigma \ne 0} \widehat{P}(\sigma)^2 \le 1$, we have*

$$\left|m^{-n}\mathrm{Tr}\,\xi(P) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\,\xi\left(P^H(\mathbf{x})\right)\right]\right| \le CB \max\left\{9m, 1/\eta^4\right\}^d \sqrt{\tau} d$$

for some absolute constant C.

For those functions that are not sufficiently smooth, if they have a mollifier, which is a smooth approximator with a bounded third derivative, then the invariance principle still holds. The following lemma proves an invariance principle for $\zeta(\cdot)$ defined in Appendix A.2.4, which has a mollifier $\zeta_\lambda(\cdot)$ guaranteed by Fact 57.

▶ **Lemma 16.** *Given $0 < \tau, \eta < 1$, $d, h, m, n \in \mathbb{Z}_{>0}$, $H \subseteq [n]$ of size $|H| = h$, a $(2, 4, \eta)$-hypercontractive (m^2, n) ensemble $\mathbf{x}$ and a degree-d $P \in \mathcal{H}_m^{\otimes n}$ satisfying $\mathrm{Inf}_i(P) \leq \tau$ for all $i \notin H$. suppose that P has a Fourier expansion $P = \sum_{\sigma \in [m^2]^n_{\geq 0}} \widehat{P}(\sigma) \mathcal{B}_\sigma$. Let $P^H(\mathbf{x}) = \sum_{\sigma \in [m^2]^n_{\geq 0}} \widehat{P}(\sigma) \mathbf{x}_{\sigma_{\overline{H}}} \mathcal{B}_{\sigma_H}$. If $\sum_{\sigma \neq 0} \widehat{P}(\sigma)^2 \leq 1$, we have*

$$\left| m^{-n} \mathrm{Tr}\, \zeta(P) - m^{-h}\, \mathbb{E}\left[\mathrm{Tr}\, \zeta\left(P^H(\mathbf{x})\right)\right] \right| \leq 3 \left(C B_3 \max\left\{9m, 1/\eta^4\right\}^d \sqrt{\tau} d \right)^{2/3}$$

for some universal constants C and B_3.

▶ Remark 17. It is possible to prove an invariance principle for a broader class of functions. For example, we can prove it for Lipschitz continuous functions using the argument in [24, Lemma 3.5]. However, it is out of the focus of this paper. We will leave it for further research.

3.3 Derandomized Invariance Principle

From Theorem 15, it is not hard to see that the non-identity basis elements can be substituted by independent Rademacher variables. In this section, we will replace those Rademacher variables with pseudorandom variables to save the randomness. It is worth noting that there is a large body of research on derandomization through invariance principles (readers may refer to[44] and the references therein). We adopt the pseudorandom generator (PRG) introduced in [36]. The PRG is constructed by pairwise uniform hash functions as follows.

For $\mathcal{F} = \{f : [n] \to [p]\}$, define $G : \mathcal{F} \times (\{-1,1\}^n)^p \to \{-1,1\}^n$ by

$$G\left(f, z^1, \ldots, z^p\right) = x, \text{ where } x_i = z_i^{f(i)} \text{ for } i \in [n]. \tag{3}$$

We define the influence of a random variable in a random matrix using the notation $\mathrm{VarInf}(\cdot)$ to distinguish from the notation for the influence of a register in Definition 38.

▶ **Definition 18.** *Given $n, p \in \mathbb{Z}_{>0}$, let $P(\mathbf{b}) = \sum_{S \subseteq [n]} \mathbf{b}_S P_S$ be a random matrix with $\mathbf{b}$ being drawn uniformly from $\{\pm 1\}^n$ and $\mathbf{b}_S = \prod_{i \in S} \mathbf{b}_i$. Then the influence of i'th coordinate of $\mathbf{b}$ is defined to be*

$$\mathrm{VarInf}_i(P(\mathbf{b})) = \sum_{S \ni i} |||P_S|||_2^2.$$

We also define the influence of a block of coordinates. Let $j \in [p]$ and $f : [n] \to [p]$ be a function, define the influence on the block $f^{-1}(j) \subseteq [n]$ to be

$$\mathrm{VarInf}_{f,j}(P(\mathbf{b})) = \sum_{S : S \cap f^{-1}(j) \neq \emptyset} |||P_S|||_2^2.$$

The following is the main theorem in this section.

▶ **Theorem 19** (Derandomized invariance principle for ζ)**.** *Given $d, h, m, n \in \mathbb{Z}_{>0}$, $m > 1$, and a random matrix $P(\mathbf{b}) = \sum_{S \subseteq [n]} \mathbf{b}_S P_S$, where $\mathbf{b} \sim_{\mathrm{u}} \{-1,1\}^n$, $\mathbb{E}_{\mathbf{b}}\left[|||P(\mathbf{b})|||_2^2\right] \leq 1$, $\mathbf{b}_S = \prod_{i \in S} \mathbf{b}_i$ and $P_S \in \mathcal{H}_m^{\otimes h}$, they satisfy $|S| + \deg(P_S) \leq d$ and $\mathrm{VarInf}_i(P(\mathbf{b})) \leq \tau$ for all $i \in [n]$.*

Let p be the smallest power of 2 satisfying $p \geq d/\tau$; $\mathcal{F} = \{f : [n] \to [p]\}$ be a family of pairwise uniform hash functions. For any $i \in [p]$, define $\mathbf{z}^i$ to be a $4d$-wise uniform random vector drawn from $\{\pm 1\}^n$, and $\mathbf{z}^i$ are independent across $i \in [p]$. Given $f \in \mathcal{F}$, denote $\mathbf{x}_f = G\left(f, \mathbf{z}^1, \ldots, \mathbf{z}^p\right)$ as in Equation (3). Then it holds that

$$\left| \frac{1}{m^h} \mathop{\mathbb{E}}_{\mathbf{b}} [\operatorname{Tr} \zeta(P(\mathbf{b}))] - \frac{1}{m^h} \mathop{\mathbb{E}}_{\mathbf{f}, \mathbf{x_f}} [\operatorname{Tr} \zeta(\mathbf{P}(\mathbf{x_f}))] \right| \leq C\sqrt{(9m)^d d\tau},$$

where $\mathbf{f}$ is drawn uniformly from $\mathcal{F}$ and C is a universal constant.

We first prove a derandomized invariance principle for the functions with bounded fourth derivative.

▶ **Theorem 20** (Derandomized invariance principle). *Given $d, h, m, n \in \mathbb{Z}_{>0}$, $m > 1$, and a random matrix $P(\mathbf{b}) = \sum_{S \subseteq [n]} \mathbf{b}_S P_S$, where $\mathbf{b} \sim_{\mathrm{u}} \{-1, 1\}^n$, $\mathbb{E}_{\mathbf{b}} \left[\|\|P(\mathbf{b})\|\|_2^2 \right] \leq 1$, $\mathbf{b}_S = \prod_{i \in S} \mathbf{b}_i$ and $P_S \in \mathcal{H}_m^{\otimes h}$, they satisfy that $|S| + \deg(P_S) \leq d$ and $\operatorname{VarInf}_i(P(\mathbf{b})) \leq \tau$ for all $i \in [n]$.*

Let p be the smallest power of 2 satisfying $p \geq d/\tau$; $\mathcal{F} = \{f : [n] \to [p]\}$ be a family of pairwise uniform hash functions. For any $i \in [p]$, define $\mathbf{z}^i$ to be a $4d$-wise uniform random vector drawn from $\{\pm 1\}^n$, and $\mathbf{z}^i$ are independent across $i \in [p]$. Given $f \in \mathcal{F}$, denote $\mathbf{x}_f = G\left(f, \mathbf{z}^1, \ldots, \mathbf{z}^p\right)$ as in Equation (3). Then for any $\xi \in \mathcal{C}^4$ with $\|\xi^{(4)}\|_\infty \leq C_0$ where C_0 is a constant, it holds that

$$\left| \frac{1}{m^h} \mathop{\mathbb{E}}_{\mathbf{b}} [\operatorname{Tr} \xi(P(\mathbf{b}))] - \frac{1}{m^h} \mathop{\mathbb{E}}_{\mathbf{f}, \mathbf{x_f}} [\operatorname{Tr} \xi(\mathbf{P}(\mathbf{x_f}))] \right| \leq 4C_1 C_0 (9m)^d d\tau,$$

where $\mathbf{f}$ is drawn uniformly from $\mathcal{F}$ and C_1 is a universal constant.

▶ Remark 21. It is also possible to generalize Theorem 19 to Lipschitz continuous functions using the argument in [24, Lemma 3.5].

Assuming Theorem 20, Theorem 19 is straightforward.

4 Positivity Tester for Low Degree Operators

In this section, we will present an algorithm deciding whether a low-degree operator is $(\beta - \delta)$-close to a positive semidefinite matrix or $(\beta + \delta)$-far from all positive semidefinite matrices, for error parameters $\beta > \delta > 0$. The input operator is given in the form of a Fourier expansion. The algorithm and the proofs can be found in Appendix B.2.

▶ **Definition 22** (Positivity testing problem). *Given $d, D, m \in \mathbb{Z}_{>0}$, $m > 1$, and real numbers $\beta > \delta > 0$, the input is a degree-d operator in $\mathcal{H}_m^{\otimes D}$ given in the form of Fourier expansion*

$$P = \sum_{\substack{\sigma \in [m^2]^D_{\geq 0} \\ \sigma : |\sigma| \leq d}} \widehat{P}(\sigma) \mathcal{B}_\sigma.$$

Distinguish the following two cases.

- *Yes: if $m^{-D} \operatorname{Tr} \zeta(P) < \beta - \delta$.*
- *No: if $m^{-D} \operatorname{Tr} \zeta(P) > \beta + \delta$.*

Notice that the number of Fourier coefficients is $\sum_{i=0}^{d} \binom{D}{i} \left(m^2 - 1\right)^i$. If we are concerned with constant-degree operators, then the dimension of the operator is exponential in the input size.

▶ **Theorem 23.** *Given $d, D, m \in \mathbb{Z}_{>0}$, $m > 1$, and real numbers $\beta > \delta > 0$, there exists a deterministic algorithm for the positivity testing problem that runs in time*

$$\exp\left(\mathrm{poly}\left(m^d, 1/\delta\right)\right) \cdot D^{O(d)}.$$

In particular, if m, d, δ are constants, then the algorithm runs in time $\mathrm{poly}(D)$.

The algorithm applies the invariance principle Lemma 16 to reduce the dimension of the matrices and then Theorem 19 to derandomize, while the distance to positive operators is approximately preserved.

5 Noisy Nonlocal Games are NP-complete

▶ **Definition 24** (Noisy Nonlocal Game Value Problem)**.** *The input consists of the description of a nonlocal game, which is a tuple $\mathfrak{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \mu, V)$, and real values ρ, β and ε. $\mathcal{X}$ and $\mathcal{Y}$ are question sets and assume $|\mathcal{X}| = |\mathcal{Y}| = s$. $\mathcal{A}$ and $\mathcal{B}$ are answer sets and assume $|\mathcal{A}| = |\mathcal{B}| = t$. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$ and $V : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B} \to \{0, 1\}$ be the predicate.*

Let $v = \mathrm{val}^*(\mathfrak{G}, \psi_{AB})$ be the value of the nonlocal game, where Alice and Bob share arbitrarily many copies of a noisy MES ψ_{AB} with the maximal correlation ρ. Let $1 > \beta > \varepsilon > 0$. The task is to distinguish the following two cases.

- *Yes: $v > \beta + \varepsilon$.*
- *No: $v < \beta - \varepsilon$.*

In this section, we show:

▶ **Theorem 25.** *The noisy nonlocal game value problem is* NP*-complete.*

It follows from the two propositions below, whose proofs can be found in Appendix B.3.

▶ **Proposition 26.** *There exists a nondeterministic algorithm that runs in time*

$$\mathrm{poly}\left(s, \mathrm{eexp}\left(t, \log\left(\frac{1}{\rho}\right), \frac{1}{\varepsilon}\right)\right)$$

that solves the noisy nonlocal game value problem. Here $\mathrm{eexp}(\cdot)$ *means doubly exponential. In particular, if t, ρ, ε are constants, then the problem is in* NP.

▶ **Proposition 27.** *For each 3-SAT instance ϕ, there is a nonlocal game $G(\phi)$ such that its noisy game value is 1 if ϕ is satisfiable, and below some constant c if ϕ is not satisfiable.*

5.1 The Nondeterministic Algorithm

We first present an upper bound on the number of noisy MES sufficient to approximate the value of a nonlocal game to an arbitrary precision. The upper bound from [46] is $D = \exp(\mathrm{poly}(s), \exp(\mathrm{poly}(t)))$. The follow-up work [47] studied fully quantum games in which both questions and answers are quantum and proved a better upper bound $D = \exp(\mathrm{poly}(s), \mathrm{poly}(t))$ using a refined Gaussian dimension reduction. We observe that this upper bound can be further improved to $D = \mathrm{poly}(s, \exp(\mathrm{poly}(t)))$ for nonlocal games.

▶ **Theorem 28.** *Given parameters $0 < \epsilon, \rho < 1$, $n, m \in \mathbb{Z}_{>0}$, $m \geq 2$, a noisy MES state ψ_{AB}, i.e., $\psi_A = \psi_B = \frac{\mathbb{1}_m}{m}$ with the maximal correlation $\rho = \rho(\psi_{AB}) < 1$ as defined in Definition 41, let $\mathfrak{G}$ be a nonlocal game with the question sets $\mathcal{X}, \mathcal{Y}$ and the answer sets $\mathcal{A}, \mathcal{B}$. Suppose the players share arbitrarily many copies of ψ_{AB}. Let $\omega_n(\mathfrak{G}, \psi_{AB})$ be the*

highest winning probability that the players can achieve when sharing n copies of ψ_{AB}. Then there exists an explicitly computable bound $D = D(|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{A}|, |\mathcal{B}|, m, \epsilon, \rho)$, such that for any $n > D$, $\omega_n(\mathfrak{G}, \psi_{AB}) - \omega_D(\mathfrak{G}, \psi_{AB}) \leq \epsilon$. In particular, one may choose

$$D = \text{poly}\left(|\mathcal{X}|, |\mathcal{Y}|, \exp\left(\text{poly}\left(|\mathcal{A}|, |\mathcal{B}|, \frac{1}{\epsilon}, \frac{1}{1-\rho}\right), \log m\right)\right).$$

The proof largely follows the framework in [46] with several refinements.[4]

Next we present the algorithm, which is deterministic provided with a certificate. By Theorem 28 we know that sharing D copies of ψ_{AB} is sufficient to approximate the game value. However, outlining a strategy that shares D copies of ψ_{AB} requires $\exp(D)$ bits, rendering it excessively costly. Despite this, we've devised a more affordable certificate. Interpreted as a degree-d pseudo-strategy, this certificate is presented through its Fourier coefficients. By pseudo-strategy we mean two sets of operators $\{P_a^x\}$ and $\{Q_b^y\}$ that may not be a valid quantum strategy. However, we can still define the winning probability on a pseudo-strategy, mathematically. The algorithm is given in Appendix B.3.

5.2 NP-Hardness

In this subsection, we first show that if $L \in$ MIP then $L \in$ noisy MIP*. Then Proposition 27 directly follows from the fact that 3-SAT $\in$ MIP[log, 1] [7].

▶ **Proposition 29.** *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_V)$ be an* MIP *protocol for a language L with perfect completeness. Then there exists a verifier V^* that is a noisy* MIP* *verifier for L with the following conditions:*

Completeness. *If* input $\in L$*, there is a value-*1 *strategy for V^*.*

Soundsness. *Given* input*, if there is a strategy for V^* with value $1-\epsilon$, then there is a strategy for V with value $1 - 2\varepsilon - \frac{16\epsilon}{1-\rho}$.*

6 MIP* Protocol for RE with $O(1)$-size Answers

In this section, we prove that there is an MIP* protocol for any language in RE with poly-size questions and constant-size answers. The key step is to develop a new answer reduction technique that can reduce the answer size of an MIP* protocol from $O(\log n)$ to $O(1)$ while maintaining other parameters of the protocol. We achieve it by modifying the answer reduction technique from [39]. Natarajan and Wright's answer reduction follows a modular design with two major components: Probabilistically checkable proofs of proximity (PCPP) and a tester of the low-degree code. Hence, to achieve constant answer size, it suffices to change the code to the Hadamard code, and derive a new tester for the Hadamard code that allows a verifier to test multiple bits of a codeword at the same time. Then in our final construction of the MIP* protocol for RE, we apply our new answer reduction with the Hadamard code to the MIP* protocol for RE from [28]. The proofs of the results of this section can be found in Appendix B.4.

[4] One may wonder why the upper bound in [47] is still exponential in the size of the question set with the refined Gaussian dimension reduction. This is because of the different treatment of the questions. When the questions are classical, we take into account the distribution of the questions. However, if the questions are quantum as considered in [47], the question registers are expressed as a linear combination of matrix basis elements, where an extra factor on the size of the question sets is introduced.

Note that [28] doesn't use the answer reduction technique of [39]. The authors of [28] use a specific PCPP tailored to the low individual-degree code in their answer reduction technique so that it fits the recursive compression framework. However, the answer reduction technique of [28] is more difficult to modify due to its less modular design.

6.1 Subset Tester for the Hadamard Code

To use the [39] answer reduction procedure with a particular error-correcting code, one must show that this code satisfies certain efficient testability properties. Here we show this for the Hadamard code. Specifically, we show that the Hadamard code has a *subset tester* in the sense of [39, Section 16], which ensures that the provers have a global Hadamard encoding of some bitstring.

First, we recall the definition and key properties of the Hadamard code.

▶ **Definition 30.** *The Hadamard code encodes $x \in \mathbb{F}_2^k$ as $\mathsf{Enc}_k(x) = (x \cdot y)_{y \in \mathbb{F}_2^k}$. Moreover,*

- *For $x \neq y \in \mathbb{F}_2^k$, $\mathsf{Enc}_k(x)$ and $\mathsf{Enc}_k(y)$ have normalized Hamming agreement at most $\eta_H = \frac{1}{2}$.*
- *There exists an embedding $\mu_k : [k] \to [2^k]$ such that for each $i \in [k]$, $\mu_k(i) = 2^{i-1}$ and $x_i = (\mathsf{Enc}(x))_{\mu_k(i)}$.*
- *There exists a decoding algorithm Dec_k such that $\mathsf{Dec}_k(\mathsf{Enc}_k(x)) = x$ and, for every w not in the range of Enc_k, $\mathsf{Dec}_k(w) = \perp$.*

The decoding algorithm Dec_k on input w, first computes $x = (w_{\mu_k(k)}, \ldots, w_{\mu_k(1)})$ outputs x if $w = \mathsf{Enc}_k(x)$ and $\perp$ otherwise. Note that both Enc_k and Dec_k run in time exponential in k.

▶ **Proposition 31.** *For the subset $F = \{x_1, \ldots, x_k\} \subseteq \mathbb{F}_2^n$ sampled according to a distribution D and a uniformly random $y \in \mathbb{F}_2^n$, if a quantum strategy with $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and measurements*

$$\left\{M_{a,c,a'}^{F,y} \mid a, a' \in \mathbb{F}_2^k, c \in \mathbb{F}_2\right\}, \left\{N_b^F \mid b \in \mathbb{F}_2^k\right\}, \{N_d^y \mid d \in \mathbb{F}_2\}$$

can pass the subset tester with probability $1 - \varepsilon$, then there is a Hilbert space $\mathcal{H}'_A \otimes \mathcal{H}'_B$, a state $|aux\rangle = |aux_A\rangle \otimes |aux_B\rangle \in \mathcal{H}'_A \otimes \mathcal{H}'_B$ and a projective measurement $\left\{\hat{G}_u \mid u \in \mathbb{F}_2^n\right\}$ on $\mathcal{H}_B \otimes \mathcal{H}'_B$ such that if we write $|\psi'\rangle = |\psi\rangle \otimes |aux\rangle$

$$\mathop{\mathbb{E}}_{F \sim D} \sum_{a \in \mathbb{F}_2^k} \|N_a^F \otimes \mathbb{1}_{\mathcal{H}'} \otimes \mathbb{1}_B |\psi'\rangle - \mathbb{1}_A \otimes \sum_{\substack{u : u \cdot x_i = a_i \\ \forall i \in [k]}} \hat{G}_u |\psi'\rangle\|^2 \leq (2k-1)^2 (45 + 12\sqrt{k})\sqrt{\varepsilon}.$$

6.2 Answer Reduction Protocol

The subset tester of the Hadamard code implies that we can replace the low-degree code of the answer reduction technique in [39, Section 17.4] by the Hadamard code. The other key ingredient of Natarajan and Wright's answer reduction is probabilistically checkable proofs of proximity, so we recall its definition and key properties that we will use later.

▶ **Definition 32** (Probabilistically checkable proofs of proximity (PCPP))**.** *For functions $r, q : \mathbb{Z}^+ \to \mathbb{Z}^+$, $t : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+$, and constants $s, \gamma \in [0, 1]$, a pair language $L \subseteq \{0,1\}^* \times \{0,1\}^*$ is in $\mathrm{PCPP}_{s,\gamma}[r, q, t]$ if there exists an (r, q, t)-restricted PCPP verifier V with the following properties:*

Completeness: *If $(x, y) \in L$, there exists a proof π such that $\Pr_R[V^{y,\pi}(x, |y|; R) = 1] = 1$ where $V^{y,\pi}(x, |y|; R)$ denotes the decision V on input $(x, |y|)$, oracle access to (y, π) with $q(|x|)$ queries, and randomness R from $r(|x|)$ coin tosses.*

Soundness: *Let $L_x = \{y \mid (x, y) \in L\}$. If (x, y) is such that y is γ-far from $L_x \cap \{0,1\}^{|y|}$, then for every π, $\Pr_R[V^{y,\pi}(x, |y|; R) = 1] \leq s$.*

We work with the PCPP such that when L is an $\mathsf{NTIME}(T)$ pair language,

Randomness complexity: $r(m) = \log_2 T(m) + O(\log_2 \log_2 T(m))$,

Query complexity: $q(m) = O(1)$, and

Verification time: $t(m, K) = \mathrm{poly}(m, \log_2 K, \log_2 T(m + K))$.

We are going to apply the PCPP defined above to the following language.

▶ **Definition 33.** *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_V)$ be an* MIP* *verifier, where Alg_Q is his algorithm to sample the questions and Alg_A is his algorithm to check the answers. Suppose on inputs of length n it has question length $\ell_Q(n)$ and answer length $\ell_A(n)$. We define*

$$L_{\mathsf{Enc}} = \left\{(\mathsf{input}, x_0, x_1, \mathsf{Enc}_{\ell_A(|\mathsf{input}|)}(y_0), \mathsf{Enc}_{\ell_A(|\mathsf{input}|)}(y_0)) \mid \mathsf{Alg}_A(\mathsf{input}, x_0, x_1, y_0, y_1) = 1\right\},$$

which are all the accepted tuples with the answers encoded by $\mathsf{Enc}_{\ell_A(|\mathsf{input}|)}$.

Note that when $|\mathsf{input}| = n$, the running time of the decider of L_{Enc} is the maximal of the running time of Alg_A and $\mathsf{Dec}_{\ell_A(n)}$ as pointed out in [39, Proposition 17.7]. Suppose $\gamma \leq \eta_H/2 = 1/4$. Then by [39, Proposition 17.8], if $(\mathsf{input}, x_0, x_1, z_0, z_1)$ does not correspond to the encoding of any assignment accepted by Alg_A, for every proof π

$$\Pr_R[V_{\mathrm{PCPP}}^{z_0,z_1,\pi}(\mathsf{input}, x_0, x_1, |z_0| + |z_1|; R) = 1] \leq s$$

where s is the soundness of V_{PCPP}.

▶ **Definition 34.** *We instantiate the answer-reduced* MIP* *protocol with the following components and notations.*

1) *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ be an* MIP* *verifier for a Language L. Suppose on inputs of size n, the verifier V has question length $\ell_{V,Q}(n)$, answer length $\ell_{V,A}(n)$.*
2) *Let $G_k(\boldsymbol{T})$ be the subset tester from Section 6.1 for the Hadamard code of $\mathbb{F}_2^k$ with the embedding μ_k, and for the subset $\boldsymbol{T}$ sampled according to some distribution D.*
3) *Let L_{Enc} be the language defined in Definition 33, and let V_{PCPP} be its* PCPP *verifier with $\gamma \leq 1/4$ and constant soundness s. Suppose on inputs of size n it has proof length $\ell_\pi(n)$.*
4) *We write $\ell_1 \coloneqq \ell_{V,A}(n)$ and $\ell_2 \coloneqq \ell_\pi(n)$.*

Next, we give the protocol of the answer reduced verifier V^{AR}, which requires the provers to encode their proof π by the Hadamard code of $\mathbb{F}_2^{\ell_2}$. The protocol is very similar to the protocol presented in [39, Figure 15], and can be found in Appendix B.4.

▶ **Theorem 35.** *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ be an* MIP* *protocol for a language L. Suppose the* PCPP *verifier is chosen so that $\gamma \leq 1/4$. Suppose further that V has the following property: for any $\mathsf{input} \in L$, the prover has a real commuting symmetric EPR strategy with a value 1. Then V^{AR} obtained by applying the the answer reduction procedure to V is also an* MIP* *verifier for L with the following two conditions:*

Completeness. *If $\mathsf{input} \in L$, there is a value-1 strategy for V^{AR}.*

Soundness. *Given input, suppose there is a strategy for V^{AR} with value $1 - \varepsilon$. Then there exists constants K_1 and K_2 such that there is a strategy for V on input with value $1 - K_1 - K_2\varepsilon^{1/96}$.*

▶ **Theorem 36.** RE *is contained in* $\mathrm{MIP}^*[\mathrm{poly}, O(1)]$ *with completeness* 1 *and a constant soundness.*

Alternatively, we can first apply the answer reduction technique from [39] to the oracularized protocol to reduce its answer size to $O(\log(n))$ and then apply our answer reduction to further reduce it to $O(1)$. Compared with the approach above, this approach gives us an MIP* protocol for RE with shorter questions but worse soundness.

References

1 Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error rate. *SIAM Journal on Computing*, 38(4):1207, 2008.

2 Dorit Aharonov, Xun Gao, Zeph Landau, Yunchao Liu, and Umesh Vazirani. A polynomial-time classical algorithm for noisy random circuit sampling. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 945–957, 2023.

3 Rotem Arnon-Friedman and Jean-Daniel Bancal. Device-independent certification of one-shot distillable entanglement. *New Journal of Physics*, 21(3):033010, 2019.

4 Rotem Arnon-Friedman, Zvika Brakerski, and Thomas Vidick. Computational entanglement theory. *arXiv preprint*, 2023. arXiv:2310.02783.

5 Rotem Arnon-Friedman and Henry Yuen. Noise-Tolerant Testing of High Entanglement of Formation. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11:1–11:12, Dagstuhl, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2018.11.

6 Srinivasan Arunachalam and Penghui Yao. Positive spectrahedra: invariance principles and pseudorandom generators. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, pages 208–221, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3519935.3519965.

7 László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, March 1991. doi:10.1007/BF01200056.

8 Ainesh Bakshi, Nadiia Chepurko, and Rajesh Jayaram. Testing positive semi-definiteness via random submatrices. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, FOCS 2020, pages 1191–1202. IEEE, 2020.

9 Salman Beigi. A new quantum data processing inequality. *Journal of Mathematical Physics*, 54(8):082202, 2013. doi:10.1063/1.4818985.

10 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.

11 Dolev Bluvstein, Simon J Evered, Alexandra A Geim, Sophie H Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, pages 1–3, 2023.

12 Sergio Boixo, Sergei V Isakov, Vadim N Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J Bremner, John M Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14(6):595–600, 2018.

13 J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions (extended abstract). In *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing*, STOC 1977, pages 106–112, New York, NY, USA, 1977. Association for Computing Machinery. doi:10.1145/800105.803400.

14 Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of NISQ. *Nature Communications*, 14(1):6001, September 2023. doi:10.1038/s41467-023-41217-6.

15 R. Cleve, P. Hoyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *Proceedings. 19th IEEE Annual Conference on Computational Complexity, 2004.*, pages 236–249, 2004. doi:10.1109/CCC.2004.1313847.

16 Rodney Coleman. *Calculus on Normed Vector Spaces*. Springer-Verlag, New York, New York, NY, 1997.

17 Yangjing Dong, Honghao Fu, Anand Natarajan, Minglong Qin, Haochen Xu, and Penghui Yao. The computational advantage of MIP* vanishes in the presence of noise. *arXiv preprint*, 2023. arXiv:2312.04360.

18 Bill Fefferman and Zachary Remscrim. Eliminating intermediate measurements in space-bounded quantum computation. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 1343–1356, 2021.

19 Honghao Fu. Constant-sized correlations are sufficient to self-test maximally entangled states with unbounded dimension. *Quantum*, 6:614, January 2022. doi:10.22331/q-2022-01-03-614.

20 Badih Ghazi, Pritish Kamath, and Prasad Raghavendra. Dimension reduction for polynomials over gaussian space and applications. In *Proceedings of the 33rd Computational Complexity Conference*, CCC '18, pages 28:1–28:37, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2018.28.

21 Insu Han, Dmitry Malioutov, Haim Avron, and Jinwoo Shin. Approximating spectral sums of large-scale matrices using stochastic Chebyshev approximations. *SIAM Journal on Scientific Computing*, 39(4):A1558–A1585, 2017.

22 Prahladh Harsha, Adam Klivans, and Raghu Meka. An invariance principle for polytopes. *J. ACM*, 59(6), January 2013. doi:10.1145/2395116.2395118.

23 Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, July 2001. doi:10.1145/502090.502098.

24 Marcus Isaksson and Elchanan Mossel. Maximally stable Gaussian partitions with discrete applications. *Israel Journal of Mathematics*, 189(1):347–396, 2012.

25 Tsuyoshi Ito, Hirotada Kobayashi, and Keiji Matsumoto. Oracularization and two-prover one-round interactive proofs against nonlocal strategies. In *Proceedings of the 2009 24th Annual IEEE Conference on Computational Complexity*, CCC 2009, pages 217–228, Washington, DC, USA, 2009. IEEE Computer Society. doi:10.1109/CCC.2009.22.

26 Tsuyoshi Ito and Thomas Vidick. A multi-prover interactive proof for NEXP sound against entangled provers. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, FOCS 2012, pages 243–252. IEEE, 2012.

27 Zhengfeng Ji. Compression of quantum multi-prover interactive proofs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 289–302, New York, NY, USA, 2017. ACM. doi:10.1145/3055399.3055441.

28 Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP* = RE. *arXiv preprint*, 2020. arXiv:2001.04383.

29 Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Quantum soundness of the classical low individual degree test. *arXiv preprint*, 2020. arXiv:2009.12982.

30 Daniel M. Kane. A Polylogarithmic PRG for Degree 2 Threshold Functions in the Gaussian Setting. In David Zuckerman, editor, *30th Conference on Computational Complexity (CCC 2015)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 567–581, Dagstuhl, Germany, 2015. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2015.567.

31 Zander Kelley and Raghu Meka. Random restrictions and prgs for ptfs in gaussian space. In *Proceedings of the 37th Computational Complexity Conference*, CCC '22, Dagstuhl, DEU, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2022.21.

32 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. doi:10.1137/090751293.

33 Julia Kempe, Oded Regev, and Ben Toner. Unique games with entangled provers are easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010. doi:10.1137/090772885.

34 Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, STOC '02, pages 767–775, New York, NY, USA, 2002. Association for Computing Machinery. doi:10.1145/509907.510017.

35 Robert Krauthgamer and Ori Sasson. Property testing of data dimensionality. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA 2003, pages 18–27, USA, 2003. Society for Industrial and Applied Mathematics.

36 Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC 2010, pages 427–436, New York, NY, USA, 2010. Association for Computing Machinery. doi:10.1145/1806689.1806749.

37 Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *46th Annual IEEE Symposium on Foundations of Computer Science*, FOCS 2005, pages 21–30. IEEE, 2005.

38 Anand Natarajan and Thomas Vidick. A quantum linearity test for robustly verifying entanglement. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 1003–1015, 2017.

39 Anand Natarajan and John Wright. NEEXP is Contained in MIP*. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science*, FOCS 2019, pages 510–518. IEEE, 2019. doi:10.1109/FOCS.2019.00039.

40 Anand Natarajan and Tina Zhang. Quantum free games. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 1603–1616, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3564246.3585208.

41 Deanna Needell, William Swartworth, and David P. Woodruff. Testing positive semidefiniteness using linear measurements. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science*, FOCS 2022, pages 87–97, 2022. doi:10.1109/FOCS54457.2022.00016.

42 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, Cambridge, UK, 2013.

43 Ryan O'Donnell, Rocco A. Servedio, and Li-Yang Tan. Fooling gaussian ptfs via local hyperconcentration. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 1170–1183, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3357713.3384281.

44 Ryan O'Donnell, Rocco A. Servedio, and Li-Yang Tan. Fooling polytopes. *J. ACM*, 69(2), January 2022. doi:10.1145/3460532.

45 Connor Paddock. Rounding near-optimal quantum strategies for nonlocal games to strategies using maximally entangled states. *arXiv preprint*, 2022. arXiv:2203.02525.

46 Minglong Qin and Penghui Yao. Nonlocal games with noisy maximally entangled states are decidable. *SIAM Journal on Computing*, 50(6):1800–1891, 2021.

47 Minglong Qin and Penghui Yao. Decidability of Fully Quantum Nonlocal Games with Noisy Maximally Entangled States. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 97:1–97:20, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2023.97.

48 Oded Regev and Liron Schiff. Impossibility of a quantum speed-up with a faulty oracle. In *International Colloquium on Automata, Languages, and Programming*, pages 773–781. Springer, 2008.

49 Ben W. Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, April 2013. doi:10.1038/nature12035.

50 Hristo S. Sendov. The higher-order derivatives of spectral functions. *Linear Algebra and its Applications*, 424(1):240–281, 2007. Special Issue in honor of Roger Horn. doi:10.1016/j.laa.2006.12.013.

51 Victor Shoup. New algorithms for finding irreducible polynomials over finite fields. *Mathematics of computation*, 54(189):435–447, 1990.

52 Anna Skripka and Anna Tomskova. *Multilinear operator integrals*. Springer, 2019.

53 William Slofstra. The set of quantum correlations is not closed. *Forum of Mathematics, Pi*, 7:e1, 2019. doi:10.1017/fmp.2018.3.

54 William Slofstra. Tsirelson's problem and an embedding theorem for groups arising from non-local games. *Journal of the American Mathematical Society*, 33:1–56, 2020. doi:/10.1090/jams/929.

55 Salil P. Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012. doi:10.1561/0400000010.

56 Thomas Vidick. Almost synchronous quantum correlations. *Journal of mathematical physics*, 63(2), 2022.

A Preliminary

For $n \in \mathbb{Z}_{>0}$, let $[n]$ and $[n]_{\geq 0}$ represent the sets $\{1, \ldots, n\}$ and $\{0, \ldots, n-1\}$, respectively. Given a finite set $\mathcal{X}$ and a natural number k, let $\mathcal{X}^k$ be the set $\mathcal{X} \times \cdots \times \mathcal{X}$, the Cartesian product of $\mathcal{X}$, k times. For any $\sigma \in \mathbb{Z}_{\geq 0}^k$, we define $|\sigma| = |\{i : \sigma_i \neq 0\}|$.

In this paper, the lowercase letters in bold $\mathbf{x}, \mathbf{y}, \cdots$ are reserved for random variables. The capital letters in bold, $\mathbf{A}, \mathbf{B}, \ldots$ are reserved for random operators.

A.1 Quantum Mechanics

A quantum system is associated with a complex finite-dimensional Hilbert space, denoted by A. A quantum state in A can be completely described by a density operator, a positive semidefinite operator with trace one. If the dimension of A is m, we denote the set of Hermitian matrices in A by $\mathcal{H}_m$. The identity matrix is denoted by $\mathbb{1}_m$ or $\mathbb{1}_A$. The state of a composite quantum system is the Kronecker product of the state spaces of the component systems. An important operation on a composite system $A \otimes B$ is the *partial trace* $\mathrm{Tr}_B(\cdot)$ which effectively derives the marginal state of the subsystem A (denoted by ψ_A) from the quantum state ψ_{AB}. The partial trace is given by

$$\psi_A = \mathrm{Tr}_B \psi_{AB} = \sum_i \left(\mathbb{1}_A \otimes \langle i|\right) \psi_{AB} \left(\mathbb{1}_A \otimes |i\rangle\right),$$

where $\{|i\rangle\}$ is an orthonormal basis in B. A linear map from a system A to a system B is *unital* if it maps $\mathbb{1}_A$ to $\mathbb{1}_B$. A *quantum measurement* is represented by a *positive operator-valued measure* (POVM), which is a set of positive semidefinite operators $\{M_1, \ldots, M_n\}$ satisfying $\sum_{i=1}^n M_i = \mathbb{1}$, where n is the number of possible measurement outcomes. Suppose that the state of the quantum system is ψ, then the probability that it produces i is $\mathrm{Tr}\, M_i \psi$. We use $\overrightarrow{M} = (M_1, \ldots, M_n)$ to represent an ordered set of operators.

A.2 Matrix Analysis

A.2.1 Matrix Spaces

Given $m \in \mathbb{Z}_{>0}$ and $M \in \mathcal{H}_m$, we use $M_{i,j}$ to represent the (i,j)-th entry of M. For $1 \leq p \leq \infty$, the p-norm of M is defined to be

$$\|M\|_p = \left(\sum_{i=1}^{m} s_i\left(M\right)^p \right)^{1/p},$$

where $\left(s_1\left(M\right), s_2\left(M\right), \ldots, s_m\left(M\right)\right)$ are the singular values of M sorted in nonincreasing order. $\|M\| = \|M\|_\infty = s_1\left(M\right)$. The *normalized p-norm* of M is defined as

$$\left|\!\left|\!\left| M \right|\!\right|\!\right|_p = \left(\frac{1}{m} \sum_{i=1}^{m} s_i\left(M\right)^p \right)^{1/p} \tag{4}$$

and $\left|\!\left|\!\left| M \right|\!\right|\!\right| = \left|\!\left|\!\left| M \right|\!\right|\!\right|_\infty = s_1\left(M\right)$.

Given $P, Q \in \mathcal{M}_m$, we define

$$\langle P, Q \rangle = \frac{1}{m} \mathrm{Tr}\ P^\dagger Q. \tag{5}$$

It is easy to verify that $\langle \cdot, \cdot \rangle$ is an inner product. $\left(\langle \cdot, \cdot \rangle, \mathcal{H}_m\right)$ forms a Hilbert space. For any $M \in \mathcal{H}_m$, $\left|\!\left|\!\left| M \right|\!\right|\!\right|_2^2 = \langle M, M \rangle$.

We say that $\{\mathcal{B}_0, \ldots, \mathcal{B}_{m^2-1}\}$ is a *standard orthonormal basis* in $\mathcal{M}_m$ if it is an orthonormal basis with all elements being Hermitian and $\mathcal{B}_0 = \mathbb{1}_m$.

▶ **Fact 37** ([46, Lemma 2.10]). *For any integer $m \geq 2$, a standard orthonormal basis exists in $\mathcal{M}_m$.*

Given a standard orthonormal basis $\mathcal{B} = \{\mathcal{B}_i\}_{i=0}^{m^2-1}$ in $\mathcal{H}_m$, every matrix $M \in \mathcal{H}_m^{\otimes n}$ has a *Fourier expansion* with respect to the basis $\mathcal{B}$ given by

$$M = \sum_{\sigma \in [m^2]_{\geq 0}^n} \widehat{M}\left(\sigma\right) \mathcal{B}_\sigma,$$

where $\mathcal{B}_\sigma = \bigotimes_{i=1}^n \mathcal{B}_{\sigma_i}$.

▶ **Definition 38.** *Let $\mathcal{B} = \{\mathcal{B}_i\}_{i=0}^{m^2-1}$ be a standard orthonormal basis in $\mathcal{H}_m$, $P \in \mathcal{H}_m^{\otimes n}$.*

1. *The* degree *of P is defined to be*

$$\deg P = \max\left\{ |\sigma| : \widehat{P}\left(\sigma\right) \neq 0 \right\}.$$

   Recall that $|\sigma|$ represents the number of nonzero entries of σ.
2. *For any $i \in [n]$, the* influence *of i-th coordinate is defined to be:*

$$\mathrm{Inf}_i(P) = \left|\!\left|\!\left| P - \mathbb{1}_m \otimes \mathrm{Tr}_i P \right|\!\right|\!\right|_2^2,$$

   where $\mathbb{1}_m$ is in the i'th quantum system, and the partial trace Tr_i derives the marginal state of the remaining $n-1$ quantum systems except for the i'th one.
3. *The total influence is defined by*

$$\mathrm{Inf}\left(P\right) = \sum_i \mathrm{Inf}_i\left(P\right).$$

▶ **Fact 39** ([46, Lemma 2.16]). *Given $P \in \mathcal{H}_m^{\otimes n}$, a standard orthonormal basis $\mathcal{B} = \{\mathcal{B}_i\}_{i=0}^{m^2-1}$ in $\mathcal{H}_m$ and a subset $S \subseteq [n]$, it holds that*

1. $\mathrm{Inf}_i(P) = \sum_{\sigma:\sigma_i \neq 0} |\widehat{P}(\sigma)|^2$;
2. $\mathrm{Inf}(P) = \sum_{\sigma} |\sigma| |\widehat{P}(\sigma)|^2 \leq \deg P \cdot \|\|P\|\|_2^2$.

The inequality in item 2 follows from Parseval's identity, which is immediate by the Fourier expansion of P (Fact 37).

▶ **Fact 40** (Parseval's identity). *For any $P \in \mathcal{H}_m^{\otimes n}$,*

$$\|\|P\|\|_2^2 = \sum_{\sigma} |\widehat{P}(\sigma)|^2.$$

Quantum maximal correlations introduced by Beigi [9] are crucial to our analysis.

▶ **Definition 41** (Maximal correlation [9]). *Given quantum systems A, B of dimension m and a bipartite state ψ_{AB} with $\psi_A = \psi_B = \frac{\mathbb{1}_m}{m}$, the maximal correlation of ψ_{AB} is defined to be*

$$\rho(\psi_{AB}) = \sup\left\{ \left|\mathrm{Tr}\left(\left(P^\dagger \otimes Q\right)\psi_{AB}\right)\right| \ : \ \begin{array}{c} P, Q \in \mathbb{C}^{m\times m}, \\ \mathrm{Tr}\ P = \mathrm{Tr}\ Q = 0, \|P\|_2 = \|Q\|_2 = 1. \end{array} \right\}$$

▶ **Fact 42** ([9]). *Given quantum systems A, B and a bipartite quantum state ψ_{AB} with $\psi_A = \mathbb{1}_{m_A}/m_A$ and $\psi_B = \mathbb{1}_{m_B}/m_B$, it holds that $\rho(\psi_{AB}) \leq 1$.*

▶ **Definition 43.** *Given quantum systems A and B with $\dim(A) = \dim(B) = m$, a bipartite state $\psi_{AB} \in \mathcal{D}(A \otimes B)$ is an m-dimensional* noisy *maximally entangled state (MES) if $\psi_A = \psi_B = \mathbb{1}_m/m$ and its maximal correlation $\rho = \rho(\psi_{AB}) < 1$.*

An interesting class of noisy MESs is the isotropic states, which are the states obtained by depolarizing MESs with arbitrarily small noise.

▶ **Fact 44** ([46, Lemma 3.9]). *For any $0 \leq \epsilon < 1$ integer $m > 1$, it holds that*

$$\rho\left((1-\epsilon)|\Psi\rangle\langle\Psi| + \epsilon \frac{\mathbb{1}_m}{m} \otimes \frac{\mathbb{1}_m}{m}\right) = 1 - \epsilon,$$

where $|\Psi\rangle = \frac{1}{\sqrt{m}} \sum_{i=0}^{m-1} |m, m\rangle$ is an m-dimensional MES.

▶ Remark 45. Fact 44 indicates the maximal correlation of an isotropic state is strictly less than 1. The class of noisy MES also contains other states. It is not hard to prove that any mixture of at least three out of the four orthogonal EPR states is a 2-dimensional noisy MES.

▶ **Fact 46** ([46, Lemma 7.4]). *Given $m \in \mathbb{Z}_{>0}$, $m \geq 2$, and a noisy m-dimensional MES ψ_{AB}. Then there exist standard orthonormal bases $\mathcal{A} = \{\mathcal{A}_i\}_{i=0}^{m^2-1}$ and $\mathcal{B} = \{\mathcal{B}_i\}_{i=0}^{m^2-1}$ in $\mathcal{H}_m$ such that*

$$\mathrm{Tr}\left(\left(\mathcal{A}_i \otimes \mathcal{B}_j\right)\psi_{AB}\right) = \begin{cases} c_i & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where $c_0 = 1 \geq c_1 = \rho(\psi_{AB}) \geq c_2 \geq \ldots c_{m^2-1} \geq 0$ and $\rho(\psi_{AB})$ is defined in Definition 41.

▶ **Definition 47.** *Given $m \in \mathbb{Z}_{>0}$, $\rho \in [0, 1]$, a noise operator $\Delta_\rho : \mathcal{H}_m \to \mathcal{H}_m$ is defined as follows. For any $P \in \mathcal{H}_m$,*

$$\Delta_\rho(P) = \rho P + \frac{1-\rho}{m}(\mathrm{Tr}\ P) \cdot \mathbb{1}_m.$$

With a slight abuse of notations, the noise operator $\Delta_\rho^{\otimes n}$ on the space $\mathcal{H}_m^{\otimes n}$ is also denoted by Δ_ρ.

▶ **Fact 48** ([46, Lemma 3.5]). *Given integers $d, n, m > 0$, $\rho \in [0,1]$, a standard orthonormal basis of $\mathcal{H}_m$: $\mathcal{B} = \{\mathcal{B}_i\}_{i=0}^{m^2-1}$, then for any $P \in \mathcal{H}_m^{\otimes n}$ with a Fourier expansion $P = \sum_{\sigma \in [m^2]_{\geq 0}^n} \widehat{P}(\sigma) \mathcal{B}_\sigma$, it holds that*

$$\Delta_\rho(P) = \sum_{\sigma \in [m^2]_{\geq 0}^n} \rho^{|\sigma|} \widehat{P}(\sigma) \mathcal{B}_\sigma.$$

A.2.2 Random Matrices

For integer $n \geq 1$, γ_n represents the distribution of an n-dimensional standard normal distribution. For any $0 \leq \rho \leq 1$, $\mathcal{G}_\rho$ represents a ρ-correlated Gaussian distribution, which is a 2-dimensional Gaussian distribution

$$(X, Y) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Namely, the marginal distributions X and Y are distributed according to γ_1 and $\mathbb{E}[XY] = \rho$.

▶ **Definition 49.** *Given $h, n, m \in \mathbb{Z}_{>0}$, we say $P(\mathbf{g})$ is a random matrix if it can be expressed as*

$$P(\mathbf{g}) = \sum_{\sigma \in [m^2]_{\geq 0}^h} p_\sigma(\mathbf{g}) \mathcal{B}_\sigma, \tag{7}$$

where $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$ is a standard orthonormal basis in $\mathcal{H}_m$, $p_\sigma : \mathbb{R}^n \to \mathbb{R}$ for all $\sigma \in [m^2]_{\geq 0}^h$ and $\mathbf{g} \sim \gamma_n$. Moreover, we say $P(\mathbf{g}) \in L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_n\right)$ if $\int_{\mathbb{R}^n} p_\sigma^2(x) \gamma_n(\mathrm{d}x) < \infty$ for all $\sigma \in [m^2]_{\geq 0}^h$.

We define the degree of random operators:

▶ **Definition 50.** *Given integers $n, h > 0, m > 1$ and random operator $\mathbf{P} \in L^p\left(\mathcal{H}_m^{\otimes h}, \gamma_n\right)$, the degree of $\mathbf{P}$, denoted by $\deg(\mathbf{P})$, is*

$$\max_{\sigma \in [m^2]_{\geq 0}^h} \deg(p_\sigma).$$

We say $\mathbf{P}$ is multilinear if $p_\sigma(\cdot)$ is multilinear for all $\sigma \in [m^2]_{\geq 0}^h$.

A.2.3 Fréchet Derivatives and Spectral Functions

The Fréchet derivatives are derivatives on Banach spaces. In this paper, we only concern ourselves with Fréchet derivatives on matrix spaces. Readers may refer to [16] for a detailed treatment.

▶ **Definition 51.** *Given a map $f : \mathcal{H}_m \to \mathcal{H}_m$ and $P, Q \in \mathcal{H}_m$, the Fréchet derivative of f at P with direction Q is defined to be*

$$Df(P)[Q] = \frac{d}{dt} f(P + tQ)|_{t=0}.$$

The k-th order Fréchet derivative of f at P with direction $(Q_1, \ldots, Q_k)$ is defined to be

$$D^k f(P)[Q_1, \ldots, Q_k] = \frac{d}{dt}\left(D^{k-1} f(P + tQ_k)[Q_1, \ldots, Q_{k-1}]\right)|_{t=0}.$$

To keep notations short, we use $D^k f(P)[Q]$ to represent $D^k f(P)[Q, \ldots, Q]$.

In this paper, we are concerned with *spectral functions*, a special class of matrix functions. We say that the function $F : \mathcal{H}_m \to \mathcal{H}_m$ is a spectral function if there exists a function $f : \mathbb{R} \to \mathbb{R}$ such that $f(P) = \sum_i f(\lambda_i) |v_i\rangle\langle v_i|$, where $P = \sum_i \lambda_i |v_i\rangle\langle v_i|$ is a spectral decomposition of P. With slight abuse of notation, we use the same notation f to represent the function on $\mathbb{R}$ and the corresponding spectral function, whenever it is clear from the context.

Given $n \in \mathbb{Z}_{>0}$, we denote $\mathcal{C}^n$ to be the space of functions continuously differentiable n times.

▶ **Definition 52.** *Let $\lambda_0, \ldots, \lambda_n \in \mathbb{R}$ and let $f \in \mathcal{C}^n$. The divided difference $f^{[n]}$ is defined recursively by*

$$f^{[n]}(\lambda_0, \lambda_1, \tilde{\lambda}) = \begin{cases} \frac{f^{[n-1]}(\lambda_0,\tilde{\lambda}) - f^{[n-1]}(\lambda_1,\tilde{\lambda})}{\lambda_0 - \lambda_1} & \text{if } \lambda_0 \neq \lambda_1, \\ \frac{\mathrm{d}}{\mathrm{d}\lambda_0} f^{[n-1]}(\lambda_0, \tilde{\lambda}) & \text{if } \lambda_0 = \lambda_1, \end{cases}$$

where $\tilde{\lambda} = (\lambda_2, \ldots, \lambda_n)$.

It is well known that $f^{[n]}$ is a symmetric function.

▶ **Fact 53** ([52, Theorem 5.3.2], [50, Theorem 6.1]). *Given $m, n \in \mathbb{Z}_{>0}$, $P, Q \in \mathcal{H}_m$. Suppose that P has a spectral decomposition*

$$P = \sum_{i=1}^{m} \lambda_i \Pi_i, \tag{8}$$

where $\lambda_1 \geq \cdots \geq \lambda_m$, $\{\Pi_i\}_{i\in[m]}$ are rank-one projectors satisfying that $\sum_{i=1}^m \Pi_i = \mathbb{1}$ and $\Pi_i \Pi_j = 0$ for all $i \neq j$. Let $f \in \mathcal{C}^n$. Then

$$D^n f(P)[Q] = \sum_{i_0,\ldots,i_n \in [m]} f^{[n]}(\lambda_{i_0}, \ldots, \lambda_{i_n}) \Pi_{i_0} Q \Pi_{i_1} Q \ldots Q \Pi_{i_n}.$$

▶ **Fact 54** ([52, Theorem 5.3.12]). *Given $m, n \in \mathbb{Z}_{>0}$, $P, Q \in \mathcal{H}_m$. Let $f \in \mathcal{C}^n$. Denote*

$$\Delta_{n,f}(P, Q) = f(P + Q) - \sum_{k=0}^{n-1} \frac{1}{k!} D^k f(P)[Q],$$

then there exists a constant c_n depending only on n such that

$$|\mathrm{Tr}[\Delta_{n,f}(P, Q)]| \leq c_n \|f^{(n)}\|_\infty \|Q\|_n^n,$$

where $\|f^{(n)}\|_\infty$ denotes the supremum of $f^{(n)}$.

A.2.4 The Distance from PSD Matrices

Define the function $\zeta : \mathbb{R} \to \mathbb{R}$ as follows.

$$\zeta(x) = \begin{cases} x^2 & \text{if } x \leq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

The function ζ measures the distance between a given matrix and its closest positive semi-definite matrix:

▶ **Fact 55** ([46, Lemma 9.1]). *Given an integer $m > 0$, $M \in \mathcal{H}_m$, $\Delta = \{X \in \mathcal{H}_m : X \geq 0\}$, let*

$$\mathcal{R}(M) = \arg\min \{\|M - X\|_2 : X \in \Delta\}$$

be a rounding map of Δ with respect to the distance $\|\cdot\|_2$. It holds that

$$\mathrm{Tr}\ \zeta(M) = \|M - \mathcal{R}(M)\|_2^2.$$

▶ **Fact 56** ([46, Lemma 10.4]). *For any Hermitian matrices P and Q, it holds that*

$$|\mathrm{Tr}\ (\zeta(P+Q) - \zeta(P))| \leq 2\left(\|P\|_2\|Q\|_2 + \|Q\|_2^2\right).$$

We will need to let ζ to be mollified[5] to get a smooth function:

▶ **Fact 57** ([37, Lemma 3.21]). *Given $\lambda > 0$, there exists a C^∞ function ζ_λ satisfying*

1. $\|\zeta_\lambda - \zeta\|_\infty \leq 2\lambda^2$,
2. *For any integer $n \geq 2$, there exists a constant B_n independent of λ such that*

$$\|(\zeta_\lambda)^{(n)}\|_\infty \leq B_n \lambda^{2-n}.$$

A.3 k-wise Uniform Hash Functions and Random Variables

▶ **Definition 58.** *A family $\mathcal{F} = \{f : [n] \to [p]\}$ of hash functions is k-wise uniform if for any $y_1, \ldots, y_k \in [p]$ and distinct $x_1, \ldots, x_k \in [n]$:*

$$\Pr_{f \in_u \mathcal{F}}[f(x_i) = y_i \wedge \cdots \wedge f(x_k) = y_k] = \frac{1}{p^k}.$$

▶ **Definition 59.** *A random vector $\mathbf{z} \in [p]^n$ is k-wise uniform if for any $y_1, \ldots, y_k \in [p]$ and distinct $x_1, \ldots, x_k \in [n]$:*

$$\Pr_{\mathbf{z}}[\mathbf{z}_{x_i} = y_i \wedge \cdots \wedge \mathbf{z}_{x_k} = y_k] = \frac{1}{p^k}.$$

▶ **Lemma 60.** *Let p be a power of 2. There exists an efficient construction of k-wise uniform hash functions $\mathcal{F} = \{f : [n] \to [p]\}$ of size $|\mathcal{F}| = O(\max(n,p)^k)$.*

Proof. For $k = 2$, efficient constructions of size $|\mathcal{F}| = O(np)$ are well known (see, e.g., [13]). For general k, let t be the minimal integer satisfying $2^t > \max(n,p)$ and consider the finite field $\mathbb{F}_{2^t}$. We can construct an irreducible polynomial in $\mathbb{F}_2$ of degree t in polynomial time, using, for example, the algorithms of Shoup [51]. Thus, the basic operations in $\mathbb{F}_{2^t}$ can be carried out efficiently. Then the k-wise uniform hash functions $\tilde{\mathcal{F}} : \left\{\tilde{f} : \mathbb{F}_{2^t} \to \mathbb{F}_{2^t}\right\}$ can be efficiently constructed, for example, using the construction in Section 3.5.5 in [55], which has size $|\mathbb{F}_{2^t}|^k = O(\max(n,p))^k$. Then k-wise uniform hash functions from $[n]$ to $\mathbb{F}_{2^t}$ can be constructed by restricting the input domain to $[n]$. k-wise uniform hash functions from $[n]$ to $[p]$ can be further constructed by cutting the output to $\log p$ bits. ◀

▶ **Corollary 61.** *There exists an efficient construction of k-wise uniform random variables $\mathbf{z} \sim \{-1,1\}^n$, which can be enumerated in $O(n^k)$ time.*

Proof. Construct k-wise uniform hash functions $\mathcal{F} = \{f : [n] \to \{-1,1\}\}$, and then define $\mathbf{z} = (f(1), \ldots, f(n))$. By the definition of k-wise uniform hash functions, $\mathbf{z}$ is k-wise uniform random variables. Moreover, the construction of $\mathcal{F}$ is efficient. Finally, the enumeration of $\mathbf{z}$ takes time $O(n^k)$ since we only need to enumerate the set $\mathcal{F}$. ◀

[5] A mollified function ζ_λ is a smooth function that is close to the original function ζ.

A.4 Lemmas for Noisy MIP*

Smoothing

The following lemma reduces the degrees of the POVMs of an MIP* strategy.

▶ **Lemma 62.** *[46, Lemma 6.1][6] Given parameters $0 \leq \rho < 1$, $0 < \delta < 1$, $n, m \in \mathbb{Z}_{>0}$, $m \geq 2$, and an m-dimensional noisy MES ψ_{AB} with the maximal correlation $\rho = \rho(\psi_{AB})$, there exists $d = d(\rho, \delta)$ and a map $f : \mathcal{H}_m^{\otimes n} \to \mathcal{H}_m^{\otimes n}$, such that for any positive semi-definite matrices $P, Q \in \mathcal{H}_m^{\otimes n}$ satisfying $\|P\|_2 \leq 1$ and $\|Q\|_2 \leq 1$. The matrices $P^{(1)} = f(P)$ and $Q^{(1)} = f(Q)$ satisfy that*

1. *$P^{(1)}$ and $Q^{(1)}$ are of degree at most d.*
2. *$\left\|P^{(1)}\right\|_2 \leq 1$ and $\left\|Q^{(1)}\right\|_2 \leq 1$.*
3. *$\left|\mathrm{Tr}\left(\left(P^{(1)} \otimes Q^{(1)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left((P \otimes Q)\psi_{AB}^{\otimes n}\right)\right| \leq \delta$.*
4. *$\frac{1}{m^n}\mathrm{Tr}\ \zeta(P^{(1)}) \leq \delta$ and $\frac{1}{m^n}\mathrm{Tr}\ \zeta(Q^{(1)}) \leq \delta$.*
5. *the map f is linear and unital.*

In particular, we can take $d = \frac{C \log^2 \frac{1}{\delta}}{\delta(1-\rho)}$ for some absolute constant C.

▶ Remark 63. It is easily verified that for the above lemma, for each $\sigma \in [m^2]^n_{\geq 0}$, we have

$$|\widehat{P^{(1)}}(\sigma)| \leq |\widehat{P}(\sigma)| \text{ and } |\widehat{Q^{(1)}}(\sigma)| \leq |\widehat{Q}(\sigma)|.$$

This is because in fact f applies depolarizing noise on P and then eliminates the high degree parts. So the Fourier coefficients are non-increasing in absolute value.

Regularization

The following lemma allows us to identify high-influence registers, and the number of such registers can be upper-bounded.

▶ **Lemma 64.** *[46, Lemma 7.4] Given $0 < \tau < 1$, $d, n, m \in \mathbb{Z}_{>0}$, $m \geq 2$, and a degree-d matrix $P \in \mathcal{H}_m^{\otimes n}$ satisfying $\|P\|_2 \leq 1$, there exists a subset $H \subseteq [n]$ of size $h = |H| \leq \frac{d}{\tau}$ such that for any $i \notin H$,*

$$\mathrm{Inf}_i\left(P^{\leq d}\right) \leq \tau.$$

Rounding

The following lemma shows that we can round a given set of matrices that sum up to $\mathbb{1}$ to a close-by POVM.

▶ **Lemma 65.** *Given $\overrightarrow{X} \in (\mathcal{H}_m^{\otimes n})^t$ satisfying that $\sum_{i=1}^t X_i = \mathbb{1}$, define*

$$\mathcal{R}\left(\overrightarrow{X}\right) = \arg\min\left\{\left\|\overrightarrow{X} - \overrightarrow{P}\right\|_2^2 : \overrightarrow{P} \text{ is a POVM}\right\}$$

It holds that

$$\left\|\mathcal{R}\left(\overrightarrow{X}\right) - \overrightarrow{X}\right\|_2^2 \leq \frac{3(t+1)}{m^n}\sum_{i=1}^t \mathrm{Tr}\ \zeta(X_i) + 6\left(\frac{t}{m^n}\sum_{i=1}^t \mathrm{Tr}\ \zeta(X_i)\right)^{1/2}.$$

[6] The statement is slightly different from that in [46, Lemma 6.1]. The difference arises due to our relocation of the truncating step, which was in [46, Lemma 10.5].

Miscellaneous Lemmas

The following lemmas are used throughout Appendix B.3.

▶ **Fact 66** ([46, Fact 2.1]). *Given registers A, B, operators $P \in \mathcal{H}(A), Q \in \mathcal{H}(B)$ and a bipartite state ψ_{AB}, it holds that*

$$|\mathrm{Tr}\left((P \otimes Q)\,\psi_{AB}\right)| \leq \left(\mathrm{Tr}\,P^2\psi_A\right)^{1/2} \cdot \left(\mathrm{Tr}\,Q^2\psi_B\right)^{1/2}.$$

▶ **Lemma 67.** *Let $\{P_a^x\}_{a\in\mathcal{A}}^{x\in\mathcal{X}}, \{Q_b^y\}_{b\in\mathcal{B}}^{y\in\mathcal{Y}}, \{\tilde{P}_a^x\}_{a\in\mathcal{A}}^{x\in\mathcal{X}}, \{\tilde{Q}_b^y\}_{b\in\mathcal{B}}^{y\in\mathcal{Y}} \subseteq \mathcal{H}_m^{\otimes n}$ be four sets of matrices. If for all $(x,y,a,b) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$,*

$$|\mathrm{Tr}\left((P_a^x \otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(\tilde{P}_a^x \otimes \tilde{Q}_b^y\right)\psi_{AB}^{\otimes n}\right)| \leq \delta |||P_a^x|||_2 |||Q_b^y|||_2$$

for some $\delta > 0$. Then

$$\left|\mathrm{val}_n\left(\{P_a^x\},\{Q_b^y\}\right) - \mathrm{val}_n\left(\left\{\tilde{P}_a^x\right\},\left\{\tilde{Q}_b^y\right\}\right)\right| \leq \delta t \left(\sum_{x,a}\mu_A(x)|||P_a^x|||_2^2\right)^{1/2}\left(\sum_{y,b}\mu_B(y)|||Q_b^y|||_2^2\right)^{1/2}.$$

Proof.

$$\begin{aligned}
&\left|\mathrm{val}_n\left(\{P_a^x\},\{Q_b^y\}\right) - \mathrm{val}_n\left(\left\{\tilde{P}_a^x\right\},\left\{\tilde{Q}_b^y\right\}\right)\right| \\
&\leq \sum_{x,y,a,b}\mu(x,y)|\mathrm{Tr}\left((P_a^x\otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(\tilde{P}_a^x\otimes\tilde{Q}_b^y\right)\psi_{AB}^{\otimes n}\right)| \\
&\leq \delta \sum_{x,y,a,b}\mu(x,y)|||P_a^x|||_2|||Q_b^y|||_2 \\
&\leq \delta\left(\sum_{x,y,a,b}\mu(x,y)|||P_a^x|||_2^2\right)^{1/2}\left(\sum_{x,y,a,b}\mu(x,y)|||Q_b^y|||_2^2\right)^{1/2} \qquad \text{(Cauchy Schwarz)} \\
&= \delta t\left(\sum_{x,a}\mu_A(x)|||P_a^x|||_2^2\right)^{1/2}\left(\sum_{y,b}\mu_B(y)|||Q_b^y|||_2^2\right)^{1/2}.
\end{aligned}$$

◀

▶ **Lemma 68** (Truncation). *Let $\{P_a^x\},\{Q_b^y\}$ be two sets of operators satisfying*

1. *For all x,y, $\sum_a P_a^x = \sum_b Q_b^y = \mathbb{1}$.*
2. *For all x,a,y,b,σ, $\left|\widehat{P}_a^x(\sigma)\right| \leq 1$ and $\left|\widehat{Q}_b^y(\sigma)\right| \leq 1$.*

Let $s_w = D\log m + \log\left(\frac{2}{\delta}\right)$. Then there exist operators $\left\{P_a^{x,(2)}\right\},\left\{Q_b^{y,(2)}\right\}$ satisfying

1. *For each x,y,a,b,σ, the Fourier coefficients of $P_a^{x,(2)}$ and $Q_b^{y,(2)}$ consists of at most s_w bits.*
2. *For all x,y, $\sum_a P_a^{x,(2)} = \sum_b Q_b^{y,(2)} = \mathbb{1}$.*
3. *For all x,y,a,b, $\left|\left|\left|P_a^{x,(2)}\right|\right|\right|_2 \leq 1$ and $\left|\left|\left|Q_b^{y,(2)}\right|\right|\right|_2 \leq 1$.*
4. *For all x,y,a,b, $\left|\mathrm{Tr}\left(\left(P_a^{x,(2)}\otimes Q_b^{y,(2)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left((P_a^x\otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right)\right| \leq \delta$.*
5. *For all x,y,a,b,*

$$\left|\frac{1}{m^D}\mathrm{Tr}\;\zeta\left(P_a^{x,(2)}\right) - \frac{1}{m^D}\mathrm{Tr}\;\zeta\left(P_a^x\right)\right| \leq \delta \text{ and } \left|\frac{1}{m^D}\mathrm{Tr}\;\zeta\left(Q_b^{y,(2)}\right) - \frac{1}{m^D}\mathrm{Tr}\;\zeta\left(Q_b^y\right)\right| \leq \delta.$$

Proof. Let $\alpha = 2^{-s_w} = \delta/(2m^D)$. For each x, y, σ, define $\widehat{P}_a^{x,(1)}(\sigma) = \lfloor \widehat{P}_a^x(\sigma)/\alpha \rfloor \alpha$. For each $x, \sigma \neq 0^D$, define integer $k_{x,\sigma}$ as

$$-\sum_a \widehat{P}_a^{x,(1)}(\sigma) = k_{x,\sigma} \cdot \alpha$$

and for $\sigma = 0^D$, define

$$1 - \sum_a \widehat{P}_a^{x,(1)}(\sigma) = k_{x,0^D} \cdot \alpha.$$

Let $t_{x,\sigma} = \left|\left\{a \in \mathcal{A} : \widehat{P}_a^{x,(1)}(\sigma) \neq \widehat{P}_a^x(\sigma)\right\}\right|$, we can see that $0 \leq k_{x,\sigma} < t_{x,\sigma}$ always holds because $\sum_a P_a^x = \mathbb{1}$ and by the fact that $\widehat{P}_a^{x,(1)}(\sigma) > \widehat{P}_a^x(\sigma) - \alpha$. Let $S_{x,\sigma}$ be an arbitrary subset of $\left\{a \in \mathcal{A} : \widehat{P}_a^{x,(1)}(\sigma) \neq \widehat{P}_a^x(\sigma)\right\}$ of size $k_{x,\sigma}$. Define $P_a^{x,(2)}$ as

$$\widehat{P}_a^{x,(2)}(\sigma) = \begin{cases} \widehat{P}_a^{x,(1)}(\sigma) & \text{if } a \notin S_{x,\sigma} \\ \widehat{P}_a^{x,(1)}(\sigma) + \alpha & \text{if } a \in S_{x,\sigma} \end{cases}$$

Then item 1 and item 2 hold for $P_a^{x,(2)}$. Also, since for $a \in S_{x,\sigma}$ we have $\widehat{P}_a^{x,(1)}(\sigma) < \widehat{P}_a^x(\sigma) \leq 1$, we have $\widehat{P}_a^{x,(1)}(\sigma) \leq 1 - \alpha$. So, it can be verified that $\left|\widehat{P}_a^{x,(2)}(\sigma)\right| \leq 1$ always holds, which implies that item 3 also holds. To prove the remaining items, we need

$$\left\|\left|P_a^x - P_a^{x,(2)}\right|\right\|_2 = \sqrt{\sum_\sigma \left(\widehat{P}_a^x(\sigma) - \widehat{P}_a^{x,(2)}(\sigma)\right)^2} < \sqrt{\sum_\sigma \alpha^2} \leq m^D \alpha.$$

We can apply the same operations to $\{Q_b^y\}$ and get $\left\{Q_b^{y,(2)}\right\}$. Then for all x, y, a, b,

$$\begin{aligned}
&\left|\mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes Q_b^{y,(2)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^x \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right)\right| \\
\leq &\left|\mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes Q_b^{y,(2)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right)\right| \\
&+ \left|\mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^x \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right)\right| \\
= &\left|\mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes \left(Q_b^{y,(2)} - Q_b^y\right)\right)\psi_{AB}^{\otimes n}\right)\right| + \left|\mathrm{Tr}\left(\left(\left(P_a^{x,(2)} - P_a^x\right) \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right)\right| \\
\leq &\left\|\left|P_a^{x,(2)}\right|\right\|_2 \left\|\left|Q_b^{y,(2)} - Q_b^y\right|\right\|_2 + \left\|\left|P_a^{x,(2)} - P_a^x\right|\right\|_2 \left\|\left|Q_b^y\right|\right\|_2 \leq 2m^D\alpha = \delta,
\end{aligned}$$

and item 4 follows. Then item 5 follows from Fact 56. ◀

A.5 Lemmas for the Answer Reduction of MIP*

This section introduces several lemmas to prove the hardness of $\mathrm{MIP}^*(\mathrm{poly}, O(1))$. We use the following notations for approximation in this section and Section 6.

- For complex numbers a and b, we write $a \approx_\delta b$ if $|a - b| \leq \delta$.
- With respect to a distribution D on $\mathcal{X}$ and state $|\psi\rangle$, we write

$$A_a^x \approx_\delta B_a^x \quad \text{if} \quad \mathop{\mathbb{E}}_{x \sim D} \sum_{a \in \mathcal{A}} \left\|\left(A_a^x - B_a^x\right)|\psi\rangle\right\|^2 \leq \delta.$$

- With respect to a distribution D on $\mathcal{X}$ and state $|\psi\rangle$, we write

$$A_a^x \simeq_\delta B_a^x \quad \text{if} \quad \mathop{\mathbb{E}}_{x \sim D} \sum_{a \in \mathcal{A}} \langle\psi| A_a^x \otimes B_a^x |\psi\rangle \geq 1 - \delta.$$

In the rest of the section, the distribution on $\mathcal{X}$ is implicit.

▶ **Lemma 69** (Fact 4.13 of [39]). *Let $\{A_a^x\}$ and $\{B_a^x\}$ be POVM measurements. If $A_a^x \otimes \mathbb{1} \simeq_\delta \mathbb{1} \otimes B_a^x$, then $A_a^x \otimes \mathbb{1} \approx_{2\delta} \mathbb{1} \otimes B_a^x$.*

▶ **Lemma 70.** *Suppose $\{A_a^x\}$ and $\{B_a^x\}$ are two measurements such that one of them is projective, and that*

$$A_a^x \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes B_a^x$$

with respect to some distribution D of x and the quantum state $|\psi\rangle$. Then

$$\left| \mathop{\mathbb{E}}_x \sum_a \langle\psi| A_a^x \otimes \mathbb{1} - \mathbb{1} \otimes B_a^x |\psi\rangle \right| \leq 2\sqrt{\delta}.$$

Proof of Lemma 70. We assume $\{A_a^x\}$ is projective. Then

$$\mathop{\mathbb{E}}_x \sum_a \langle\psi| \mathbb{1} \otimes B_a^x |\psi\rangle \geq \mathop{\mathbb{E}}_x \sum_a \langle\psi| \mathbb{1} \otimes (B_a^x)^2 |\psi\rangle \geq 0,$$

which implies that

$$|\mathop{\mathbb{E}}_x \sum_a \langle\psi| A_a^x \otimes \mathbb{1} |\psi\rangle - \langle\psi| \mathbb{1} \otimes B_a^x |\psi\rangle| \leq |\mathop{\mathbb{E}}_x \sum_a \langle\psi| A_a^x \otimes \mathbb{1} |\psi\rangle - \langle\psi| \mathbb{1} \otimes (B_a^x)^2 |\psi\rangle|.$$

We can bound the second quantity in two steps.

$$\begin{aligned}
&|\mathop{\mathbb{E}}_x \sum_a \langle\psi| A_a^x \otimes \mathbb{1} |\psi\rangle - \langle\psi| A_a^x \otimes B_a^x |\psi\rangle| \\
&\leq \sqrt{\mathop{\mathbb{E}}_x \sum_a \|A_a^x |\psi\rangle\|^2} \sqrt{\mathop{\mathbb{E}}_x \sum_a \|(A_a^x \otimes \mathbb{1} - \mathbb{1} \otimes B_a^x) |\psi\rangle\|^2} \leq \sqrt{\delta},
\end{aligned}$$

and similarly

$$|\mathop{\mathbb{E}}_x \sum_a \langle\psi| A_a^x \otimes B_a^x |\psi\rangle - \langle\psi| \mathbb{1} \otimes (B_a^x)^2 |\psi\rangle| \leq \sqrt{\delta}.$$

By the triangle inequality, the second quantity is at most $2\sqrt{\delta}$. So is the first one. ◀

▶ **Lemma 71** (Fact 4.14 of [39]). *Suppose $\{A_a^x\}$ and $\{B_a^x\}$ are two measurements such that $A_a^x \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes B_a^x$. Suppose that either A or B is a projective measurement and the other is a POVM measurement. Then $A_a^x \otimes \mathbb{1} \simeq_{\sqrt{\delta}} \mathbb{1} \otimes B_a^x$.*

▶ **Lemma 72** (Proposition 4.26 of [29]). *Let $\left\{C_{a,b}^x\right\} \subseteq \mathcal{L}(\mathcal{H})$ be a set of matrices such that $\sum_b (C_{a,b}^x)^\dagger C_{a,b}^x \leq \mathbb{1}$ for all x and a. Then*

$$A_a^x \approx_\delta B_a^x \quad \textit{implies that} \quad C_{a,b}^x A_a^x \approx_\delta C_{a,b}^x B_a^x.$$

▶ **Lemma 73** (Proposition 4.28 of [29]). *Suppose $A_i = \{(A_i)_a^x\}$ be a set of matrices such that $(A_i)_a^x \approx_{\delta_i} (A_{i+1})_a^x$ for $i \in [k+1]$. Then*

$$(A_1)_a^x \approx_{k(\delta_1+\ldots+\delta_k)} (A_{k+1})_a^x.$$

▶ **Lemma 74** (Fact 4.33 of [39]). *Let $k \geq 0$ be a constant. Let $\left\{A^x_{a_1,\ldots,a_k}\right\}$ be a projective measurement. For $1 \leq j \leq k$, let $\left\{(B_j)^x_{a_j}\right\}$ be a projective measurement, and suppose that*

$$A^x_{a_j} \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes (B_j)^x_{a_j}.$$

Define the POVM measurement $\left\{J^x_{a_1,\ldots,a_k}\right\}$ as

$$J^x_{a_1,\ldots,a_k} = (B_k)^x_{a_k} \ldots (B_2)^x_{a_2}(B_1)^x_{a_1}(B_2)^x_{a_2} \ldots (B_k)^x_{a_k}.$$

Then

$$A^x_{a_1,\ldots,a_k} \otimes \mathbb{1} \approx_{(2k-1)^2\delta} \mathbb{1} \otimes J^x_{a_1,\ldots,a_k}.$$

Proof of Lemma 74. We start with

$$A^x_{a_1,\ldots,a_k} = A^x_{a_k} \cdots A^x_{a_2} A^x_{a_1} A^x_{a_2} \cdots A^x_{a_k}.$$

Because $A^x_{a_k} \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes (B_k)^x_{a_k}$, To apply Lemma 72, we can set $C^x_{a,b} = A^x_{a_k} \cdots A^x_{a_2} A^x_{a_1} A^x_{a_2} \cdots A^x_{a_{k-1}} \otimes \mathbb{1}$ with $a = a_k$ and $b = (a_1, \ldots, a_{k-1})$. Then $\sum_b (C^x_{a,b})^\dagger C^x_{a,b} \leq \mathbb{1}$. Hence by Lemma 72

$$A^x_{a_1,\ldots,a_k} \otimes \mathbb{1} \approx_\delta A^x_{a_k} \cdots A^x_{a_2} A^x_{a_1} A^x_{a_2} \cdots A^x_{a_{k-1}} \otimes (B_k)^x_{a_k}$$

We can apply Lemma 72 again with $C^x_{a,b} = A^x_{a_k} \cdots A^x_{a_2} A^x_{a_1} A^x_{a_2} \cdots A^x_{a_{k-2}} \otimes B^{(a_k)}_k$ with $a = a_{k-1}$ and $b = (a_1, \ldots, a_{k-2}, a_k)$. Because $A^x_{a_{k-1}} \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes (B_{k-1})^{(a_{k-1})}$, we can get that

$$A^x_{a_k} \cdots A^x_{a_2} A^x_{a_1} A^x_{a_2} \cdots A^x_{a_{k-1}} \otimes (B_k)^x_{a_k} \approx_\delta A^x_{a_k} \cdots A^x_{a_2} A^x_{a_1} A^x_{a_2} \cdots A^x_{a_{k-2}} \otimes (B_k)^x_{a_k} (B_{k-1})^x_{a_{k-1}}.$$

Continuing similarly, we can get that

$$A^x_{a_k} \cdots A^x_{a_2} A^x_{a_1} \otimes (B_k)^x_{a_k} \cdots (B_2)^x_{a_2} \approx_\delta A^x_{a_k} \cdots A^x_{a_2} \otimes (B_k)^x_{a_k} \cdots (B_1)^x_{a_1}.$$

With another $(k-2)$ steps we can get that

$$A^x_{a_k} \otimes (B_k)^x_{a_k} \cdots (B_2)^x_{a_2}(B_1)^x_{a_1}(B_2)^x_{a_2} \cdot (B_{k-1})^x_{a_{k-1}} \approx_\delta \mathbb{1} \otimes (B_k)^x_{a_k} \cdots (B_2)^x_{a_2}(B_1)^x_{a_1}(B_2)^x_{a_2} \cdot (B_k)^x_{a_k}.$$

Combining all the steps above with Lemma 73

$$A^x_{a_1,\ldots,a_k} \otimes \mathbb{1} \approx_{(2k-1)^2\delta} \mathbb{1} \otimes (B_k)^x_{a_k} \cdots (B_2)^x_{a_2}(B_1)^x_{a_1}(B_2)^x_{a_2} \cdot (B_k)^x_{a_k},$$

which completes the proof. ◀

▶ **Lemma 75** (Fact 4.35 of [39]). *Let $k \geq 0$ be a constant. Let D be a distribution on questions $(x, y_1, \ldots, y_k)$, where each $y_i \in \mathcal{Y}_i$. For each $1 \leq i \leq k$, let $\mathcal{G}_i$ be a set of functions $g_i : \mathcal{Y}_i \to \mathcal{R}_i$, and let $\left\{(G_i)^x_g \mid g \in \mathcal{G}_i\right\}$ be a projective measurement. Suppose that the set $\mathcal{G}_i$ has the following distance property: fix a question $z = (x, y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_k)$, and let D_z be the distribution on y_i conditioned on z. Then for any two nonequal $g_i, g'_i \in \mathcal{G}_i$, the probability that $g_i(\boldsymbol{y}_i) = g'_i(\boldsymbol{y}_i)$, over a random $\boldsymbol{y}_i \sim D_z$, is at most ε.*

Let $\left\{A^{x,y_1,\ldots,y_k}_{a_1,\ldots,a_k}\right\}$ be a projective measurement with outcomes $a_i \in \mathcal{R}_i$. For each $1 \leq i \leq k$, suppose that

$$A^{x,y_1,\ldots,y_k}_{a_i} \otimes \mathbb{1} \simeq_\delta \mathbb{1} \otimes (G_i)^x_{[g_i(y_i)=a_i]} \tag{10}$$

$$(G_i)^x_{[g_i(y_i)=a_i]} \otimes \mathbb{1} \simeq_\delta \mathbb{1} \otimes A^{x,y_1,\ldots,y_k}_{a_i}. \tag{11}$$

Also suppose that

$$A_{a_i}^{x,y_1,\ldots,y_k} \otimes \mathbb{1} \simeq_\delta \mathbb{1} \otimes A_{a_i}^{x,y_1,\ldots,y_k}. \tag{12}$$

Define the POVM $\left\{J^x_{g_1,\ldots,g_k}\right\}$ *as*

$$J^x_{g_1,\ldots,g_k} := (G_k)^x_{g_k} \cdots (G_2)^x_{g_2} \cdot (G_1)^x_{g_1} \cdot (G_2)^x_{g_2} \cdots (G_k)^x_{g_k}.$$

Then

$$A_{a_1,\ldots,a_k}^{x,y_1,\ldots,y_k} \otimes \mathbb{1} \approx_{\mathrm{poly}(\exp(k),\delta^{1/4k},\varepsilon^{1/2k})} \mathbb{1} \otimes J^x_{[g_1(y_1),\ldots,g_k(y_k)=a_1,\ldots,a_k]}.$$

This proof is the same as the original one, but we rewrite it to keep better track of the approximation errors.

The original proof. We first show the $k = 2$ case. Notice that

$$J^{x,y_1,y_2}_{[g_1(y_1),g_2(y_2)=a_1,a_2]} = \sum_{g_2:g_2(y_2)=a_2} (G_2)^x_{g_2} \left(\sum_{g_1:g_1(y_1)=a_1} (G_1)^x_{g_1} \right) (G_2)^x_{g_2}.$$

Our goal is to bound

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \langle\psi| A^{x,y_1,y_2}_{a_1,a_2} \otimes J^{x,y_1,y_2}_{[g_1(y_1),g_2(y_2)=a_1,a_2]} |\psi\rangle \\
&= \mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \langle\psi| A^{x,y_1,y_2}_{a_1,a_2} \otimes \sum_{g_2:g_2(y_2)=a_2} (G_2)^x_{g_2} (G_1)^x_{[g_1(y_1)=a_1]} (G_2)^x_{g_2} |\psi\rangle \\
&= \mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| A^{x,y_1,y_2}_{a_1,g_2(y_2)} \otimes (G_2)^x_{g_2} (G_1)^x_{[g_1(y_1)=a_1]} (G_2)^x_{g_2} |\psi\rangle .
\end{aligned}$$

First notice that

$$\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| A^{x,y_1,y_2}_{a_1,g_2(y_2)} \otimes (G_2)^x_{g_2} (G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle \approx_{2\sqrt{2\delta}} \mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \langle\psi| A^{x,y_1,y_2}_{a_1,a_2} \otimes \mathbb{1} |\psi\rangle = 1.$$

This is because

$$\begin{aligned}
&\Big|\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| A^{x,y_1,y_2}_{a_1,g_2(y_2)} \otimes (G_2)^x_{g_2} |\psi\rangle - \langle\psi| A^{x,y_1,y_2}_{a_1,g_2(y_2)} \otimes (G_2)^x_{g_2} (G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle\Big| \\
&= \Big|\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| A^{x,y_1,y_2}_{a_1,g_2(y_2)} \otimes (G_2)^x_{g_2} (A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} - \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}) |\psi\rangle\Big| \\
&\le \sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \|A^{x,y_1,y_2}_{a_1,g_2(y_2)} \otimes (G_2)^x_{g_2} |\psi\rangle\|^2} \cdot \\
&\quad \sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| (A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} - \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}) A^{x,y_1,y_2}_{a_1,g_2(y_2)} (A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} - \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}) |\psi\rangle} \\
&\le \sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \|A^{x,y_1,y_2}_{a_1,g_2(y_2)} \otimes (G_2)^x_{g_2} |\psi\rangle\|^2} \cdot \\
&\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1} \langle\psi| (A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} - \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}) \sum_{g_2} A^{x,y_1,y_2}_{a_1,g_2(y_2)} (A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} - \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}) |\psi\rangle} \\
&\le 1 \cdot \sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1} \|(A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} - \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}) |\psi\rangle\|^2} \\
&\le \sqrt{2\delta}
\end{aligned}$$

and

$$
\begin{aligned}
&|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\langle\psi|\,A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes(G_2)^x_{g_2}\,|\psi\rangle-\langle\psi|\,A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes\mathbb{1}\,|\psi\rangle|\\
&=|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,A^{x,y_1,y_2}_{a_1,a_2}\cdot(\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}-A^{x,y_1,y_2}_{a_2}\otimes\mathbb{1})\,|\psi\rangle|\\
&\le\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\|A^{x,y_1,y_2}_{a_1,a_2}\,|\psi\rangle\|^2}\cdot\\
&\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,(\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}-A^{x,y_1,y_2}_{a_2}\otimes\mathbb{1})A^{x,y_1,y_2}_{a_1,a_2}(\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}-A^{x,y_1,y_2}_{a_2}\otimes\mathbb{1})\,|\psi\rangle}\\
&\le\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\|A^{x,y_1,y_2}_{a_1,a_2}\,|\psi\rangle\|^2}\cdot\\
&\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_2}\langle\psi|\,(\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}-A^{x,y_1,y_2}_{a_2}\otimes\mathbb{1})\sum_{a_1}A^{x,y_1,y_2}_{a_1,a_2}(\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}-A^{x,y_1,y_2}_{a_2}\otimes\mathbb{1})\,|\psi\rangle}\\
&\le 1\cdot\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_2}\|(\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}-A^{x,y_1,y_2}_{a_2}\otimes\mathbb{1})\,|\psi\rangle\|^2}\\
&\le\sqrt{2\delta},
\end{aligned}
$$

Hence, we focus on proving

$$
\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\|\mathbb{1}\otimes\left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}-(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}\right)|\psi\rangle\|^2\le C_1\sqrt{\delta}+C_2\varepsilon \tag{13}
$$

for some constants C_1 and C_2, which will imply that

$$
\begin{aligned}
&|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\langle\psi|\,A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes(G_2)^x_{g_2}\left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}-(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}\right)|\psi\rangle|\\
&\le\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\|A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes(G_2)^x_{g_2}\,|\psi\rangle\|^2}\cdot\\
&\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\|\langle\psi|\,\mathbb{1}\otimes\left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}-(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}\right)|\psi\rangle\|^2}\\
&\le\sqrt{C_1\sqrt{\delta}+C_2\varepsilon}
\end{aligned}
$$

and

$$
|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,A^{x,y_1,y_2}_{a_1,a_2}\otimes J^{x,y_1,y_2}_{[g_1(y_1),g_2(y_2)=a_1,a_2]}\,|\psi\rangle-1|\le 2\sqrt{2\delta}+\sqrt{C_1\sqrt{\delta}+C_2\varepsilon}.
$$

To prove Equation (13), we start with Equation (10)

$$
\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_i}\|(A^{x,y_1,y_2}_{a_i}\otimes\mathbb{1}-\mathbb{1}\otimes(G_i)^x_{[g_i(y_i)=a_i]})\,|\psi\rangle\|^2\le 2\delta
$$

for $i = 1, 2$. Then by Lemma 72

$$\begin{aligned}
&\mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} |\psi\rangle \\
&\approx_{2\delta} A^{x,y_1,y_2}_{a_2} \otimes (G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle \\
&\approx_{2\delta} A^{x,y_1,y_2}_{a_2} A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} |\psi\rangle \\
&= A^{x,y_1,y_2}_{a_1} A^{x,y_1,y_2}_{a_1} \otimes \mathbb{1} |\psi\rangle \\
&\approx_{2\delta} A^{x,y_1,y_2}_{a_2} \otimes (G_2)^x_{[g_2(y_2)=a_2]} |\psi\rangle \\
&\approx_{2\delta} \mathbb{1} \otimes (G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle .
\end{aligned}$$

Chaining the inequalities together using Lemma 73 gives

$$\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \left\| \mathbb{1} \otimes \left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} - (G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} \right) |\psi\rangle \right\|^2 \leq 32\delta.$$

Let

$$\begin{aligned}
S_1 &= \mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g} \left\| \mathbb{1} \otimes \left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2} - (G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]} \right) |\psi\rangle \right\|^2 \\
S_2 &= \mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \left\| \mathbb{1} \otimes \left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} - (G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} \right) |\psi\rangle \right\|^2.
\end{aligned}$$

We are going to show that S_1 is close to S_2. Expanding $S_1 - S_2$, we get $|S_1 - S_2| \leq \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$, where

$$\begin{aligned}
\Delta_1 = \Big| &\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| \mathbb{1} \otimes (G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2} |\psi\rangle \\
&- \sum_{a_1,a_2} \langle\psi| \mathbb{1} \otimes (G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} |\psi\rangle \Big| \\
\Delta_2 = \Big| &\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle \\
&- \sum_{a_1,a_2} \langle\psi| \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]}(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle \Big| \\
\Delta_3 = \Big| &\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| \mathbb{1} \otimes (G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle \\
&- \sum_{a_1,a_2} \langle\psi| \mathbb{1} \otimes (G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]} |\psi\rangle \Big| \\
\Delta_4 = \Big| &\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2} |\psi\rangle \\
&- \sum_{a_1,a_2} \langle\psi| \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} |\psi\rangle \Big|.
\end{aligned}$$

First of all

$$\Delta_1 = \Big| 1 - \mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \langle\psi| \mathbb{1} \otimes (G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} |\psi\rangle \Big|.$$

By Equation (11),

$$\mathbb{1} \otimes (G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]} |\psi\rangle \approx_{18\delta} A^{x,y_1,y_2}_{a_1,a_2} \otimes \mathbb{1} |\psi\rangle ,$$

then Lemma 70 implies that

$$|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,A^{x,y_1,y_2}_{a_1,a_2}\otimes\mathbb{1}\,|\psi\rangle-\langle\psi|\,\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]}\,|\psi\rangle|\le 6\sqrt{2\delta}.$$

Since $\mathbb{E}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,\mathbb{1}\otimes A^{x,y_1,y_2}_{a_1,a_2}\,|\psi\rangle=1$, $\Delta_1\le 6\sqrt{2\delta}$. Next, observe that $\Delta_2=0$ as $(G_2)^x_{g_2}$ and $(G_2)^x_{[g_2(y_2)=a_2]}$ are projective measurements. Lastly, observe that $\Delta_3=\Delta_4$, so we focus on bounding Δ_3. First notice that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\langle\psi|\,\mathbb{1}\otimes(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}\,|\psi\rangle\\
\approx_{3\sqrt{2\delta}}&\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\langle\psi|\,(G_1)^x_{[g_1(y_1)=a_1]}\otimes(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}\,|\psi\rangle\\
&\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,\mathbb{1}\otimes(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}\,|\psi\rangle\\
\approx_{3\sqrt{2\delta}}&\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,(G_1)^x_{[g_1(y_1)=a_1]}\otimes(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]}\,|\psi\rangle
\end{aligned}$$

The reason why $\mathbb{1}\otimes(G_1)^x_{[g_1(y_1)=a_1]}\approx_{18\delta}(G_1)^x_{[g_1(y_1)=a_1]}\otimes\mathbb{1}$ is the following. Applying Lemma 69 to Equations (10) and (11) we get

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y_1,\ldots,y_k}\sum_{a_i}\|(A^{x,y_1,\ldots,y_k}_{a_i}\otimes\mathbb{1}-\mathbb{1}\otimes(G_i)^x_{[g_i(y_i)=a_i]})\,|\psi\rangle\|^2\le 2\delta\\
&\mathop{\mathbb{E}}_{x,y_1,\ldots,y_k}\sum_{a_i}\|((G_i)^x_{[g_i(y_i)=a_i]}\otimes\mathbb{1}-\mathbb{1}\otimes A^{x,y_1,\ldots,y_k}_{a_i})\,|\psi\rangle\|^2\le 2\delta.
\end{aligned}$$

Notice that for any $i\in[k]$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y_1,\ldots,y_k}\sum_{a_i}\langle\psi|\,A^{x,y_1,\ldots,y_k}_{a_i}\otimes A^{x,y_1,\ldots,y_k}_{a_i}\,|\psi\rangle\\
&\qquad\ge\mathop{\mathbb{E}}_{x,y_1,\ldots,y_k}\sum_{a_1,\ldots,a_k}\langle\psi|\,A^{x,y_1,\ldots,y_k}_{a_1,\ldots,a_k}\otimes A^{x,y_1,\ldots,y_k}_{a_1,\ldots,a_k}\,|\psi\rangle\ge 1-\delta
\end{aligned}$$

because $A^{x,y_1,\ldots,y_k}_{a_1,\ldots,a_k}\otimes A^{x,y_1,\ldots,y_k}_{b_1,\ldots,b_k}\ge 0$ for any $a_1,\ldots,a_k,b_1,\ldots,b_k$. Then Lemma 69 also implies that

$$\mathop{\mathbb{E}}_{x,y_1,\ldots,y_k}\sum_{a_1}\|(A^{x,y_1,\ldots,y_k}_{a_i}\otimes\mathbb{1}-\mathbb{1}\otimes A^{x,y_1,\ldots,y_k}_{a_i})\,|\psi\rangle\|^2\le 2\delta.$$

Hence, Lemma 73 implies that for all $i\in[k]$.

$$\mathop{\mathbb{E}}_{x,y_1,\ldots,y_k}\sum_{a_i}\|((G_i)^x_{[g_i(y_i)=a_i]}\otimes\mathbb{1}-\mathbb{1}\otimes(G_i)^x_{[g_i(y_i)=a_i]})\,|\psi\rangle\|^2\le 18\delta.$$

Also, notice that

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\langle\psi|\,(G_2)^x_{[g_2(y_2)=a_2]}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{[g_2(y_2)=a_2]}\otimes(G_1)^x_{[g_1(y_1)=a_1]}\,|\psi\rangle\\
&-\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\langle\psi|\,(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}\otimes(G_1)^x_{[g_1(y_1)=a_1]}\,|\psi\rangle|\\
&=|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1}\sum_{g_2,g_2'}\langle\psi|\,(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2'}\otimes(G_1)^x_{[g_1(y_1)=a_1]}\,|\psi\rangle\,\mathbb{1}[g_2(y_2)=g_2'(y_2)]|\\
&\le\varepsilon|\mathop{\mathbb{E}}_{x,y_1}\sum_{a_1}\langle\psi|\,(G_1)^x_{[g_1(y_1)=a_1]}\otimes(G_1)^x_{[g_1(y_1)=a_1]}\,|\psi\rangle|\\
&\le\varepsilon.
\end{aligned}$$

Therefore, $\Delta_3 = \Delta_4 \leq 6\sqrt{2\delta} + \varepsilon$, and

$$|S_1 - S_2| \leq \sum_{j=1}^{4} \Delta_j \leq 18\sqrt{2\delta} + 2\varepsilon,$$

and

$$S_1 \leq 32\delta + 18\sqrt{2\delta} + 2\varepsilon.$$

In conclusion,

$$\begin{aligned}
&\Big|\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \langle\psi| A^{x,y_1,y_2}_{a_1,a_2} \otimes J^{x,y_1,y_2}_{[g_1(y_1),g_2(y_2)=a_1,a_2]} |\psi\rangle - 1\Big| \\
&\leq 2\sqrt{2\delta} + \sqrt{32\delta + 18\sqrt{2\delta} + 2\varepsilon} \leq 11\delta^{1/4} + 2\sqrt{\varepsilon},
\end{aligned}$$

and equivalently

$$A^{x,y_1,y_2}_{a_1,a_2} \otimes \mathbb{1} \approx_{22\delta^{1/4}+4\sqrt{\epsilon}} \mathbb{1} \otimes J^{x,y_1,y_2}_{[g_1(y_1),g_2(y_2)=a_1,a_2]}.$$

Switching the roles of Alice and Bob, the same proof gives us that

$$J^{x,y_1,y_2}_{[g_1(y_1),g_2(y_2)=a_1,a_2]} \otimes \mathbb{1} \approx_{22\delta^{1/4}+4\sqrt{\epsilon}} \mathbb{1} \otimes A^{x,y_1,y_2}_{a_1,a_2}.$$

For the general case, assume

$$\begin{aligned}
&A^{x,y_1,\ldots,y_i}_{a_1,\ldots,a_i} \otimes \mathbb{1} \approx_{f(i,\delta,\varepsilon)} \mathbb{1} \otimes J^{x}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]} \text{ and} \\
&\mathbb{1} \otimes A^{x,y_1,\ldots,y_i}_{a_1,\ldots,a_i} \approx_{f(i,\delta,\varepsilon)} J^{x}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]} \otimes \mathbb{1},
\end{aligned}$$

which imply that

$$\mathbb{1} \otimes J^{x}_{[g_1(y_1),\ldots,g_i(y_i)} \approx_{3(2\delta+2f(i,\delta,\varepsilon))} J^{x}_{[g_1(y_1),\ldots,g_i(y_i)} \otimes \mathbb{1}.$$

Since δ and ε are fixed, we write $f(i,\delta,\varepsilon)$ as $f(i)$ in the rest of the proof and proceed to the $i+1$ case. As in the base case, our goal is to bound

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_{i+1}} \langle\psi| A^{x,y_1,\ldots,y_{i+1}}_{a_1,\ldots,a_{i+1}} \otimes J^{x,y_1,\ldots,y_{i+1}}_{[g_1(y_1),\ldots,g_{i+1}(y_{i+1})=a_1,\ldots,a_{i+1}]} |\psi\rangle \\
&= \mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_i,g_{i+1}} \langle\psi| A^{x,y_1,\ldots,y_{i+1}}_{a_1,\ldots,a_i,g_{i+1}(y_{i+1})} \otimes (G_{i+1})^x_{g_{i+1}} J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]} (G_{i+1})^x_{g_{i+1}} |\psi\rangle.
\end{aligned}$$

by relating it to

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_i,g_{i+1}} \langle\psi| A^{x,y_1,\ldots,y_{i+1}}_{a_1,\ldots,a_i,g_{i+1}(y_{i+1})} \otimes (G_{i+1})^x_{g_{i+1}} J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]} |\psi\rangle \\
&\approx_{\sqrt{2\delta}+\sqrt{f(i)}} \mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_{i+1}} \langle\psi| A^{x,y_1,\ldots,y_{i+1}}_{a_1,\ldots,a_{i+1}} \otimes \mathbb{1} |\psi\rangle = 1.
\end{aligned}$$

So the central step is bounding

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_i,g_{i+1}} \Big\| \mathbb{1} \otimes \Big(& J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]} (G_{i+1})^x_{g_{i+1}} \\
& - (G_{i+1})^x_{g_{i+1}} J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]} \Big) |\psi\rangle \Big\|^2.
\end{aligned}$$

As in the base case, we can use similar arguments to show

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_i,g_{i+1}} \|\mathbb{1}\otimes\Big(&J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}(G_{i+1})^x_{[g_{i+1}(y_{i+1})=a_{i+1}]}\\ &-(G_{i+1})^x_{[g_{i+1}(y_{i+1})=a_{i+1}]}J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}\Big)|\psi\rangle\|^2\\ \leq\ & 4(2f(i)+4\delta),\end{aligned}$$

and

$$\begin{aligned}\Big|\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_i,g_{i+1}} \|\mathbb{1}\otimes\Big(&J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}(G_{i+1})^x_{g_{i+1}}\\ &-(G_{i+1})^x_{g_{i+1}}J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}\Big)|\psi\rangle\|^2-\\ \mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_i,g_{i+1}} \|\mathbb{1}\otimes\Big(&J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}(G_{i+1})^x_{[g_{i+1}(y_{i+1})=a_{i+1}]}\\ &-(G_{i+1})^x_{[g_{i+1}(y_{i+1})=a_{i+1}]}J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}\Big)|\psi\rangle\|^2\Big|\\ \leq\ & 2\sqrt{2f(i)+4\delta}+2\sqrt{6f(i)+4\delta}+2\varepsilon.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_i,g_{i+1}} \|\mathbb{1}\otimes\Big(&J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}(G_{i+1})^x_{g_{i+1}}\\ &-(G_{i+1})^x_{g_{i+1}}J^{x,y_1,\ldots,y_i}_{[g_1(y_1),\ldots,g_i(y_i)=a_1,\ldots,a_i]}\Big)|\psi\rangle\|^2\\ \leq\ & 4(2f(i)+4\delta)+2\sqrt{2f(i)+4\delta}+2\sqrt{6f(i)+4\delta}+2\varepsilon,\end{aligned}$$

and

$$\begin{aligned}&\Big|\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_{i+1}} \langle\psi|A^{x,y_1,\ldots,y_{i+1}}_{a_1,\ldots,a_{i+1}}\otimes J^{x,y_1,\ldots,y_{i+1}}_{[g_1(y_1),\ldots,g_{i+1}(y_{i+1})=a_1,\ldots,a_{i+1}]}|\psi\rangle-1\Big|\\ &\leq\sqrt{2\delta}+\sqrt{f(i)}+\sqrt{16\sqrt{f(i)}+24\sqrt{\delta}+2\varepsilon}\end{aligned}$$

That is $f(i+1)=5f(i)^{1/4}+7\delta^{1/4}+\sqrt{2\varepsilon}$. Then the lemma follows. ◀

B Proofs of Theorems

B.1 Invariance Principle for Matrix Spaces

▶ **Fact 76** ([37, Remark 3.10]). *If* $\mathbf{x}$ *is* (p,q,η)*-hypercontractive, then it is* (p,q,η')*-hypercontractive for any* $0<\eta'\leq\eta$.

▶ **Lemma 77.** *Given* $m,n\in\mathbb{Z}_{>0}$, $0<\eta<1$, *a* $(2,4,\eta)$*-hypercontractive* (m,n) *ensemble* $\mathbf{x}$, *it holds that*

$$\mathbb{E}\left[\left(\sum_{i=1}^{k}\left(T_\eta p_i\right)(\mathbf{x})^2\right)^2\right]\leq\left(\mathbb{E}\left[\sum_{i=1}^{k}p_i(\mathbf{x})^2\right]\right)^2,$$

for any multilinear polynomials $p_1,\ldots p_k$.

Proof of Lemma 77. Let $q_i = T_\eta p_i$. Then

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{k}\left(T_\eta p_i\right)(\mathbf{x})^2\right)^2\right] &= \sum_{i,j}\mathbb{E}\left[q_i(\mathbf{x})^2 q_j(\mathbf{x})^2\right] \\
&\leq \sum_{i,j}\|q_i\|_4^2\|q_j\|_4^2 \qquad \text{(Cauchy-Schwarz inequality)} \\
&\leq \sum_{i,j}\|p_i\|_2^2\|p_j\|_2^2 \qquad (\mathbf{x} \text{ is } (2,4,\eta)\text{-hypercontractive}) \\
&= \left(\sum_i \|p_i\|_2^2\right)^2 = \left(\mathbb{E}\left[\sum_{i=1}^{k} p_i(\mathbf{x})^2\right]\right)^2. \qquad \blacktriangleleft
\end{aligned}$$

The lemma below follows directly from Definition 10 and Fact 48.

▶ **Lemma 78.** *Given $0 \leq \gamma \leq 1$, $h, n, m \in \mathbb{Z}_{>0}$, $m \geq 2$, an (m^2, n) ensemble $\mathbf{x}$, and a random matrix*

$$P(\mathbf{x}) = \sum_{\sigma \in [m^2]^h_{\geq 0}} p_\sigma(\mathbf{x})\,\mathcal{B}_\sigma,$$

where $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$ is a standard orthonormal basis and p_σ is a real multilinear polynomial for all $\sigma \in [m^2]^h_{\geq 0}$, suppose that for all $\sigma \in [m^2]^h_{\geq 0}$, p_σ has an expansion

$$p_\sigma(\mathbf{x}) = \sum_{\tau \in [m^2]^n_{\geq 0}} \widehat{p_\sigma}(\tau)\mathbf{x}_\tau.$$

It holds that

$$\Gamma_\gamma(P(\mathbf{x})) = \sum_{\sigma \in [m^2]^h_{\geq 0}} \sum_{\tau \in [m^2]^n_{\geq 0}} \gamma^{|\sigma|+|\tau|}\widehat{p_\sigma}(\tau)\mathbf{x}_\tau \mathcal{B}_\sigma. \tag{14}$$

We need the hypercontractivity inequality for Hermitian matrices.

▶ **Fact 79** ([46, Lemma 8.3])**.** *Given $h, n, m \in \mathbb{Z}_{>0}$, $m \geq 2$, $0 \leq \gamma \leq (9m)^{-1/4}$ and $P \in \mathcal{H}_m^{\otimes n}$, it holds that*

$$\left|\left|\left|\Delta_\gamma^{\otimes n}(P)\right|\right|\right|_4 \leq |||P|||_2,$$

where $\Delta_\gamma(\cdot)$ is defined in Definition 47.

Proof of Theorem 13. Set $Q(\mathbf{x}) = \sum_{\sigma \in [m^2]^h_{\geq 0}} \left(T_\gamma p_\sigma\right)(\mathbf{x})\,\mathcal{B}_\sigma$. Then by the definition of Γ_γ,

$$\Gamma_\gamma(P(\mathbf{x})) = \Delta_\gamma(Q(\mathbf{x})).$$

Using Fact 79,

$$\mathbb{E}\left[|||\Delta_\gamma(Q(\mathbf{x}))|||_4^4\right] \leq \mathbb{E}\left[|||Q(\mathbf{x})|||_2^4\right]. \tag{15}$$

Denote $q_\sigma = T_\gamma p_\sigma$. Notice that

$$\begin{aligned}
\mathbb{E}\left[|||Q(\mathbf{x})|||_2^4\right] &= m^{-2h}\,\mathbb{E}\left[\left(\sum_{\sigma \in [m^2]^h_{\geq 0}} q_\sigma(\mathbf{x})^2\right)^2\right] \leq m^{-2h}\left(\mathbb{E}\left[\sum_{\sigma \in [m^2]^h_{\geq 0}} p_\sigma(\mathbf{x})^2\right]\right)^2 \\
&= \left(\mathbb{E}\left[|||P(\mathbf{x})|||_2^2\right]\right)^2,
\end{aligned}$$

where the inequality follows from Fact 76 and Lemma 77. We conclude the result by combining it with Equation (15). $\blacktriangleleft$

Proof of Theorem 14. Suppose that for all $\sigma \in [m^2]^h_{\geq 0}$, p_σ has an expansion

$$p_\sigma(\mathbf{x}) = \sum_{\tau \in [m^2]^n_{\geq 0}} \widehat{p_\sigma}(\tau)\mathbf{x}_\tau.$$

Set

$$P^{=i}(\mathbf{x}) = \sum_{\substack{\sigma \in [m^2]^h_{\geq 0}, \tau \in [m^2]^n_{\geq 0}: \\ |\sigma|+|\tau|=i}} \widehat{p_\sigma}(\tau)\,\mathbf{x}_\tau \mathcal{B}_\sigma.$$

Set $\gamma = \min\left\{\eta, (9m)^{-1/4}\right\}$. Applying Lemma 78 and Theorem 13,

$$\mathbb{E}\left[\|\!|P(\mathbf{x})|\!\|_4^4\right] = \mathbb{E}\left[\left\|\!\left|\Gamma_\gamma\left(\sum_{i=1}^d \gamma^{-i}P^{=i}(\mathbf{x})\right)\right|\!\right\|_4^4\right] \leq \left(\mathbb{E}\left[\left\|\!\left|\sum_{i=1}^d \gamma^{-i}P^{=i}(\mathbf{x})\right|\!\right\|_2^2\right]\right)^2$$

By the orthogonality of $\mathbf{x}$ and $\mathcal{B}$, if $i \neq j$, we have

$$\mathbb{E}\left[\mathrm{Tr}\ P^{=i}(\mathbf{x})P^{=j}(\mathbf{x})\right] = 0.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\|\!|P(\mathbf{x})|\!\|_4^4\right] &\leq \left(\sum_{i=1}^d \gamma^{-2i}\,\mathbb{E}\left[\|\!|P^{=i}(\mathbf{x})|\!\|_2^2\right]\right)^2 \\ &\leq \gamma^{-4d}\left(\sum_{i=1}^d \mathbb{E}\left[\|\!|P^{=i}(\mathbf{x})|\!\|_2^2\right]\right)^2 = \gamma^{-4d}\left(\mathbb{E}\left[\|\!|P(\mathbf{x})|\!\|_2^2\right]\right)^2.\end{aligned}$$

◀

Proof of Theorem 15. Without loss of generality, we assume $\overline{H} = [n-h]$. We prove this by a hybrid argument. For any $0 \leq i \leq n-h$, define the hybrid basis elements and the hybrid random operators as follows.

$$\mathcal{X}_\sigma^{(i)} = \mathbf{x}_{\sigma_{\leq i}} \cdot \mathcal{B}_{\sigma_{>i}} \text{ for } \sigma \in [m^2]^n_{\geq 0}; \tag{16}$$

$$P^{(i)}(\mathbf{x}) = \sum_{\sigma \in [m^2]^n_{\geq 0}} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i)}, \tag{17}$$

where $\mathbf{x}_{\sigma_{\leq i}} = \mathbf{x}_{\sigma_1} \cdots \mathbf{x}_{\sigma_i}$ and $\mathcal{B}_{\sigma_{>i}} = \mathcal{B}_{\sigma_{i+1}} \otimes \ldots \otimes \mathcal{B}_{\sigma_n}$. Then $P = P^{(0)}(\mathbf{x})$ and $P^H(\mathbf{x}) = P^{(n-h)}(\mathbf{x})$. Note that

$$\begin{aligned}P^{(i)}(\mathbf{x}) &= \sum_{\sigma:\sigma_{i+1}=0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i)} + \sum_{\sigma:\sigma_{i+1}\neq 0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i)}, \\ P^{(i+1)}(\mathbf{x}) &= \sum_{\sigma:\sigma_{i+1}=0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i+1)} + \sum_{\sigma:\sigma_{i+1}\neq 0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i+1)},\end{aligned}$$

Set

$$\mathbf{A} = \sum_{\sigma:\sigma_{i+1}=0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i)}; \qquad \mathbf{B} = \sum_{\sigma:\sigma_{i+1}\neq 0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i)};$$

$$\mathbf{C} = \sum_{\sigma:\sigma_{i+1}=0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i+1)}; \qquad \mathbf{D} = \sum_{\sigma:\sigma_{i+1}\neq 0} \widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i+1)}.$$

Then we have

$$P^{(i)}(\mathbf{x}) = \mathbf{A} + \mathbf{B};\ P^{(i+1)}(\mathbf{x}) = \mathbf{C} + \mathbf{D}.$$

Notice that $\mathbf{A} = \mathbb{1}_m \otimes \mathbf{C}$, where $\mathbb{1}_m$ is placed in the $(i+1)$-th register. Thus,

$$\mathrm{Tr}\ \xi(\mathbf{A}) = m \cdot \mathrm{Tr}\ \xi(\mathbf{C}). \tag{18}$$

From Fact 54 and then Equation (18),

$$\begin{aligned}
&\left| m^{i+1-n}\,\mathbb{E}\left[\mathrm{Tr}\ \xi\left(P^{(i+1)}(\mathbf{x})\right)\right] - m^{i-n}\,\mathbb{E}\left[\mathrm{Tr}\ \xi\left(P^{(i)}(\mathbf{x})\right)\right]\right| \\
&= \left|\mathbb{E}\left[\begin{array}{c} m^{i+1-n}\left(\mathrm{Tr}\ \xi(C) + \mathrm{Tr}\ D\xi(\mathbf{C})[\mathbf{D}] + \frac{1}{2}\mathrm{Tr}\ D^2\xi(\mathbf{C})[\mathbf{D}] + \Delta_{3,\xi}(\mathbf{C},\mathbf{D})\right) - \\ m^{i-n}\left(\mathrm{Tr}\ \xi(A) + \mathrm{Tr}\ D\xi(\mathbf{A})[\mathbf{B}] + \frac{1}{2}\mathrm{Tr}\ D^2\xi(\mathbf{A})[\mathbf{B}] + \Delta_{3,\xi}(\mathbf{A},\mathbf{B})\right) \end{array}\right]\right| \\
&= \left|\mathbb{E}\left[\begin{array}{c} m^{i+1-n}\left(\mathrm{Tr}\ D\xi(\mathbf{C})[\mathbf{D}] + \frac{1}{2}\mathrm{Tr}\ D^2\xi(\mathbf{C})[\mathbf{D}] + \Delta_{3,\xi}(\mathbf{C},\mathbf{D})\right) - \\ m^{i-n}\left(\mathrm{Tr}\ D\xi(\mathbf{A})[\mathbf{B}] + \frac{1}{2}\mathrm{Tr}\ D^2\xi(\mathbf{A})[\mathbf{B}] + \Delta_{3,\xi}(\mathbf{A},\mathbf{B})\right) \end{array}\right]\right|
\end{aligned}$$

Both the first-order and second-order derivatives cancel out because of the following claim.

▷ Claim 80. It holds that

$$\mathbb{E}[\mathrm{Tr}\ D\xi(\mathbf{A})[\mathbf{B}]] = m\,\mathbb{E}[\mathrm{Tr}\ D\xi(\mathbf{C})[\mathbf{D}]];$$

$$\mathbb{E}\left[\mathrm{Tr}\ D^2\xi(\mathbf{A})[\mathbf{B}]\right] = m\,\mathbb{E}\left[\mathrm{Tr}\ D^2\xi(\mathbf{C})[\mathbf{D}]\right].$$

By Fact 54, there exists a universal constant $c_3 > 0$ such that

$$\begin{aligned}
&\left|\mathbb{E}\left[m^{i+1-n}\mathrm{Tr}\ \xi\left(P^{(i+1)}(\mathbf{x})\right) - m^{i-n}\mathrm{Tr}\ \xi\left(P^{(i)}(\mathbf{x})\right)\right]\right| \\
&\le c_3 B\left(\mathbb{E}\left[|||\mathbf{B}|||_3^3\right] + \mathbb{E}\left[|||\mathbf{D}|||_3^3\right]\right) \\
&\le c_3 B\left(\mathbb{E}\left[|||\mathbf{B}|||_2|||\mathbf{B}|||_4^2\right] + \mathbb{E}\left[|||\mathbf{D}|||_2|||\mathbf{D}|||_4^2\right]\right) \quad \text{(Hölder's)} \\
&\le c_3 B\left(\left(\mathbb{E}\left[|||\mathbf{B}|||_2^2\right]\mathbb{E}\left[|||\mathbf{B}|||_4^4\right]\right)^{1/2} + \left(\mathbb{E}\left[|||\mathbf{D}|||_2^2\right]\mathbb{E}\left[|||\mathbf{D}|||_4^4\right]\right)^{1/2}\right) \quad \text{(Cauchy-Schwartz)} \\
&\le c_3 B\theta^d\left(\left(\mathbb{E}\left[|||\mathbf{B}|||_2^2\right]\right)^{3/2} + \left(\mathbb{E}\left[|||\mathbf{D}|||_2^2\right]\right)^{3/2}\right) \quad \text{(Theorem 14)},
\end{aligned}$$

where $\theta = \max\left\{9m, 1/\eta^4\right\}$. Notice that

$$\mathbb{E}\left[|||\mathbf{B}|||_2^2\right] = \mathbb{E}\left[|||\mathbf{D}|||_2^2\right] = \sum_{\sigma:\sigma_{i+1}\neq 0}\left|\widehat{P}(\sigma)^2\right| = \mathrm{Inf}_{i+1}(P).$$

Therefore,

$$\left|\mathbb{E}\left[m^{i+1-n}\mathrm{Tr}\ \xi\left(P^{(i+1)}(\mathbf{x})\right) - m^{i-n}\mathrm{Tr}\ \xi\left(P^{(i)}(\mathbf{x})\right)\right]\right| \le 2c_3 B\theta^d \mathrm{Inf}_{i+1}(P)^{3/2}.$$

Summing over $i \in [n-h]_{\ge 0}$, we have

$$\begin{aligned}
&\left|m^{-n}\mathrm{Tr}\ \xi(P) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\ \xi\left(P^H(\mathbf{x})\right)\right]\right| \\
&\le 2c_3 B\theta^d \sum_{i\notin H}\mathrm{Inf}_i(P)^{3/2} \\
&\le 2c_3 B\theta^d\sqrt{\tau}\sum_{i\notin H}\mathrm{Inf}_i(P) \\
&\le 2c_3 B\theta^d\sqrt{\tau}d\sum_{\sigma\neq 0}\widehat{P}(\sigma)^2 \\
&\le 2c_3 B\theta^d\sqrt{\tau}d.
\end{aligned}$$

◀

Proof of Claim 80. Note that $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$ can be expressed as

$$\mathbf{A} = \mathbb{1}_m \otimes \mathbf{C}; \qquad \mathbf{B} = \sum_{\sigma \in [m^2]_{\geq 0} : \sigma \neq 0} \mathcal{B}_\sigma \otimes \mathbf{X}_\sigma; \qquad \mathbf{D} = \sum_{\sigma \in [m^2]_{\geq 0} : \sigma \neq 0} \mathbf{x}_{i+1,\sigma} \mathbf{X}_\sigma$$

for some random matrices $\mathbf{X}_\sigma$'s which are independent of $\mathbf{x}_{i+1,\sigma}$'s, where $\mathbb{1}_m$ and $\mathbf{B}_\sigma$'s are in the $(i+1)$-th register.

Suppose that $\mathbf{C}$ has a spectral decomposition

$$\mathbf{C} = \sum_{j=1}^{m'} \mathbf{a}_i \mathbf{\Pi}_i,$$

where m' is the dimension of $\mathbf{C}$, $\mathbf{a}_1 \geq \cdots \geq \mathbf{a}_{m'}$, $\{\Pi_i\}_{i \in [m']}$ are rank-one projectors satisfying that $\sum_{i=1}^{m'} \Pi_i = \mathbb{1}$ and $\Pi_i \Pi_j = 0$ for all $i \neq j$.

By Fact 53, we have

$$\begin{aligned}
&\mathbb{E}[\mathrm{Tr}\ D\xi\left(\mathbf{A}\right)[\mathbf{B}]] \\
=& \sum_{j,k \in [m']} \mathbb{E}\left[\xi^{[1]}\left(\mathbf{a}_j, \mathbf{a}_k\right) \mathrm{Tr}\left(\left(\mathbb{1} \otimes \mathbf{\Pi}_j\right) \mathbf{B} \left(\mathbb{1} \otimes \mathbf{\Pi}_k\right)\right)\right] \\
=& \sum_{j,k \in [m']} \mathbb{E}\left[\xi^{[1]}\left(\mathbf{a}_j, \mathbf{a}_k\right) \mathrm{Tr}\left(\left(\mathbb{1} \otimes \mathbf{\Pi}_j \mathbf{\Pi}_k\right) \mathbf{B}\right)\right] \\
=& \sum_{j \in [m']} \mathbb{E}[\xi'\left(\mathbf{a}_j\right) \mathrm{Tr}\left(\left(\mathbb{1} \otimes \mathbf{\Pi}_j\right) \mathbf{B}\right)] \\
=& \mathbb{E}[\mathrm{Tr}\ \xi'\left(\mathbf{A}\right) \mathbf{B}] \\
=& \sum_{\sigma \in [m^2]_{\geq 0} : \sigma \neq 0} \mathbb{E}[\mathrm{Tr}\ \left(\mathbb{1}_m \otimes \xi'\left(\mathbf{C}\right)\right)\left(\mathcal{B}_\sigma \otimes \mathbf{X}_\sigma\right)] \\
=& \sum_{\sigma \in [m^2]_{\geq 0} : \sigma \neq 0} \mathbb{E}[\mathrm{Tr}\ \mathcal{B}_\sigma \cdot \mathrm{Tr}\ \xi'\left(\mathbf{C}\right) \mathbf{X}_\sigma] = 0,
\end{aligned}$$

where the last equality follows from the orthogonality of $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$.

$$\begin{aligned}
\mathbb{E}[\mathrm{Tr}\ D\xi\left(\mathbf{C}\right)[\mathbf{D}]] =& \mathbb{E}[\mathrm{Tr}\ \xi'\left(\mathbf{C}\right) \mathbf{D}] \\
=& \sum_{\sigma \in [m^2]_{\geq 0} : \sigma \neq 0} \mathbb{E}[\mathbf{x}_{i+1,\sigma} \cdot \mathrm{Tr}\ \xi'\left(\mathbf{C}\right) \mathbf{X}_\sigma] \\
=& \sum_{\sigma \in [m^2]_{\geq 0} : \sigma \neq 0} \mathbb{E}[\mathbf{x}_{i+1,\sigma}] \cdot \mathbb{E}[\mathrm{Tr}\ \xi'\left(\mathbf{C}\right) \mathbf{X}_\sigma] = 0,
\end{aligned}$$

where the last equality follows from the orthogonality of $\mathbf{x}$.

By Fact 53, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Tr}\ D^2\xi\left(\mathbf{A}\right)[\mathbf{B}]\right] \\
=& \sum_{j,k,\ell \in [m']} \mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j, \mathbf{a}_k, \mathbf{a}_\ell\right) \mathrm{Tr}\left(\left(\mathbb{1} \otimes \mathbf{\Pi}_j\right) \mathbf{B} \left(\mathbb{1} \otimes \mathbf{\Pi}_k\right) \mathbf{B} \left(\mathbb{1} \otimes \mathbf{\Pi}_\ell\right)\right)\right] \\
=& \sum_{\sigma,\tau \neq 0} \sum_{j,k,\ell \in [m']} \mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j, \mathbf{a}_k, \mathbf{a}_\ell\right) \mathrm{Tr}\left(\mathcal{B}_\sigma \mathcal{B}_\tau\right) \cdot \mathrm{Tr}\left(\mathbf{\Pi}_j \mathbf{X}_\sigma \mathbf{\Pi}_k \mathbf{X}_\tau \mathbf{\Pi}_\ell\right)\right] \\
=& \sum_{\sigma \neq 0} \sum_{j,k,\ell \in [m']} \mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j, \mathbf{a}_k, \mathbf{a}_\ell\right) \mathrm{Tr}\left(\mathbf{\Pi}_j \mathbf{X}_\sigma \mathbf{\Pi}_k \mathbf{X}_\sigma \mathbf{\Pi}_\ell\right)\right],
\end{aligned}$$

where the last equality follows from the orthogonality of $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$.

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Tr}\ D^2\xi\left(\mathbf{C}\right)\left[\mathbf{D}\right]\right] \\
&= \sum_{j,k,\ell\in[m']} \mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j,\mathbf{a}_k,\mathbf{a}_\ell\right)\mathrm{Tr}\left(\mathbf{\Pi}_j\mathbf{D}\mathbf{\Pi}_k\mathbf{D}\mathbf{\Pi}_\ell\right)\right] \\
&= \sum_{\sigma,\tau\neq 0}\sum_{j,k,\ell\in[m']} \mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j,\mathbf{a}_k,\mathbf{a}_\ell\right)\mathbf{x}_{i+1,\sigma}\mathbf{x}_{i+1,\tau}\cdot\mathrm{Tr}\left(\mathbf{\Pi}_j\mathbf{X}_\sigma\mathbf{\Pi}_k\mathbf{X}_\tau\mathbf{\Pi}_\ell\right)\right] \\
&= \sum_{\sigma,\tau\neq 0}\sum_{j,k,\ell\in[m']} \mathbb{E}[\mathbf{x}_{i+1,\sigma}\mathbf{x}_{i+1,\tau}]\,\mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j,\mathbf{a}_k,\mathbf{a}_\ell\right)\cdot\mathrm{Tr}\left(\mathbf{\Pi}_j\mathbf{X}_\sigma\mathbf{\Pi}_k\mathbf{X}_\tau\mathbf{\Pi}_\ell\right)\right] \\
&= \sum_{\sigma\neq 0}\sum_{j,k,\ell\in[m']} \mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j,\mathbf{a}_k,\mathbf{a}_\ell\right)\mathrm{Tr}\left(\mathbf{\Pi}_j\mathbf{X}_\sigma\mathbf{\Pi}_k\mathbf{X}_\sigma\mathbf{\Pi}_\ell\right)\right],
\end{aligned}$$

where the last equality follows from the orthogonality of $\mathbf{x}$. ◁

Proof of Lemma 16. Let $\lambda > 0$ be determined later, and ζ_λ be defined as in Fact 57. By Theorem 15 and Fact 57,

$$\left|m^{-n}\mathrm{Tr}\ \zeta_\lambda\left(P\right) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\ \zeta_\lambda\left(P^H(\mathbf{x})\right)\right]\right| \leq CB_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d/\lambda,$$

where C, B_3 are universal constants. By Fact 57 we also have

$$\left|m^{-n}\mathrm{Tr}\ \zeta\left(P\right) - m^{-n}\mathrm{Tr}\ \zeta_\lambda\left(P\right)\right| \leq 2\lambda^2$$

and

$$\left|m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\ \zeta\left(P^H(\mathbf{x})\right)\right] - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\ \zeta_\lambda\left(P^H(\mathbf{x})\right)\right]\right| \leq 2\lambda^2.$$

By the triangle inequality, we have

$$\left|m^{-n}\mathrm{Tr}\ \zeta\left(P\right) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\ \zeta\left(P^H(\mathbf{x})\right)\right]\right| \leq 4\lambda^2 + CB_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d/\lambda.$$

Choosing $\lambda = \left(CB_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d/8\right)^{1/3}$, we have

$$\left|m^{-n}\mathrm{Tr}\ \zeta\left(P\right) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\ \zeta\left(P^H(\mathbf{x})\right)\right]\right| \leq 3\left(CB_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d\right)^{2/3}.$$ ◀

Proof of Theorem 19. Let $\lambda > 0$ be determined later and let ζ_λ be defined as in Fact 57. By Theorem 20 and Fact 57,

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\mathrm{Tr}\ \zeta_\lambda\left(P(\mathbf{b})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x_f}}[\mathrm{Tr}\ \zeta_\lambda\left(\mathbf{P}(\mathbf{x_f})\right)]\right| \leq 4C_1B_4\lambda^{-2}(9m)^d d\tau,$$

where C_1, B_4 are universal constants. By Fact 57 we also have

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\mathrm{Tr}\ \zeta\left(P(\mathbf{b})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\mathrm{Tr}\ \zeta_\lambda\left(P(\mathbf{b})\right)]\right| \leq 2\lambda^2$$

and

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x_f}}[\mathrm{Tr}\ \zeta_\lambda\left(\mathbf{P}(\mathbf{x_f})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x_f}}[\mathrm{Tr}\ \zeta\left(\mathbf{P}(\mathbf{x_f})\right)]\right| \leq 2\lambda^2.$$

By the triangle inequality, we have

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\mathrm{Tr}\ \zeta\left(P(\mathbf{b})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x_f}}\left[\mathrm{Tr}\ \zeta\left(\mathbf{P}(\mathbf{x_f})\right)\right]\right| \leq 4\lambda^2 + 4C_1B_4\lambda^{-2}(9m)^d d\tau.$$

Choosing $\lambda = \left(C_1B_4(9m)^d d\tau\right)^{1/4}$, we have

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\mathrm{Tr}\ \zeta\left(P(\mathbf{b})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x_f}}\left[\mathrm{Tr}\ \zeta\left(\mathbf{P}(\mathbf{x_f})\right)\right]\right| \leq 8\left(C_1B_4(9m)^d d\tau\right)^{1/2}.$$

Let $C = 8\sqrt{C_1B_4}$, we conclude the result. ◀

▶ **Lemma 81.** *Given $d, n \in \mathbb{Z}_{>0}$, and a random matrix*

$$P(\mathbf{b}) = \sum_{S\subseteq[n]:|S|\leq d} \mathbf{b}_S P_S,$$

where $\mathbf{b}$ is a $2d$-wise uniform random vector from $\{\pm1\}^n$ and $\mathbb{E}_{\mathbf{b}}\left[\|\|P(\mathbf{b})\|\|_2^2\right] \leq 1$, it holds that

$$\sum_{i=1}^{n} \mathrm{VarInf}_i\left(P(\mathbf{b})\right) \leq d.$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{VarInf}_i\left(P(\mathbf{b})\right) &= \sum_{i=1}^{n}\sum_{S\ni i} \|\|P_S\|\|_2^2 \\
&= \sum_{S\subseteq[n]:|S|\leq d} |S|\,\|\|P_S\|\|_2^2 \\
&\leq d \sum_{S\subseteq[n]:|S|\leq d} \|\|P_S\|\|_2^2 \\
&= d\,\mathop{\mathbb{E}}_{\mathbf{b}}\left[\|\|P(\mathbf{b})\|\|_2^2\right] \quad \leq d.
\end{aligned}$$

◀

The following lemma is crucial to our proof. The proof follows closely to the proof of [36, Lemma 5.4].

▶ **Lemma 82.** *Given $d, n, p \in \mathbb{Z}_{>0}$, and a random matrix*

$$P(\mathbf{b}) = \sum_{S\subseteq[n]:|S|\leq d} \mathbf{b}_S P_S,$$

satisfying $\mathbb{E}_{\mathbf{b}}\left[\|\|P(\mathbf{b})\|\|_2^2\right] \leq 1$, where $\mathbf{b}$ is a $2d$-wise uniform random vector drawn from $\{\pm1\}^n$, let $\mathcal{F} = \{f : [n] \to [p]\}$ be a family of pairwise uniform hash functions. Then for $\mathbf{f} \sim_{\mathrm{u}} \mathcal{F}$,

$$\mathop{\mathbb{E}}_{\mathbf{f}}\left[\sum_{j=1}^{p} \mathrm{VarInf}_{\mathbf{f},j}\left(P(\mathbf{b})\right)^2\right] \leq \sum_{i=1}^{n} \mathrm{VarInf}_i\left(P(\mathbf{b})\right)^2 + \frac{d^2}{p}.$$

Proof. Fix $j \in [p]$ and for $1 \leq i \leq n$, let $\mathbf{X}_i$ be the indicator variable that is 1 if $f(i) = j$ and 0 otherwise. For brevity, let $\tau_i = \mathrm{VarInf}_i\left(P(\mathbf{b})\right)$ for $i \in [n]$. Now,

$$\begin{aligned}
\mathrm{VarInf}_{\mathbf{f},j}\left(P(\mathbf{b})\right) &= \sum_{S:S\cap f^{-1}(j)\neq\emptyset} \|\|P_S\|\|_2^2 \leq \sum_{S} \|\|P_S\|\|_2^2 \left(\sum_{i\in S} \mathbf{X}_i\right) \\
&= \sum_{i\in[n]} \mathbf{X}_i \sum_{S\ni i} \|\|P_S\|\|_2^2 = \sum_{i\in[n]} \mathbf{X}_i\tau_i
\end{aligned}$$

Thus

$$\mathrm{VarInf}_{\mathbf{f},j}\left(P(\mathbf{b})\right)^2 \leq \left(\sum_{i\in[n]} \mathbf{X}_i\tau_i\right)^2 = \sum_{i\in[n]} \mathbf{X}_i^2\tau_i^2 + \sum_{i\neq k} \mathbf{X}_i\mathbf{X}_k\tau_i\tau_k.$$

Note that $\mathbb{E}[\mathbf{X}_i] = 1/p$ and for $i \neq k$, $\mathbb{E}[\mathbf{X}_i\mathbf{X}_k] = 1/p^2$. Thus

$$\mathbb{E}\left[\mathrm{VarInf}_{\mathbf{f},j}\left(P(\mathbf{b})\right)^2\right] \leq \frac{1}{p}\sum_i \tau_i^2 + \sum_{i\neq k}\tau_i\tau_k\frac{1}{p^2} \leq \frac{1}{p}\sum_i \tau_i^2 + \frac{1}{p^2}\left(\sum_i \tau_i\right)^2.$$

The lemma follows by using Lemma 81 and summing all $j \in [p]$. ◀

We are ready to prove Theorem 20.

Proof of Theorem 20. We prove this by a hybrid argument.
Denote $\mathbf{b}^{(0)} = \mathbf{b} = G(f, \mathbf{b}, \ldots, \mathbf{b})$. For $j \in [p]$, define $\mathbf{b}^{(j)} = G\left(f, \mathbf{z}^1, \ldots, \mathbf{z}^j, \mathbf{b}, \ldots, \mathbf{b}\right)$, i.e., substituting $\mathbf{b}^{(j-1)}|_{f^{-1}(j)}$ with $\mathbf{z}^j_{f^{-1}(j)}$. Then $\mathbf{b}^{(p)} = \mathbf{x}_{\mathbf{f}}$, and

$$\mathbf{P}(\mathbf{b}^{(j-1)}) = \sum_{S:S\cap f^{-1}(j)=\emptyset} \mathbf{b}_S^{(j-1)}P_S + \sum_{S:S\cap f^{-1}(j)\neq\emptyset} \mathbf{b}_S^{(j-1)}P_S$$
$$\mathbf{P}(\mathbf{b}^{(j)}) = \sum_{S:S\cap f^{-1}(j)=\emptyset} \mathbf{b}_S^{(j)}P_S + \sum_{S:S\cap f^{-1}(j)\neq\emptyset} \mathbf{b}_S^{(j)}P_S.$$

Note that for $S \cap f^{-1}(j) = \emptyset$, $\mathbf{b}_S^{(j-1)} = \mathbf{b}_S^{(j)}$. Denote

$$\mathbf{A} = \sum_{S:S\cap f^{-1}(j)=\emptyset} \mathbf{b}_S^{(j)}P_S, \qquad \mathbf{B} = \sum_{S:S\cap f^{-1}(j)\neq\emptyset} \mathbf{b}_S^{(j-1)}P_S, \qquad \mathbf{C} = \sum_{S:S\cap f^{-1}(j)\neq\emptyset} \mathbf{b}_S^{(j)}P_S.$$

We have

$$\begin{aligned}
&\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j-1)}}\left[\mathrm{Tr}\ \xi\left(\mathbf{P}(\mathbf{b}^{(j-1)})\right)\right] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j)}}\left[\mathrm{Tr}\ \xi\left(\mathbf{P}(\mathbf{b}^{(j)})\right)\right]\right| \\
&= \left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j-1)}}\left[\mathrm{Tr}\ \xi\left(\mathbf{A}+\mathbf{B}\right)\right] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j)}}\left[\mathrm{Tr}\ \xi\left(\mathbf{A}+\mathbf{C}\right)\right]\right| \\
&= \left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j-1)}}\left[\sum_{k=0}^{3}\frac{1}{k!}\mathrm{Tr}\ D^k\xi(\mathbf{A})[\mathbf{B}] + \mathrm{Tr}\ \Delta_{4,\xi}(\mathbf{A},\mathbf{B})\right]\right. \\
&\qquad\qquad \left. - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j)}}\left[\sum_{k=0}^{3}\frac{1}{k!}\mathrm{Tr}\ D^k\xi(\mathbf{A})[\mathbf{C}] + \mathrm{Tr}\ \Delta_{4,\xi}(\mathbf{A},\mathbf{C})\right]\right|
\end{aligned}$$

By Fact 53 and the fact that $\mathbf{z}_j$ is $4d$-wise uniform, we have for $k = 0, 1, 2, 3$,

$$\mathop{\mathbb{E}}_{\mathbf{b}^{(j-1)}}\left[\mathrm{Tr}\ D^k\xi(\mathbf{A})\left[\mathbf{B}\right]\right] = \mathop{\mathbb{E}}_{\mathbf{b}^{(j)}}\left[\mathrm{Tr}\ D^k\xi(\mathbf{A})\left[\mathbf{C}\right]\right].$$

Thus,

$$\begin{aligned}
&\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j-1)}}\left[\mathrm{Tr}\ \xi\left(\mathbf{P}(\mathbf{b}^{(j-1)})\right)\right] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j)}}\left[\mathrm{Tr}\ \xi\left(\mathbf{P}(\mathbf{b}^{(j)})\right)\right]\right| \\
&\leq \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j-1)}}\left[|\mathrm{Tr}\ \Delta_{4,\xi}(\mathbf{A},\mathbf{B})|\right] + \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j)}}\left[|\mathrm{Tr}\ \Delta_{4,\xi}(\mathbf{A},\mathbf{C})|\right] \\
&\leq C_1C_0\left(\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j-1)}}\left[\|\mathbf{B}\|_4^4\right] + \mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j)}}\left[\|\mathbf{C}\|_4^4\right]\right),
\end{aligned}$$

where the last inequality is from Fact 54, and C_1 is a universal constant. Because $\mathbf{z}_j$ is $4d$-wise uniform, we have $\mathbb{E}_{\mathbf{b}^{(j-1)}}\left[\|\|\mathbf{B}\|\|_4^4\right] = \mathbb{E}_{\mathbf{b}^{(j)}}\left[\|\|\mathbf{C}\|\|_4^4\right]$. Using Theorem 14 with $\eta \leftarrow 1/\sqrt{3}$,

$$\mathop{\mathbb{E}}_{\mathbf{b}^{(j-1)}}\left[\|\|\mathbf{B}\|\|_4^4\right] \leq (9m)^d \left(\mathop{\mathbb{E}}_{\mathbf{b}^{(j)}}\left[\|\|\mathbf{B}\|\|_2^2\right]\right)^2.$$

So we have

$$\begin{aligned}
&\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j-1)}}\left[\operatorname{Tr}\xi\left(\mathbf{P}(\mathbf{b}^{(j-1)})\right)\right] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{b}^{(j)}}\left[\operatorname{Tr}\xi\left(\mathbf{P}(\mathbf{b}^{(j)})\right)\right]\right| \\
\leq\ & 2C_1C_0(9m)^d \mathop{\mathbb{E}}_{\mathbf{f}}\left[\left(\mathop{\mathbb{E}}_{\mathbf{b}^{(j-1)}}\left[\|\|\mathbf{B}\|\|_2^2\right]\right)^2\right] \\
=\ & 2C_1C_0(9m)^d \mathop{\mathbb{E}}_{\mathbf{f}}\left[\operatorname{VarInf}_{\mathbf{f},j}\left(P(\mathbf{b})\right)^2\right].
\end{aligned}$$

Summing over $j \in [p]$ and by Lemma 82, we have

$$\begin{aligned}
&\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\operatorname{Tr}\zeta\left(P(\mathbf{b})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x_f}}\left[\operatorname{Tr}\zeta\left(\mathbf{P}(\mathbf{x_f})\right)\right]\right| \\
\leq\ & 2C_1C_0(9m)^d\left(\sum_{i=1}^n \operatorname{VarInf}_i\left(P(\mathbf{b})\right)^2 + \frac{d^2}{p}\right) \\
\leq\ & 2C_1C_0(9m)^d\left(\tau\sum_{i=1}^n \operatorname{VarInf}_i\left(P(\mathbf{b})\right) + \frac{d^2}{p}\right) \\
\leq\ & 4C_1C_0(9m)^d d\tau,
\end{aligned}$$

where the last inequality is by Lemma 81 and $p \geq d/\tau$. ◀

B.2 Positivity Tester for Low Degree Operators

Proof of Theorem 23. Consider the algorithm below

Input: Parameters given in Definition 22.
Algorithm:

1. **Regularization**: Compute $\tau = \delta^3/\left(8\cdot 3^{2d}m^d d^2\right)$. For each i, compute the influence
$$\operatorname{Inf}_i\left(P\right) = \sum_{\sigma:\sigma_i\neq 0}\widehat{P}(\sigma)^2.$$
Let $H = \{i : \operatorname{Inf}_i\left(P\right) > \tau\}$.
2. **Derandomized invariance principle**: Let p be the smallest power of 2 satisfying $p \geq d/\tau$. Let $n = \left(m^2-1\right)\left(D - |H|\right)$ and $\mathcal{F} = \{f : [n] \to [p]\}$ be a family of pairwise uniform hash functions. For any $i \in [p]$, let $\mathbf{z}^i$ be $4d$-wise uniform random variables of length n and $\left(\mathbf{z}^i\right)$'s be independent across $i \in [p]$. For any $f \in \mathcal{F}$, set $\mathbf{x}_f = G\left(f, \mathbf{z}^1, \ldots, \mathbf{z}^p\right)$ as defined in Theorem 19. Define the random operator
$$P'(f,\mathbf{z}) = \sum_{\sigma\in[m^2]^D_{\geq 0}:|\sigma|\leq d}\widehat{P}(\sigma)\mathbf{x}_{f,\sigma_{\bar{H}}}\mathcal{B}_{\sigma_H}, \tag{19}$$
where $\mathbf{x}_{f,\sigma_{\bar{H}}} = \prod_{i\notin H}\left(\mathbf{x}_f\right)_{\left(m^2-1\right)(i-1)+\sigma_i}$ and $\mathcal{B}_{\sigma_H} = \bigotimes_{i\in H}\mathcal{B}_{\sigma_i}$.

3. Compute the distance to PSD: For each $f, \mathbf{z}$, compute

$$\delta_{f,\mathbf{z}} = m^{-|H|} \operatorname{Tr} \zeta(P'(f,\mathbf{z})).$$

4. Accept if

$$\mathop{\mathbb{E}}_{f,\mathbf{z}}[\delta_{f,\mathbf{z}}] < \beta.$$

Time complexity

1. Given that each computation of $\mathrm{Inf}_i(P)$ entails calculating a sum of products of Fourier coefficients, the time required can be expressed as $\sum_{i=0}^{d} \binom{D}{i} \left(m^2-1\right)^i \leq dm^{2d}D^d$. In addition, the time needed to determine the set H is at most D.
2. When fixing f and $\mathbf{z}$, computing $\delta_{f,\mathbf{z}}$ takes time

$$\exp(|H|) = \exp(|d/\tau|) = \exp\left(\mathrm{poly}\left(m^d, 1/\delta\right)\right).$$

3. By Lemma 60 and Corollary 61, the enumeration over $\mathcal{F}$ and $\mathbf{z}$ takes time polynomial in D, thus computing the expectation of $\delta_{f,\mathbf{z}}$ also takes time polynomial in D.

Correctness

By the choice of τ, it holds that

$$\left(3^d m^{d/2}\sqrt{\tau}d\right)^{2/3} \leq \delta/2, \tag{20}$$

$$C\sqrt{(9m)^d d\tau} \leq \delta/2. \tag{21}$$

Let $\mathbf{b} \in \{-1,1\}^n$ be uniformly distributed. Consider the operator $P^{(1)}$ obtained by replacing the basis outside of H by random bits. That is,

$$P^{(1)}(\mathbf{b}) = \sum_{\sigma \in [m^2]^D_{\geq 0}, |\sigma| \leq d} \widehat{P}(\sigma) \mathbf{b}_{\sigma_{\bar{H}}} \mathcal{B}_{\sigma_H},$$

where $\mathbf{b}_{\sigma_{\bar{H}}} = \prod_{i \notin H} \mathbf{b}_{(m^2-1)(i-1)+\sigma_i}$ and $\mathcal{B}_{\sigma_H} = \bigotimes_{i \in H} \mathcal{B}_{\sigma_i}$.

By Eq. (20) and Lemma 16, we have

$$\left| \frac{1}{m^{|H|}} \mathop{\mathbb{E}}_{\mathbf{b}} \left[\operatorname{Tr} \zeta(P^{(1)}(\mathbf{b})) \right] - \frac{1}{m^D} \operatorname{Tr} \zeta(P) \right| \leq \delta/2.$$

Then we define $P^{(2)}$ to be the operator obtained by replacing $\mathbf{b}$ with $\mathbf{x}_{f,\mathbf{z}}$, which is the operator in Equation (19). By Eq. (21) and Theorem 19,

$$\left| \frac{1}{m^{|H|}} \mathop{\mathbb{E}}_{\mathbf{b}} \left[\operatorname{Tr} \zeta(P^{(2)}(\mathbf{x}_{f,\mathbf{z}})) \right] - \frac{1}{m^{|H|}} \mathop{\mathbb{E}}_{f,\mathbf{z}} \left[\operatorname{Tr} \zeta(P^{(1)}(\mathbf{b})) \right] \right| \leq \delta/2.$$

Thus by triangle inequality, we have

$$\left| \frac{1}{m^{|H|}} \mathop{\mathbb{E}}_{f,\mathbf{z}} \left[\operatorname{Tr} \zeta(P^{(2)}(\mathbf{x}_{f,\mathbf{z}})) \right] - \frac{1}{m^D} \operatorname{Tr} \zeta(P) \right| \leq \delta. \tag{22}$$

The algorithm computes $m^{-|H|} \mathbb{E}_{f,\mathbf{z}} \left[\operatorname{Tr} \zeta(P^{(2)}(\mathbf{x}_{f,\mathbf{z}})) \right]$. By Eq.(22), the value is smaller than β if $m^{-D} \operatorname{Tr} \zeta(P) < \beta - \delta$; or greater than β if $m^{-D} \operatorname{Tr} \zeta(P) > \beta + \delta$. Therefore, the algorithm distinguishes the two cases correctly. ◀

B.3 Noisy Nonlocal Games are NP-complete

B.3.1 The nondeterministic algorithm

First, we prove an upper bound on the Number of Noisy MES's for Nonlocal Games. The proof follows closely to that of [46]. The major difference is that in the proof of [46], each pair of questions (x, y) is treated independently. Then, a union bound is applied to all possible questions. To improve the upper bound, we take into account the distribution of the questions, combined with a better Gaussian dimension reduction in [47]. Then our new upper bound below only depends polynomially on the size of the question set whereas the previous one has an exponential dependence.

Gaussian Dimension Reduction

The following lemma is a simplified version of [47, Lemma 5.13], with the questions and answers being classical. In the proof of Theorem 28, we will use this lemma, after we replace the low-influence registers by Gaussian random variables, to further reduce the dimension of the Gaussian space. The only difference is in Item 3 of Lemma 83, where we preserve the expectation of the ζ function value over the random variable $\mathbf{M}$. In the previous version (Item 2 of [47, Lemma 5.13]), we used Markov's inequality on the expectation value. As the notations are considerably different, we include a new proof for completeness.

▶ **Lemma 83** ([47, Lemma 5.13]). *Given parameters $\rho \in [0,1]$, $\delta > 0$, $d, n, h \in \mathbb{Z}_{>0}$, $m \geq 2$, an m-dimensional noisy MES ψ_{AB} with the maximal correlation $\rho = \rho(\psi_{AB})$, and degree-d multilinear joint random matrices*

$$(P(\mathbf{g}), Q(\mathbf{h})) = \left(\sum_{S \subseteq [n]} \mathbf{g}_S P_S, \sum_{S \subseteq [n]} \mathbf{h}_S Q_S \right)_{(\mathbf{g},\mathbf{h}) \sim \mathcal{G}_\rho^{\otimes n}},$$

where $\mathbf{g}_S = \prod_{i \in S} \mathbf{g}_i, \mathbf{h}_S = \prod_{i \in S} \mathbf{h}_i$ and $P_S, Q_S \in \mathcal{H}_m^{\otimes h}$ for all $S \subseteq [n]$, satisfying

$$\mathop{\mathbb{E}}_{\mathbf{g}} \left[\|\!| P(\mathbf{g}) |\!\|_2^2 \right] \leq 1 \text{ and } \mathop{\mathbb{E}}_{\mathbf{h}} \left[\|\!| Q(\mathbf{h}) |\!\|_2^2 \right] \leq 1.$$

Let $L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_n\right)$ be the space of random operators whose Fourier coefficients are square-integrable with respect to the measure γ_n. Then there exists an explicitly computable $n_0 = n_0(d, \delta)$ and maps $f_M, g_M : L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_n\right) \to L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_n\right)$ for $M \in \mathbb{R}^{n \times n_0}$ and joint random operators $(P(M\tilde{\mathbf{x}}), Q(M\tilde{\mathbf{y}})) = (f_M(P(\mathbf{g})), g_M(Q(\mathbf{h})))$:

$$(P(M\tilde{\mathbf{x}}), Q(M\tilde{\mathbf{y}})) = \left(\sum_{S \subseteq [n]} \mathbf{u}_S P_S, \sum_{S \subseteq [n]} \mathbf{v}_S Q_S \right)_{(\mathbf{x},\mathbf{y}) \sim \mathcal{G}_\rho^{\otimes n_0}},$$

where $\tilde{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|_2$, $\tilde{\mathbf{y}} = \mathbf{y}/\|\mathbf{y}\|_2$, $\mathbf{u}_S = \prod_{i \in S} \langle m_i, \tilde{\mathbf{x}} \rangle$, $\mathbf{v}_S = \prod_{i \in S} \langle m_i, \tilde{\mathbf{y}} \rangle$, $\langle \cdot, \cdot \rangle$ denotes the standard inner product over $\mathbb{R}^{n_0}$ and m_i denotes the i'th row of M, such that if we sample $\mathbf{M} \sim \gamma_{n \times n_0}$, then the following hold:

1. *With probability at least $1 - 2\delta$, we have*

$$\mathop{\mathbb{E}}_{\mathbf{x}} \left[\|\!| P(\mathbf{M}\tilde{\mathbf{x}}) |\!\|_2^2 \right] \leq 1 + \delta \quad \text{and} \quad \mathop{\mathbb{E}}_{\mathbf{y}} \left[\|\!| Q(\mathbf{M}\tilde{\mathbf{y}}) |\!\|_2^2 \right] \leq 1 + \delta.$$

2. *With probability at least $1 - \delta$, we have*

$$\left| \mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}} \left[\mathrm{Tr}\left(\left(P(\mathbf{M}\tilde{\mathbf{x}}) \otimes Q(\mathbf{M}\tilde{\mathbf{y}}) \right) \psi_{AB}^{\otimes h} \right) \right] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}} \left[\mathrm{Tr}\left(\left(P(\mathbf{g}) \otimes Q(\mathbf{h}) \right) \psi_{AB}^{\otimes h} \right) \right] \right| \leq \delta.$$

3. $\mathop{\mathbb{E}}_{\mathbf{g}}[\mathrm{Tr}\,\zeta\,(P(\mathbf{g}))] = \mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x}}[\mathrm{Tr}\,\zeta\,(P(\mathbf{M}\tilde{\mathbf{x}}))]$ *and* $\mathop{\mathbb{E}}_{\mathbf{h}}[\mathrm{Tr}\,\zeta\,(Q(\mathbf{h}))] = \mathop{\mathbb{E}}_{\mathbf{M},\mathbf{y}}[\mathrm{Tr}\,\zeta\,(Q(\mathbf{M}\tilde{\mathbf{y}}))]$.
4. *the maps f_M, g_M are linear and unital for any nonzero $M \in \mathbb{R}^{n\times n_0}$.*

In particular, one may take $n_0 = \frac{d^{O(d)}}{\delta^6}$.

For $M \in \mathbb{R}^{n\times n_0}$, denote $F(M) = \mathbb{E}_{\mathbf{x},\mathbf{y}}\big[\mathrm{Tr}\left((P(M\tilde{\mathbf{x}}) \otimes Q(M\tilde{\mathbf{y}}))\,\psi_{AB}^{\otimes h}\right)\big]$. To prove Lemma 83 item 2, we need the following lemma.

▶ **Lemma 84.** *In the setting of Lemma 83, given $d \in \mathbb{Z}_{>0}$, $\delta > 0$, there exists $n_0 = \frac{d^{O(d)}}{\delta^2}$ such that the following holds: For $\mathbf{M} \sim \gamma_{n\times n_0}$,*

$$\left|\mathbb{E}[F(\mathbf{M})] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}\left[\mathrm{Tr}\left((P(\mathbf{g}) \otimes Q(\mathbf{h}))\,\psi_{AB}^{\otimes h}\right)\right]\right| \le \delta,$$
$$\mathrm{Var}\,[F(\mathbf{M})] \le \delta.$$

We use the following lemma to prove Lemma 84.

▶ **Lemma 85** ([20, Lemma A.8,A.9])**.** *Given parameters d and δ, there exists an explicitly computable $n_0(d,\delta)$ such that the followings hold:*

- *For any subsets $S, T \subseteq [n]$ satisfying $|S|, |T| \le d$, it holds that*

$$\text{if } S \neq T: \quad \mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}[\mathbf{u}_S \mathbf{v}_T] = 0,$$
$$\text{if } S = T: \quad \left|\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}[\mathbf{u}_S \mathbf{v}_T] - \rho^{|S|}\right| \le \delta.$$

- *Let $(\mathbf{x}', \mathbf{y}') \sim \mathcal{G}_\rho^{\otimes n_0}$ be independent of $(\mathbf{x}, \mathbf{y})$, and let $\mathbf{u}'_S = \prod_{i\in S}\left\langle m_i, \frac{\mathbf{x}'}{\|\mathbf{x}'\|_2}\right\rangle$, $\mathbf{v}'_S = \prod_{i\in S}\left\langle m_i, \frac{\mathbf{y}'}{\|\mathbf{y}'\|_2}\right\rangle$. For any subsets $S, T, S', T' \subseteq [n]$ satisfying $|S|, |T|, |S'|, |T'| \le d$, it holds that*

$$\text{if } S \triangle T \triangle S' \triangle T' \neq \emptyset: \quad \left|\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y},\mathbf{x}',\mathbf{y}'}[\mathbf{u}_S \mathbf{v}_T \mathbf{u}'_{S'} \mathbf{v}'_{T'}] - \left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}[\mathbf{u}_S \mathbf{v}_T]\right)\left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x}',\mathbf{y}'}[\mathbf{u}'_{S'} \mathbf{v}'_{T'}]\right)\right| = 0,$$

$$\text{if } S \triangle T \triangle S' \triangle T' = \emptyset: \quad \left|\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y},\mathbf{x}',\mathbf{y}'}[\mathbf{u}_S \mathbf{v}_T \mathbf{u}'_{S'} \mathbf{v}'_{T'}] - \left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}[\mathbf{u}_S \mathbf{v}_T]\right)\left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x}',\mathbf{y}'}[\mathbf{u}'_{S'} \mathbf{v}'_{T'}]\right)\right| \le \delta.$$

Here, $S \triangle T \triangle S' \triangle T'$ is the symmetric difference of the sets S, T, S', T', equivalently, the set of all $i \in [n]$ which appear an odd number of times in the multiset $S \sqcup T \sqcup S' \sqcup T'$.

In particular, one may take $n_0 = \frac{d^{O(d)}}{\delta^2}$.

Proof of Lemma 84. Use Lemma 85 with parameters d and δ, we have

$$
\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{M}}[F(\mathbf{M})] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}\left[\mathrm{Tr}\left(\left(P(\mathbf{g})\otimes Q(\mathbf{h})\right)\psi_{AB}^{\otimes h}\right)\right]\right| \\
&= \left|\sum_{S,T\subseteq[n]}\left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}[\mathbf{u}_S\mathbf{v}_T] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}[\mathbf{g}_S\mathbf{h}_T]\right)\mathrm{Tr}\left(\left(P_S\otimes Q_T\right)\psi_{AB}^{\otimes h}\right)\right| \\
&= \left|\sum_{S\subseteq[n]}\left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}[\mathbf{u}_S\mathbf{v}_S] - \rho^{|S|}\right)\mathrm{Tr}\left(\left(P_S\otimes Q_S\right)\psi_{AB}^{\otimes h}\right)\right| \\
&\le \delta\sum_{S\subseteq[n]}\left|\mathrm{Tr}\left(\left(P_S\otimes Q_S\right)\psi_{AB}^{\otimes h}\right)\right| \quad \text{(Lemma 85)} \\
&\le \delta\sum_{S\subseteq[n]}|||P_S|||_2|||Q_S|||_2 \quad \text{(Fact 66)} \\
&\le \delta\sqrt{\sum_{S\subseteq[n]}|||P_S|||_2^2\cdot\sum_{S\subseteq[n]}|||Q_S|||_2^2} \\
&= \delta\left(\mathop{\mathbb{E}}_{\mathbf{g}}\left[|||P(\mathbf{g})|||_2^2\right]\mathop{\mathbb{E}}_{\mathbf{g}}\left[|||Q(\mathbf{h})|||_2^2\right]\right)^{1/2} \qquad \le \delta.
\end{aligned}
$$

Use Lemma 85 with parameters d and $\delta \leftarrow \delta/9^d$, we have

$$
\begin{aligned}
&\mathrm{Var}\left[F(\mathbf{M})\right] \\
&= \mathop{\mathbb{E}}_{\mathbf{M}}\left[F(\mathbf{M})^2\right] - \left(\mathop{\mathbb{E}}_{\mathbf{M}}[F(\mathbf{M})]\right)^2 \\
&\le \sum_{S,T,S',T'\subseteq[n]}\left|\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y},\mathbf{x}',\mathbf{y}'}[\mathbf{u}_S\mathbf{v}_T\mathbf{u}'_{S'}\mathbf{v}'_{T'}] - \left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}[\mathbf{u}_S\mathbf{v}_T]\right)\left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x}',\mathbf{y}'}[\mathbf{u}'_{S'}\mathbf{v}'_{T'}]\right)\right| \\
&\quad \left|\mathrm{Tr}\left(\left(P_S\otimes Q_S\right)\psi_{AB}^{\otimes h}\right)\mathrm{Tr}\left(\left(P_{S'}\otimes Q_{S'}\right)\psi_{AB}^{\otimes h}\right)\right| \\
&\le \frac{\delta}{9^d}\sum_{\substack{S,T,S',T'\subseteq[n]\\ S\triangle T\triangle S'\triangle T'=\emptyset}}|||P_S|||_2|||Q_T|||_2|||P_{S'}|||_2|||Q_{T'}|||_2
\end{aligned}
$$

To finish the proof, we will show that,

$$
\sum_{\substack{S,T,S',T'\subseteq[n]\\ S\triangle T\triangle S'\triangle T'=\emptyset}}|||P_S|||_2|||Q_T|||_2|||P_{S'}|||_2|||Q_{T'}|||_2 \le 9^d\mathop{\mathbb{E}}_{\mathbf{g}}\left[|||P(\mathbf{g})|||_2^2\right]\mathop{\mathbb{E}}_{\mathbf{g}}\left[|||Q(\mathbf{h})|||_2^2\right]
$$

Define functions $f, g : \{1,-1\}^n \to \mathbb{R}$ over the boolean hypercube as,

$$
f(x) = \sum_{\substack{S\subseteq[n]\\ |S|\le d}}|||P_S|||_2\chi_S(x) \quad \text{and} \quad g(x) = \sum_{\substack{T\subseteq[n]\\ |T|\le d}}|||Q_T|||_2\chi_T(x)
$$

By the hypercontractivity inequality over the boolean hypercube [42, Page 240]

$$
\mathop{\mathbb{E}}_{x}\left[f(x)^4\right] \le 9^d\left(\mathop{\mathbb{E}}_{x}\left[f(x)^2\right]\right)^2 \quad \text{and} \quad \mathop{\mathbb{E}}_{x}\left[g(x)^4\right] \le 9^d\left(\mathop{\mathbb{E}}_{x}\left[g(x)^2\right]\right)^2,
$$

we have

$$\begin{aligned}
&\sum_{\substack{S,T,S',T'\subseteq[n]\\ S\triangle T\triangle S'\triangle T'=\emptyset}} |||P_S|||_2|||Q_T|||_2|||P_{S'}|||_2|||Q_{T'}|||_2 \\
&= \mathop{\mathbb{E}}_x\left[f(x)^2 g(x)^2\right] \\
&\le \sqrt{\mathop{\mathbb{E}}_x[f(x)^4]\mathop{\mathbb{E}}_x[g(x)^4]} \\
&\le 9^d \mathop{\mathbb{E}}_x\left[f(x)^2\right]\mathop{\mathbb{E}}_x\left[g(x)^2\right] \\
&= 9^d \sum_{S\subseteq[n]} |||P_S|||_2^2 \sum_{S\subseteq[n]} |||Q_S|||_2^2 \\
&= 9^d \mathop{\mathbb{E}}_{\mathbf{g}}\left[|||P(\mathbf{g})|||_2^2\right]\mathop{\mathbb{E}}_{\mathbf{g}}\left[|||Q(\mathbf{h})|||_2^2\right] \quad \le 9^d.
\end{aligned}$$

Thus $\mathrm{Var}\left[F(\mathbf{M})\right] \le \delta$. ◀

To prove Lemma 83 Item 1, we need the following lemma whose proof is similar to that of Lemma 84. We omit the proof here.

▶ **Lemma 86.** *In the setting of Lemma 83, given $d \in \mathbb{Z}_{>0}$, $\delta > 0$, there exists $n_0 = \frac{d^{O(d)}}{\delta^2}$ such that the following holds: For $\mathbf{M} \sim \gamma_{n\times n_0}$,*

$$\begin{aligned}
\left|\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x}}\left[|||P(\mathbf{M}\tilde{\mathbf{x}})|||_2^2\right] - \mathop{\mathbb{E}}_{\mathbf{g}}\left[|||P(\mathbf{g})|||_2^2\right]\right| &\le \delta, \\
\mathrm{Var}\left[\mathop{\mathbb{E}}_{\mathbf{x}}\left[|||P(\mathbf{M}\tilde{\mathbf{x}})|||_2^2\right]\right] &\le \delta, \\
\left|\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{y}}\left[|||Q(\mathbf{M}\tilde{\mathbf{y}})|||_2^2\right] - \mathop{\mathbb{E}}_{\mathbf{h}}\left[|||Q(\mathbf{h})|||_2^2\right]\right| &\le \delta, \\
\mathrm{Var}\left[\mathop{\mathbb{E}}_{\mathbf{y}}\left[|||Q(\mathbf{M}\tilde{\mathbf{y}})|||_2^2\right]\right] &\le \delta.
\end{aligned}$$

Proof of Lemma 83. For item 2, we invoke Lemma 84 with parameters d and $\delta \leftarrow \delta^3/2$. Using Chebyshev's inequality, we have that for any $\eta > 0$,

$$\Pr_{\mathbf{M}}\left[\left|F(\mathbf{M}) - \mathop{\mathbb{E}}_{\mathbf{M}}[F(\mathbf{M})]\right| > \eta\right] \le \frac{\delta^3}{2\eta^2}.$$

Using the triangle inequality, we get

$$\begin{aligned}
&\Pr_{\mathbf{M}}\left[\left|F(\mathbf{M}) - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}\left[\mathrm{Tr}\left(\left(P(\mathbf{g})\otimes Q(\mathbf{h})\right)\psi_{AB}^{\otimes h}\right)\right]\right| > \delta\right] \\
\le\ &\Pr_{\mathbf{M}}\left[\left|F(\mathbf{M}) - \mathop{\mathbb{E}}_{\mathbf{M}}[F(\mathbf{M})]\right| + \left|\mathop{\mathbb{E}}_{\mathbf{M}}[F(\mathbf{M})] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}\left[\mathrm{Tr}\left(\left(P(\mathbf{g})\otimes Q(\mathbf{h})\right)\psi_{AB}^{\otimes h}\right)\right]\right| > \delta\right] \\
\le\ &\Pr_{\mathbf{M}}\left[\left|F(\mathbf{M}) - \mathop{\mathbb{E}}_{\mathbf{M}}[F(\mathbf{M})]\right| > \delta - \delta^3/2\right] \le \delta.
\end{aligned}$$

By Lemma 86, we can similarly argue for item 1. For item 3, note that for any fixed $x \in \mathbb{R}^{n_0}$, the distribution of $\mathbf{M}x/\|x\|_2$ is identical to γ_n. It is easy to verify Item 4. ◀

We are now ready to prove Theorem 28.

Proof of Theorem 28. The proof follows that in [46] step by step, except that the Gaussian dimension reduction step in the original proof is replaced by Lemma 83. Here, we include the proof for completeness.

Suppose the players use the strategy $\left(\left\{P_a^{x,(0)}\right\}_{a\in\mathcal{A}}^{x\in\mathcal{X}}, \left\{Q_b^{y,(0)}\right\}_{b\in\mathcal{B}}^{y\in\mathcal{Y}}\right)$ to achieve the highest winning probability when sharing n copies of ψ_{AB}, where $P_a^{x,(0)}$ is the POVM element of Alice corresponding to the answer a upon receiving the question x, and $Q_b^{y,(0)}$ is the POVM element of Bob corresponding to the answer b upon receiving the question y. Then for all $(x,y,a,b)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$, $P_a^{x,(0)}\geq 0$, $Q_b^{y,(0)}\geq 0$, $\sum_a P_a^{x,(0)}=\mathbb{1}$, $\sum_b Q_b^{y,(0)}=\mathbb{1}$, and $\omega_n(\mathfrak{G},\psi_{AB})=\mathrm{val}_n\left(\left\{P_a^{x,(0)}\right\},\left\{Q_b^{y,(0)}\right\}\right)$.

Let δ,τ be parameters which are chosen later. The proof is composed of several steps.

- **Smoothing.** This step allows us to restrict ourselves to strategies with low-degree POVMs.

  More specifically, for any $(x,y,a,b)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$, we apply the map $f^{(1)}$ implied by Lemma 62 to $P_a^{x,(0)}$ and $Q_b^{y,(0)}$ to get $P_a^{x,(1)}$ and $Q_b^{y,(1)}$, respectively. Note that for all x,y,a,b, $\left|\left|\left|P_a^{x,(0)}\right|\right|\right|_2^2\leq 1$ and $\left|\left|\left|Q_b^{y,(0)}\right|\right|\right|_2^2\leq 1$. Let $d=\frac{C\log^2\frac{1}{\delta}}{\delta(1-\rho)}$, by Lemma 62 Item 3 and Item 4,

  $$\left|\mathrm{Tr}\left(\left(P_a^{x,(1)}\otimes Q_b^{y,(1)}\right)\psi_{AB}^{\otimes n}\right)-\mathrm{Tr}\left(\left(P_a^{x,(0)}\otimes Q_b^{y,(0)}\right)\psi_{AB}^{\otimes n}\right)\right|\leq\delta$$

  and

  $$\frac{1}{m^n}\mathrm{Tr}\,\zeta(P_a^{x,(1)})\leq\delta,\quad \frac{1}{m^n}\mathrm{Tr}\,\zeta(Q_b^{y,(1)})\leq\delta.$$

  By Lemma 67 and Lemma 62 items 1, 2 and 5, the following hold.

  1. For any x,y,a,b, $P_a^{x,(1)}$ and $Q_b^{y,(1)}$ are of degree at most d.
  2. For any x,y,a,b, $\left|\left|\left|P_a^{x,(1)}\right|\right|\right|_2\leq 1$ and $\left|\left|\left|Q_b^{y,(1)}\right|\right|\right|_2\leq 1$.
  3. $\left|\mathrm{val}_n\left(\left\{P_a^{x,(1)}\right\},\left\{Q_b^{y,(1)}\right\}\right)-\mathrm{val}_n\left(\left\{P_a^{x,(0)}\right\},\left\{Q_b^{y,(0)}\right\}\right)\right|\leq\delta t^2$,
  4. $\frac{1}{m^n}\sum_{x,a}\mu_A(x)\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right)\leq\delta t$ and $\frac{1}{m^n}\sum_{y,b}\mu_B(y)\mathrm{Tr}\,\zeta\left(Q_b^{y,(1)}\right)\leq\delta t$.
  5. For any x,y, $\sum_{a\in\mathcal{A}}P_a^{x,(1)}=\sum_{b\in\mathcal{B}}Q_b^{y,(1)}=\mathbb{1}$.

- **Regularization.** In this step, we identify the set H of high-influence registers for all POVM elements.

  For any $(x,y,a,b)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$, we apply Lemma 64 to $P_a^{x,(1)}$ and $Q_b^{y,(1)}$ to get sets $H_{x,a}$ and $H_{y,b}$ of size at most d/τ, respectively, such that

  $$(\forall i\notin H_{x,a})\ \mathrm{Inf}_i\left(P_a^{x,(1)}\right)\leq\tau\quad\text{and}\quad(\forall i\notin H_{y,b})\ \mathrm{Inf}_i\left(Q_b^{y,(1)}\right)\leq\tau.$$

  Set $H=\left(\bigcup_{x,a}H_{x,a}\right)\cup\left(\bigcup_{y,b}H_{y,b}\right)$, then $h=|H|\leq\frac{2std}{\tau}$, and

  $$(\forall i\notin H)\ \mathrm{Inf}_i\left(P_a^{x,(1)}\right)\leq\tau\quad\text{and}\quad\mathrm{Inf}_i\left(Q_b^{y,(1)}\right)\leq\tau.$$

- **Invariance from $\mathcal{H}_m^{\otimes n}$ to $L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_{(m^2-1)(n-h)}\right)$.** In this step, we only keep the quantum registers in H and replace the rest of the quantum registers by Gaussian random variables. Hence, the number of quantum registers is reduced from n to $h = |H| = d/\tau$. For any $(x, y, a, b) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$, applying [46, Lemma 10.5] to $P_a^{x,(1)}$, $Q_b^{y,(1)}$ and H, we obtain joint random matrices

$$\left(P_a^{x,(2)}(\mathbf{g}), Q_b^{y,(2)}(\mathbf{h})\right) \in L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_{(m^2-1)(n-h)}\right) \times L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_{(m^2-1)(n-h)}\right),$$

where $(\mathbf{g}, \mathbf{h}) \sim \mathcal{G}_\rho^{\otimes 2(m^2-1)(n-h)}$, such that the following hold.

1. For any x, y, a, b, $\mathop{\mathbb{E}}_{\mathbf{g}}\left[\left\|\left|P_a^{x,(2)}(\mathbf{g})\right|\right\|_2^2\right] \le 1$ and $\mathop{\mathbb{E}}_{\mathbf{h}}\left[\left\|\left|Q_b^{y,(2)}(\mathbf{h})\right|\right\|_2^2\right] \le 1$.
2. $\mathbb{E}_{\mathbf{g},\mathbf{h}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(2)}(\mathbf{g})\right\}, \left\{Q_b^{y,(2)}(\mathbf{g})\right\}\right)\right] = \mathrm{val}_n\left(\left\{P_a^{x,(1)}\right\}, \left\{Q_b^{y,(1)}\right\}\right)$.
3. $\sum_{x,a} \mu_A(x) \left|\frac{1}{m^h}\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(2)}(\mathbf{g})\right)\right] - \frac{1}{m^n}\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right)\right| \le O\left(t\left(3^d m^{d/2}\sqrt{\tau}d\right)^{2/3}\right)$ and $\sum_{y,b} \mu_B(y) \left|\frac{1}{m^h}\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(2)}(\mathbf{h})\right)\right] - \frac{1}{m^n}\mathrm{Tr}\,\zeta\left(Q_b^{y,(1)}\right)\right| \le O\left(t\left(3^d m^{d/2}\sqrt{\tau}d\right)^{2/3}\right)$.
4. For any x, y, $\sum_{a\in\mathcal{A}} P_a^{x,(2)}(\mathbf{g}) = \sum_{b\in\mathcal{B}} Q_b^{y,(2)}(\mathbf{h}) = \mathbb{1}$.

- **Gaussian dimension reduction.** In this step, we apply Lemma 83 to further reduce the number of Gaussian random variables. This is the only part different from the proof in [46].

  Let n_0 be determined later. For any $(x, y, a, b) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$ and $M \in \mathbb{R}^{n\times n_0}$, applying Lemma 83 to $P_a^{x,(2)}(\mathbf{g})$ and $Q_b^{y,(2)}(\mathbf{h})$ with $\delta \leftarrow \delta/\left(2s^2t^2\right)$, $d \leftarrow d$, $n \leftarrow 2(m^2-1)(n-h)$, we get joint random matrices $P_a^{x,(3)}(M\tilde{\mathbf{x}})$ and $Q_b^{y,(3)}(M\tilde{\mathbf{y}})$. If we sample $\mathbf{M} \sim \gamma_{n\times n_0}$, by Lemma 83 item 3 we have

$$\sum_{x,a} \mu_A(x) \mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(3)}(\mathbf{M}\tilde{\mathbf{x}})\right)\right] = \sum_{x,a} \mu_A(x) \mathop{\mathbb{E}}_{\mathbf{g}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(2)}(\mathbf{g})\right)\right]$$

and

$$\sum_{y,b} \mu_B(y) \mathop{\mathbb{E}}_{\mathbf{M},\mathbf{y}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(3)}(\mathbf{M}\tilde{\mathbf{y}})\right)\right] = \sum_{y,b} \mu_B(y) \mathop{\mathbb{E}}_{\mathbf{h}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(2)}(\mathbf{h})\right)\right].$$

Then by Markov's inequality, with probability each at most $1/6$,

$$\sum_{x,a} \mu_A(x) \mathop{\mathbb{E}}_{\mathbf{x}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(3)}(\mathbf{M}\tilde{\mathbf{x}})\right)\right] > 6\sum_{x,a} \mu_A(x) \mathop{\mathbb{E}}_{\mathbf{g}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(2)}(\mathbf{g})\right)\right]$$

and

$$\sum_{y,b} \mu_B(y) \mathop{\mathbb{E}}_{\mathbf{y}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(3)}(\mathbf{M}\tilde{\mathbf{y}})\right)\right] > 6\sum_{y,b} \mu_B(y) \mathop{\mathbb{E}}_{\mathbf{h}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(2)}(\mathbf{h})\right)\right].$$

By Lemma 83 item 1, 2, and using a union bound, with probability at least $2/3 - \delta$ the following hold:

1. For any x, y, a, b, $\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|\left|P_a^{x,(3)}(M\tilde{\mathbf{x}})\right|\right\|_2^2\right] \le 2$ and $\mathop{\mathbb{E}}_{\mathbf{y}}\left[\left\|\left|Q_b^{y,(3)}(M\tilde{\mathbf{y}})\right|\right\|_2^2\right] \le 2$.
2. $\left|\mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(3)}(M\tilde{\mathbf{x}})\right\}, \left\{Q_b^{y,(3)}(M\tilde{\mathbf{y}})\right\}\right)\right] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(2)}(\mathbf{g})\right\}, \left\{Q_b^{y,(2)}(\mathbf{g})\right\}\right)\right]\right| \le \delta t^2$.

3. $\sum_{x,a}\mu_A(x)\mathop{\mathbb{E}}_{\mathbf{x}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(3)}(M\tilde{\mathbf{x}})\right)\right]\le 6\sum_{x,a}\mu_A(x)\mathop{\mathbb{E}}_{\mathbf{g}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(2)}(\mathbf{g})\right)\right]$ and $\sum_{y,b}\mu_B(y)\mathop{\mathbb{E}}_{\mathbf{y}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(3)}(M\tilde{\mathbf{y}})\right)\right]\le 6\sum_{y,b}\mu_B(y)\mathop{\mathbb{E}}_{\mathbf{h}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(2)}(\mathbf{h})\right)\right]$.
4. For any x,y, $\sum_{a\in\mathcal{A}}P_a^{x,(3)}(M\tilde{\mathbf{x}})=\sum_{b\in\mathcal{B}}Q_b^{y,(3)}(M\tilde{\mathbf{y}})=\mathbb{1}$.

Here $n_0=\frac{d^{O(d)}s^{12}t^{12}}{\delta^6}$.

- **Smoothing random matrices.** In this step, we reduce $\deg(P_a^{x,(3)})$ and $\deg(Q_b^{y,(3)})$ for any $(x,y,a,b)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$. We apply [46, Lemma 12.1] to $P_a^{x,(3)}(M\tilde{\mathbf{x}})$ and $Q_b^{y,(3)}(M\tilde{\mathbf{y}})$ with $\delta\leftarrow\delta$, $h\leftarrow h$, $n\leftarrow n_0$ and obtain joint random matrices $P_a^{x,(4)}(\mathbf{x}),Q_b^{y,(4)}(\mathbf{y})\in L^2\left(\mathcal{H}_m^{\otimes h},\gamma_{n_0}\right)$ such that the following holds.
  1. For any x,y,a,b, the entries of $P_a^{x,(4)}(\mathbf{x})$ and $Q_b^{y,(4)}(\mathbf{y})$ are polynomials of degree at most d.
  2. For any $(x,y,a,b)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$, $\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|\left|P_a^{x,(4)}(\mathbf{x})\right|\right\|_2^2\right]\le 2$ and $\mathop{\mathbb{E}}_{\mathbf{y}}\left[\left\|\left|Q_b^{y,(4)}(\mathbf{y})\right|\right\|_2^2\right]\le 2$.
  3. $\left|\mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(4)}(\mathbf{x})\right\},\left\{Q_b^{y,(4)}(\mathbf{x})\right\}\right)\right]-\mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(3)}(M\tilde{\mathbf{x}})\right\},\left\{Q_b^{y,(3)}(M\tilde{\mathbf{y}})\right\}\right)\right]\right|\le\delta t^2$.
  4. $\left|\sum_{x,a}\mu_A(x)\mathop{\mathbb{E}}_{\mathbf{x}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(4)}(\mathbf{x})\right)\right]-\sum_{x,a}\mu_A(x)\mathop{\mathbb{E}}_{\mathbf{x}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(3)}(M\tilde{\mathbf{x}})\right)\right]\right|\le\delta t$ and $\left|\sum_{y,b}\mu_B(y)\mathop{\mathbb{E}}_{\mathbf{y}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(4)}(\mathbf{y})\right)\right]-\sum_{y,b}\mu_B(y)\mathop{\mathbb{E}}_{\mathbf{y}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(3)}(M\tilde{\mathbf{y}})\right)\right]\right|\le\delta t$.
  5. For any x,y, $\sum_{a\in\mathcal{A}}P_a^{x,(4)}(\mathbf{x})=\sum_{b\in\mathcal{B}}Q_b^{y,(4)}(\mathbf{y})=\mathbb{1}$.
- **Multilinearization.** For any $(x,y,a,b)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$, we apply [46, Lemma 13.1] to $P_a^{x,(4)}(\mathbf{x})$ and $Q_b^{y,(4)}(\mathbf{y})$ with $d\leftarrow d$, $\delta\leftarrow\tau$, $h\leftarrow h$, $n\leftarrow n_0$ and obtain joint random matrices $P_a^{x,(5)}(\mathbf{x}),Q_b^{y,(5)}(\mathbf{y})\in L^2\left(\mathcal{H}_m^{\otimes h},\gamma_{n_0n_1}\right)$ such that the following holds.
  1. For any x,y,a,b, the entries of $P_a^{x,(5)}(\mathbf{x})$ and $Q_b^{y,(5)}(\mathbf{y})$ are multilinear polynomials of degree at most d, and every variable in $P_a^{x,(5)}(\mathbf{x})$ and $Q_b^{y,(5)}(\mathbf{x})$ has influence at most τ.
  2. For any x,y,a,b, $\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|\left|P_a^{x,(5)}(\mathbf{x})\right|\right\|_2^2\right]\le 2$ and $\mathop{\mathbb{E}}_{\mathbf{y}}\left[\left\|\left|Q_b^{y,(5)}(\mathbf{y})\right|\right\|_2^2\right]\le 2$.
  3. $\left|\mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(5)}(\mathbf{x})\right\},\left\{Q_b^{y,(5)}(\mathbf{x})\right\}\right)\right]-\mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(4)}(\mathbf{x})\right\},\left\{Q_b^{y,(4)}(\mathbf{y})\right\}\right)\right]\right|\le\tau t^2$.
  4. $\left|\sum_{x,a}\mu_A(x)\mathop{\mathbb{E}}_{\mathbf{x}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(5)}(\mathbf{x})\right)\right]-\sum_{x,a}\mu_A(x)\mathop{\mathbb{E}}_{\mathbf{x}}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(4)}(\mathbf{x})\right)\right]\right|\le\tau t$ and $\left|\sum_{y,b}\mu_B(y)\mathop{\mathbb{E}}_{\mathbf{y}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(5)}(\mathbf{y})\right)\right]-\sum_{y,b}\mu_B(y)\mathop{\mathbb{E}}_{\mathbf{y}}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(4)}(\mathbf{y})\right)\right]\right|\le\tau t$.
  5. For any x,y, $\sum_{a\in\mathcal{A}}P_a^{x,(5)}(\mathbf{x})=\sum_{b\in\mathcal{B}}Q_b^{y,(5)}(\mathbf{y})=\mathbb{1}$.

Here $n_1=O\left(\frac{d^2}{\tau^2}\right)$.

- **Invariance from $L^2\left(\mathcal{H}_m^{\otimes h},\gamma_{n_0n_1}\right)$ to $\mathcal{H}_m^{\otimes h+n_0n_1}$.** In this step, we transform all the random matrices from the previous step to matrices without any classical randomness. In particular, we replace all the Gaussian random variables with n_0n_1 quantum registers, so after this step, the number of quantum registers is $h+n_0n_1$.
  For any $(x,y,a,b)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$, applying [46, Lemma 10.11] to $P_a^{x,(5)}(\mathbf{x})$, $Q_b^{y,(5)}(\mathbf{y})$ with $n\leftarrow n_0n_1$, $h\leftarrow h$, $d\leftarrow 2d$, $\tau\leftarrow\tau$ to get $P_a^{x,(6)},Q_b^{y,(6)}\in\mathcal{H}_m^{\otimes h+n_0n_1}$ satisfying the following.

1. For any x, y, a, b, $\left|\left|\left|P_a^{x,(6)}\right|\right|\right|_2^2 \leq 2$ and $\left|\left|\left|Q_b^{y,(6)}\right|\right|\right|_2^2 \leq 2$.
2. $\mathrm{val}_{h+n_0n_1}\left(\left\{P_a^{x,(6)}\right\}, \left\{Q_b^{y,(6)}\right\}\right) = \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(5)}(\mathbf{x})\right\}, \left\{Q_b^{y,(5)}(\mathbf{y})\right\}\right)\right]$.
3. $\sum_{x,a}\mu_A(x)\left|\frac{1}{m^{h+n_0n_1}}\mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right) - \frac{1}{m^h}\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(5)}(\mathbf{x})\right)\right]\right| \leq O\left(t\left(9^d m^d\sqrt{\tau}d\right)^{2/3}\right)$ and $\sum_{y,b}\mu_B(y)\left|\frac{1}{m^{h+n_0n_1}}\mathrm{Tr}\,\zeta\left(Q_b^{y,(6)}\right) - \frac{1}{m^h}\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(5)}(\mathbf{y})\right)\right]\right| \leq O\left(t\left(9^d m^d\sqrt{\tau}d\right)^{2/3}\right)$.
4. For any x, y, $\sum_{a\in\mathcal{A}} P_a^{x,(6)} = \sum_{b\in\mathcal{B}} Q_b^{y,(6)} = \mathbb{1}$.

- **Rounding.** Note that the matrices from the previous step may not form valid POVMs, so in this step we round them to close POVMs. In this step, the number of quantum registers remains the same as $h + n_0n_1$.
  By Lemma 65 there exist operators $\left\{P_a^{x,(7)}\right\}$ and $\left\{Q_b^{y,(7)}\right\}$ satisfying for all x

$$\sum_a \left|\left|\left|P_a^{x,(7)} - P_a^{x,(6)}\right|\right|\right|_2^2 \leq \frac{3(t+1)}{m^D}\sum_a \mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right) + 6\sqrt{t}\left(\frac{1}{m^D}\sum_a \mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right)\right)^{1/2}$$
$$\leq 10t\left(\frac{1}{m^D}\sum_a \mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right)\right)^{1/2}. \tag{23}$$

  Similarly, for all y, we have

$$\sum_a \left|\left|\left|Q_b^{y,(7)} - Q_b^{y,(6)}\right|\right|\right|_2^2 \leq 10t\left(\frac{1}{m^D}\sum_b \mathrm{Tr}\,\zeta\left(Q_b^{y,(6)}\right)\right)^{1/2}. \tag{24}$$

  Then

$$\begin{aligned}
&\left|\mathrm{val}_D\left(\left\{P_a^{x,(7)}\right\}, \left\{Q_b^{y,(7)}\right\}\right) - \mathrm{val}_D\left(\left\{P_a^{x,(6)}\right\}, \left\{Q_b^{y,(6)}\right\}\right)\right| \\
\leq\ &\left|\mathrm{val}_D\left(\left\{P_a^{x,(7)} - P_a^{x,(6)}\right\}, \left\{Q_b^{y,(7)}\right\}\right)\right| + \left|\mathrm{val}_D\left(\left\{P_a^{x,(6)}\right\}, \left\{Q_b^{y,(7)} - Q_b^{y,(6)}\right\}\right)\right| \\
\leq\ &\sum_{x,y,a,b}\mu(x,y)\left(\left|\left|\left|P_a^{x,(7)} - P_a^{x,(6)}\right|\right|\right|_2\left|\left|\left|Q_b^{y,(7)}\right|\right|\right|_2 + \left|\left|\left|P_a^{x,(6)}\right|\right|\right|_2\left|\left|\left|Q_b^{y,(7)} - Q_b^{y,(6)}\right|\right|\right|_2\right) \\
\leq\ &\left(\sum_b\sum_{x,a}\mu_A(x)\left|\left|\left|P_a^{x,(7)} - P_a^{x,(6)}\right|\right|\right|_2^2\right)^{1/2}\left(\sum_a\sum_{y,b}\mu_B(y)\left|\left|\left|Q_b^{y,(7)}\right|\right|\right|_2^2\right)^{1/2} \\
&\quad + \left(\sum_b\sum_{x,a}\mu_A(x)\left|\left|\left|P_a^{x,(6)}\right|\right|\right|_2^2\right)^{1/2}\left(\sum_a\sum_{y,b}\mu_B(y)\left|\left|\left|Q_b^{y,(7)} - Q_b^{y,(6)}\right|\right|\right|_2^2\right)^{1/2} \\
&\quad \text{(Cauchy Schwarz)} \\
\leq\ &\sqrt{10}t^2\left(\sum_x\mu_A(x)\left(\frac{1}{m^D}\sum_a \mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right)\right)^{1/2}\right)^{1/2} \\
&\quad + 2\sqrt{5}t^2\left(\sum_y\mu_B(y)\left(\frac{1}{m^D}\sum_b \mathrm{Tr}\,\zeta\left(Q_b^{y,(6)}\right)\right)^{1/2}\right)^{1/2} \\
\leq\ &\sqrt{10}t^2\left(\frac{1}{m^D}\sum_{x,a}\mu_A(x)\mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right)\right)^{1/4} + 2\sqrt{5}t^2\left(\frac{1}{m^D}\sum_{y,b}\mu_B(y)\mathrm{Tr}\,\zeta\left(Q_b^{y,(6)}\right)\right)^{1/4},
\end{aligned}$$

  where in the second last inequality, we use $\left|\left|\left|P_a^{x,(6)}\right|\right|\right| \leq 2$, $\left|\left|\left|Q_b^{y,(7)}\right|\right|\right| \leq 1$, and Equations (23) and (24). The last inequality follows from concavity of the function $x \mapsto \sqrt{x}$.

Keeping track of the parameters in the construction, we can upper bound $\frac{1}{m^D}\sum_{x,a}\mu_A(x)\mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right)$ and $\frac{1}{m^D}\sum_{y,b}\mu_B(y)\mathrm{Tr}\,\zeta\left(Q_b^{y,(6)}\right)$. We choose

$$\delta=\frac{\epsilon^4}{300t^9},\tau=\frac{\epsilon^{12}}{t^{27}}\exp\left(-\frac{300t^9\log m}{\epsilon^4(1-\rho)}\log^2\left(\frac{t}{\epsilon}\right)\right) \tag{25}$$

such that the difference in the game value at the final step matches that of the previous steps, remaining on the order of $O(\delta t^2)$. We conclude that the number of quantum registers is

$$\begin{aligned}D=h+n_0n_1&=\frac{d}{\tau}+\frac{d^{O(d)}s^{12}t^{12}}{\delta^6}\cdot O\left(\frac{d^2}{\tau^2}\right)\\&=O\left(\frac{s^{12}t^{120}}{\epsilon^{48}}\exp\left(\frac{600t^9\log m}{\epsilon^4(1-\rho)}\log^2\left(\frac{t}{\epsilon(1-\rho)}\right)\right)\right),\end{aligned}$$

which completes the proof. ◀

Next, we give the non-deterministic algorithm with the following parameters.

$C_{pt}=300$

$\varepsilon_{rd}=\varepsilon^2/(4t^3)$

$\delta=\dfrac{\varepsilon_{rd}^2}{C_{pt}t(t+1)}$

$d=\dfrac{C_{sm}\log^2\frac{1}{\delta}}{\delta\log(1/\rho)}$ as in Lemma 62.

$s_w=D\log m+\log\left(\dfrac{2}{\delta}\right)$ as in Lemma 68.

$D=$ is the polynomial specified in Theorem 28 with $\varepsilon\leftarrow\varepsilon/2$.

Proof of Proposition 26. Consider the algorithm below, with the parameters above.

Input: Parameters in Definition 24.

Certificate: Let $\{(\mathcal{A}_i,\mathcal{B}_i)\}_{i=0}^{m^2-1}$ be a pair of standard orthonormal basis satisfying Fact 46. A tuple of real numbers of width s_w, which are non-zero Fourier coefficients of a degree-d pseudo-strategy on D copies of ψ_{AB}. For each $x\in\mathcal{X},a\in\mathcal{A}$ and $\sigma\in[m^2]^D_{\geq 0}$ satisfying $|\sigma|\leq d$, the certificate should contain the coefficient $\widehat{P}_a^x(\sigma)$. Similarly, for $y\in\mathcal{Y},b\in\mathcal{B}$ and σ, the certificate should contain the coefficient $\widehat{Q}_b^y(\sigma)$. Then the degree-d pseudo-strategy can be written as P_a^x and Q_b^y satisfying

$$P_a^x=\sum_{|\sigma|\leq d}\widehat{P}_a^x(\sigma)\mathcal{A}_\sigma\text{ and }Q_b^y=\sum_{|\sigma|\leq d}\widehat{Q}_b^y(\sigma)\mathcal{B}_\sigma.$$

Algorithm:

1. Compute the winning probability on the pseudo-strategy, which is

$$\mathrm{val}_D\left(\{P_a^x\},\{Q_b^y\}\right)=\sum_{x,y,a,b}\mu(x,y)\cdot V(x,y,a,b)\sum_{\sigma\in[m^2]^D_{\geq 0}}c_\sigma\widehat{P}_a^x(\sigma)\cdot\widehat{Q}_b^y(\sigma),$$

where $c_\sigma=c_{\sigma_1}\cdots c_{\sigma_D}$, and $\{c_i\}_{i=0}^{m^2-1}$ is given in Fact 46. Reject if

$$\mathrm{val}_D\left(\{P_a^x\},\{Q_b^y\}\right)<\beta.$$

2. Check if the operators sum up to the identity by checking
   - For all x, y and $\sigma \neq 0^D$, it should hold that
   $$\sum_a \widehat{P}_a^x(\sigma) = \sum_b \widehat{Q}_b^y(\sigma) = 0.$$
   - For all x, y, and $\sigma = 0^D$, it should hold that
   $$\sum_a \widehat{P}_a^x(\sigma) = \sum_b \widehat{Q}_b^y(\sigma) = 1.$$

   Reject if any of the above equalities fails.
3. For each x, y, a, b, run the positivity testing algorithm described in Section 4 on P_a^x and Q_b^y with parameters $\beta \leftarrow 4\delta$ and $\delta \leftarrow 2\delta$. Reject if any of the positivity testings fails.
4. Accept.

Time complexity

We upper bound the time complexity of each step.

1. Certificate length: The certificate contains the non-zero Fourier coefficients of degree-d operators acting on D qudits. Each degree-d operator consists of
   $$\sum_{d=0}^{d} \binom{D}{d} \cdot (m^2-1)^d \leq d(m^2-1)^d D^d$$
   coefficients, each s_w bits. Hence, the length of the certificate is $O(stdm^{2d}D^d s_w)$.
2. To compute the game value, we need to enumerate over all x, y, a, b, σ and compute a sum of products. This takes time
   $$s^2t^2(m^2-1)^d D^d.$$
3. Checking if the operators sum up to the identity takes linear time in certificate length as it involves only summation over Fourier coefficients.
4. Each positivity testing takes time as specified in Theorem 23, which is
   $$\exp\left(\text{poly}\left(m^d, 1/\delta\right)\right) \cdot D^{O(d)}.$$

By the choices of parameters, the overall running time is upper bounded by

$$\text{poly}\left(s, \text{eexp}\left(t, \log\left(\frac{1}{\rho}\right), \frac{1}{\varepsilon}\right)\right).$$

Completeness

Suppose $\omega^*(G, \psi_{AB}) \geq \beta + \varepsilon$. Then by Theorem 28, there exists a strategy (P_a^x, Q_b^y) that uses D copies of ψ_{AB} with game value $\text{val}_D\left(\{P_a^x\}, \{Q_b^y\}\right) \geq \beta + \varepsilon/2$. Let f be the smoothing map in Lemma 62, and let $P_a^{x,(1)} = f(P_a^x)$ and $Q_b^{y,(1)} = f(Q_b^y)$. Then $\left\{P_a^{x,(1)}\right\}, \left\{Q_b^{y,(1)}\right\}$ are of degree at most d and satisfy

1. For all x, y, we have $\sum_a = P_a^{x,(1)} = \sum_b Q_b^{y,(1)} = \mathbb{1}$ (since f is linear and unital)
2. For all x, y, a, b, $\left|\left|\left|P_a^{x,(1)}\right|\right|\right|_2 \leq 1$ and $\left|\left|\left|Q_b^{y,(1)}\right|\right|\right|_2 \leq 1$.
3. For all x, y, a, b, $\left|\mathrm{Tr}\left(\left(P_a^{x,(1)} \otimes Q_b^{y,(1)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^x \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right)\right| \leq \delta$.
4. For all x, y, a, b, $m^{-D}\,\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right) \leq \delta$ and $m^{-D}\,\mathrm{Tr}\,\zeta\left(Q_b^{y,(1)}\right) \leq \delta$.

We observe that Lemma 62 also guarantees the Fourier coefficients of $P_a^{x,(1)}$ and $Q_b^{y,(1)}$ have absolute values bounded by 1. This allows us to truncate the strategy. For each Fourier coefficient we preserve s_w digits and by Lemma 68 get $\left\{P_a^{x,(2)}\right\}, \left\{Q_b^{y,(2)}\right\}$ satisfying

1. For all x, y, $\sum_a P_a^{x,(2)} = \sum_b Q_b^{y,(2)} = \mathbb{1}$.
2. For all x, y, a, b, $\left|\left|\left|P_a^{x,(2)}\right|\right|\right|_2 \leq 1$ and $\left|\left|\left|Q_b^{y,(2)}\right|\right|\right|_2 \leq 1$;
3. For all x, y, a, b, $\left|\mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes Q_b^{y,(2)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^{x,(1)} \otimes Q_b^{y,(1)}\right)\psi_{AB}^{\otimes n}\right)\right| \leq \delta$,
4. For all x, y, a, b, $m^{-D}\,\mathrm{Tr}\,\zeta\left(P_a^{x,(2)}\right) \leq 2\delta$ and $m^{-D}\,\mathrm{Tr}\,\zeta\left(Q_b^{y,(2)}\right) \leq 2\delta$.

This pseudo-strategy is the certificate. Specifically, by Lemma 67 the game value is

$$\mathrm{val}_D\left(\left\{P_a^{x,(2)}\right\}, \left\{Q_b^{y,(2)}\right\}\right) \geq \beta + \varepsilon/2 - 2\delta t^2 = \beta + \varepsilon/2 - \frac{\varepsilon^2}{2tC_{pt}} \geq \beta,$$

and the first check is passed. Also, by item 4, the positivity testings can also be passed.

Soundness

Suppose that there exists a certificate that passes all the testings, then there exists a degree-d pseudo-strategy $\left\{P_a^{x,(1)}\right\}, \left\{Q_b^{y,(1)}\right\}$ satisfying

- By the game value testing,

$$\mathrm{val}_D\left(\left\{P_a^{x,(1)}\right\}, \left\{Q_b^{y,(1)}\right\}\right) \geq \beta.$$

- By "summing up to the identity" testings, for all x, y

$$\sum_a P_a^{x,(1)} = \mathbb{1}, \text{ and } \sum_b Q_b^{y,(1)} = \mathbb{1}.$$

- By the positivity testings, for all x, y, a, b

$$\frac{1}{m^D}\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right) \leq 6\delta, \text{ and } \frac{1}{m^D}\mathrm{Tr}\,\zeta\left(Q_b^{y,(1)}\right) \leq 6\delta.$$

We then apply Lemma 65 to get a strategy $\left\{P_a^{x,(2)}\right\}$ and $\left\{Q_b^{y,(2)}\right\}$. It holds that for each $x \in \mathcal{X}$

$$\begin{aligned}\sum_{a\in\mathcal{A}} \left|\left|\left|P_a^{x,(2)} - P_a^{x,(1)}\right|\right|\right|_2^2 &\leq 3(t+1)\left(\sum_{a\in A}\frac{1}{m^D}\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right)\right) + 6\sqrt{t}\left(\sum_{a\in A}\frac{1}{m^D}\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right)\right)^{1/2} \\ &\leq 18t(t+1)\delta + 6\sqrt{6}t\sqrt{\delta} \leq \frac{18\varepsilon_{rd}^2}{C_{pt}} + \frac{6\sqrt{6}\varepsilon_{rd}}{\sqrt{C_{pt}}} \leq \frac{18 + 6\sqrt{6C_{pt}}}{C_{pt}}\varepsilon_{rd} \leq \varepsilon_{rd}.\end{aligned}$$

Similarly, for each $y \in \mathcal{Y}$ we have

$$\sum_{b\in\mathcal{B}} \left|\left|\left|Q_b^{y,(2)} - Q_b^{y,(1)}\right|\right|\right|_2^2 \leq \varepsilon_{rd}.$$

We get a strategy $\left\{P_a^{x,(2)}\right\}$ and $\left\{Q_b^{y,(2)}\right\}$ with game value

$$\begin{aligned}
&\left|\mathrm{val}_D\left(\left\{P_a^{x,(2)}\right\},\left\{Q_b^{y,(2)}\right\}\right)-\mathrm{val}_D\left(\left\{P_a^{x,(1)}\right\},\left\{Q_b^{y,(1)}\right\}\right)\right| \\
\leq\ &\left|\mathrm{val}_D\left(\left\{P_a^{x,(2)}-P_a^{x,(1)}\right\},\left\{Q_b^{y,(2)}\right\}\right)\right|+\left|\mathrm{val}_D\left(\left\{P_a^{x,(1)}\right\},\left\{Q_b^{y,(2)}-Q_b^{y,(1)}\right\}\right)\right| \\
\leq\ &\sum_{x,y,a,b}\mu(x,y)\left(\left\|P_a^{x,(2)}-P_a^{x,(1)}\right\|_2\left\|Q_b^{y,(2)}\right\|_2+\left\|P_a^{x,(1)}\right\|_2\left\|Q_b^{y,(2)}-Q_b^{y,(1)}\right\|_2\right) \\
\leq\ &\left(\sum_b\sum_{x,a}\mu_A(x)\left\|P_a^{x,(2)}-P_a^{x,(1)}\right\|_2^2\right)^{1/2}\left(\sum_a\sum_{y,b}\mu_B(y)\left\|Q_b^{y,(2)}\right\|_2^2\right)^{1/2} \\
&\quad+\left(\sum_b\sum_{x,a}\mu_A(x)\left\|P_a^{x,(1)}\right\|_2^2\right)^{1/2}\left(\sum_a\sum_{y,b}\mu_B(y)\left\|Q_b^{y,(2)}-Q_b^{y,(1)}\right\|_2^2\right)^{1/2} \\
&\qquad\text{(Cauchy-Schwarz)} \\
\leq\ &2t\sqrt{t\varepsilon_{rd}}.
\end{aligned}$$

Thus there exists a strategy with game value

$$\mathrm{val}_D\left(\left\{P_a^{x,(2)}\right\},\left\{Q_b^{y,(2)}\right\}\right)>\beta-2t\sqrt{t\varepsilon_{rd}}=\beta-\varepsilon.$$ ◀

B.3.2 NP-hardness

Proof of Proposition 29. The noisy MIP* verifier V^* from an MIP verifier $V=(\mathsf{Alg}_Q,\mathsf{Alg}_V)$

Setup: Flip two unbiased coins $\boldsymbol{b},\boldsymbol{c}\sim\{0,1\}$. Sample questions $(\boldsymbol{x},\boldsymbol{y})\sim\mathsf{Alg}_Q(\mathsf{input})$. With probability 1/2 each, perform one of the following ten tests.

Verify: Distribute the questions as follows

- Player $\boldsymbol{b}$: give $\boldsymbol{x}$; receive $\boldsymbol{a}$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{y}$; receive $\boldsymbol{b}$

Accept if $V(\mathsf{input},\boldsymbol{x},\boldsymbol{y})$ accepts on $\boldsymbol{a},\boldsymbol{b}$.

Consistency: Distribute the questions as follows: if $\boldsymbol{c}=0$

- Player $\boldsymbol{b}$: give $\boldsymbol{x}$; receive $\boldsymbol{a}$,
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}$; receive $\boldsymbol{b}$,

otherwise

- Player $\boldsymbol{b}$: give $\boldsymbol{y}$; receive $\boldsymbol{a}$,
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{y}$; receive $\boldsymbol{b}$,

Accept if $\boldsymbol{a}=\boldsymbol{b}$.

Completeness. If input is satisfiable, the value-1 strategy for V is also a value-1 strategy for V^*.

Soundness. In the consistency test, with probability 1/2 both provers get a question x. Hence the probability for the provers to pass the consistency test of x is at least $1-4\epsilon$. If Alice and Bob sharing n copies of a noisy m-dimensional MES ψ_{AB}, it means that

$$\mathop{\mathbb{E}}_x\sum_{a\in\mathcal{A}}\mathrm{Tr}\left(\left(P_a^x\otimes P_a^x\right)\psi_{AB}^{\otimes n}\right)\geq 1-4\epsilon.$$

Using the Fourier expansion of $P_a^x = \sum_\sigma \widehat{P_a^x}(\sigma)\,\mathcal{P}_\sigma$. the condition above is equivalent to

$$\mathop{\mathbb{E}}_x \sum_a \sum_\sigma \rho^{|\sigma|} \widehat{P_a^x}(\sigma)^2 \geq 1 - 4\epsilon.$$

Notice that $\|\!|P_a^x|\!\|_2^2 = \sum_\sigma \widehat{P_a^x}(\sigma)^2$, and for all x, $\sum_a \|\!|P_a^x|\!\|_2^2 \leq 1$. Hence

$$\begin{aligned}
\mathop{\mathbb{E}}_x \sum_a \sum_\sigma \rho^{|\sigma|} \widehat{P_a^x}(\sigma)^2 &\leq \mathop{\mathbb{E}}_x \sum_a \left[\widehat{P_a^x}(\emptyset)^2 + \rho \sum_{\sigma \neq \emptyset} \widehat{P_a^x}(\sigma)^2 \right] \\
&\leq \mathop{\mathbb{E}}_x \sum_a \left[\widehat{P_a^x}(\emptyset)^2 + \rho(\|\!|P_a^x|\!\|_2^2 - \widehat{P_a^x}(\emptyset)^2) \right] \\
&\leq \rho + (1-\rho) \mathop{\mathbb{E}}_x \sum_a \widehat{P_a^x}(\emptyset)^2 .
\end{aligned}$$

Therefore,

$$\mathop{\mathbb{E}}_x \sum_a \widehat{P_a^x}(\emptyset)^2 \geq 1 - \frac{4\epsilon}{1-\rho}. \tag{26}$$

On the other hand, for all x, $\sum_a \widehat{P_a^x}(\emptyset) = 1$. For each x, let a_x be the answer that maximizes $\widehat{P_a^x}(\emptyset)$. Then $\sum_a \widehat{P_a^x}(\emptyset)^2 \leq \widehat{P_{a_x}^x}(\emptyset) \sum_a \widehat{P_a^x}(\emptyset) = \widehat{P_{a_x}^x}(\emptyset)$, and

$$\mathop{\mathbb{E}}_x \widehat{P_{a_x}^x}(\emptyset) \geq 1 - \frac{4\epsilon}{1-\rho}.$$

Similarly, for each y, let b_y be the answer that maximizes $\widehat{Q_b^y}(\emptyset)$, and then

$$\mathop{\mathbb{E}}_{x,y} \widehat{Q_{b_y}^y}(\emptyset) \geq 1 - \frac{4\epsilon}{1-\rho}.$$

In the deterministic strategy, Alice answers a_x for question x and Bob answers b_y for question y. The difference in the probability of satisfying V between the original strategy and the deterministic strategy is

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{x,y} \sum_{a,b} \mathrm{Tr}\left[(P_a^x \otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right] V(x,y,a,b) - \mathop{\mathbb{E}}_{x,y} V(x,y,a_x,b_y)| \\
&= \mathop{\mathbb{E}}_{x,y} \left(1 - \mathrm{Tr}\left[\left(P_{a_x}^x \otimes Q_{b_y}^y\right)\psi_{AB}^{\otimes n}\right]\right) V(x,y,a_x,b_y) \\
&\qquad + \mathop{\mathbb{E}}_{x,y} \sum_{\substack{a \neq a_x \\ b \neq b_y}\ \text{or}} \mathrm{Tr}\left[(P_a^x \otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right] V(x,y,a,b) \\
&\leq \mathop{\mathbb{E}}_{x,y} \left(1 - \mathrm{Tr}\left[\left(P_{a_x}^x \otimes Q_{b_y}^y\right)\psi_{AB}^{\otimes n}\right]\right) + \mathop{\mathbb{E}}_{x,y} \sum_{\substack{a \neq a_x \\ b \neq b_y}\ \text{or}} \mathrm{Tr}\left[(P_a^x \otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right]
\end{aligned}$$

where we use the fact that $V(x,y,a,b) \leq 1$ for all x,y,a,b.
Writing $1 = \sum_{a,b} \mathrm{Tr}\left[(P_a^x \otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right]$, we get that the expression above equals

$$\begin{aligned}
&2 \mathop{\mathbb{E}}_{x,y} \sum_{\substack{a \neq a_x \\ b \neq b_y}\ \text{or}} \mathrm{Tr}\left[(P_a^x \otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right] \\
&= 2 \mathop{\mathbb{E}}_{x,y} \sum_{a \neq a_x, b} \mathrm{Tr}\left[(P_a^x \otimes Q_b^y)\,\psi_{AB}^{\otimes n}\right] + 2 \mathop{\mathbb{E}}_{x,y} \sum_{b \neq b_y} \mathrm{Tr}\left[\left(P_{a_x}^x \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right] \\
&\leq 2 \mathop{\mathbb{E}}_{x,y} \left[\sum_{a \neq a_x} \widehat{P_a^x}(\emptyset) + \sum_{b \neq b_y} \widehat{Q_b^y}(\emptyset) \right] \\
&\leq \frac{16\epsilon}{1-\rho}.
\end{aligned}$$

The probability for the original strategy to satisfy V is at least $1-2\varepsilon$, so the probability for the deterministic strategy to satisfy V is at least $1-2\varepsilon-16\varepsilon/(1-\rho)$. ◀

B.4 MIP* Protocol for RE with $O(1)$-size Answers

Subset tester for the Hadamard code

Let $k \le n$ and D be a distribution on the subsets of $\mathbb{F}_2^n$ with size k. Flip an unbiased coin $\boldsymbol{b} \sim \{0,1\}$. Sample $\boldsymbol{F} = \{x_1, \ldots, x_k\} \sim D$ and a uniformly random $\boldsymbol{y} \in \mathbb{F}_2^n$, Perform one of the following three subtests with equal probability.

Subtest 1: Perform one of the following checks with equal probability.

Check 1: Distribute the question as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{F}$ and $\boldsymbol{y}$; receive $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{c}, \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_k) \in \mathbb{F}_2^{2k+1}$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{F}$, receive $(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_k) \in \mathbb{F}_2^k$.

Accept if $\boldsymbol{a}_i + \boldsymbol{c} = \boldsymbol{a}'_i$ and $\boldsymbol{a}_i = \boldsymbol{d}_i$ for all i.

Check 2: Distribute the question as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{F}$ and $\boldsymbol{y}$; receive $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{c}, \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_k) \in \mathbb{F}_2^{2k+1}$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{y}$, receive $\boldsymbol{e} \in \mathbb{F}_2$.

Accept if $\boldsymbol{a}_i + \boldsymbol{c} = \boldsymbol{a}'_i$ for all i, and $\boldsymbol{e} = \boldsymbol{c}$.

Check 3: Distribute the question as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{F}$ and $\boldsymbol{y}$; receive $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{c}, \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_k) \in \mathbb{F}_2^{2k+1}$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{F} + \boldsymbol{y} = \{\boldsymbol{x}_1 + \boldsymbol{y}, \ldots, \boldsymbol{x}_k + \boldsymbol{y}\}$, receive $(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_k) \in \mathbb{F}_2^k$.

Accept if $\boldsymbol{a}_i + \boldsymbol{c} = \boldsymbol{a}'_i$ and $\boldsymbol{a}'_i = \boldsymbol{d}_i$ for all i.

Subtest 2: Distribute the question as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{F} + \boldsymbol{y} = \{\boldsymbol{x}_1 + \boldsymbol{y}, \ldots, \boldsymbol{x}_k + \boldsymbol{y}\}$; receive $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}_i + \boldsymbol{y}$ for a random i, receive $\boldsymbol{d}$.

Accept if $\boldsymbol{a}_i = \boldsymbol{d}$.

Subtest 3: Perform one of the following checks with equal probability

Check 1: Distribute the question as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{F}$; receive $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{F}$; receive $(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_k)$.

Accept if $\boldsymbol{a}_i = \boldsymbol{d}_i$ for all i.

Check 2: Distribute the question as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{x}_i + \boldsymbol{y}$ for a random i; receive $\boldsymbol{a}$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}_i + \boldsymbol{y}$ for a random i; receive $\boldsymbol{d}$.

Accept if $\boldsymbol{a} = \boldsymbol{d}$.

Proof of Proposition 31. Let $F + y = (x_1 + y, \ldots, x_k + y)$. Let

$$\Omega = \{(a, c, a') \mid a_i + c = a'_i \text{ for all } i \in [k]\}.$$

The set Ω is the set of valid answer tuples for Alice in **Subtest 1**; we also use Ω to denote the *event* that Alice's answers are valid. Winning the subset tester with probability $1-\varepsilon$ implies that winning each subtest with a probability of at least $1-3\varepsilon$. Furthermore, winning **Subtest 1** with a probability of at least $1-3\varepsilon$ implies that when Alice gets question (F, y) and Bob gets Player 1's questions:

$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\sim D_{\mathsf{Unif}}}\Pr[a_1=b_1\wedge\ldots\wedge a_k=b_k\wedge\Omega\mid q_A=(F,y),q_B=F]\geq 1-18\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\sim D_{\mathsf{Unif}}}\Pr[c=d\wedge\Omega\mid q_A=(F,y),q_B=y]\geq 1-18\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\sim D_{\mathsf{Unif}}}\Pr[a'_1=b_1\wedge\ldots\wedge a'_k=b_k\wedge\Omega\mid q_A=(F,y),q_B=F+y]\geq 1-18\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\sim D_{\mathsf{Unif}}}\Pr[\Omega\mid q_A=(F,y)]\geq 1-6\varepsilon,$$

for all $i\in[k]$; winning **Subtest 2** with a probability of at least $1-3\varepsilon$ implies that when Alice gets Player 0's question and Bob gets Player 1's question

$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\sim D_{\mathsf{Unif}}}\Pr[a_i=d\mid q_A=F+y,q_B=x_i+y]\geq 1-6k\varepsilon;$$

and winning **Subtest 3** with a probability of at least $1-3\varepsilon$ implies that when Alice gets Player 0's question and Bob gets Player 1's question

$$\mathop{\mathbb{E}}_{F\sim D}\Pr[a_1=b_1\wedge\ldots\wedge a_k=b_k\mid q_A=q_B=F]\geq 1-12\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\sim D_{\mathsf{Unif}}}\Pr[a=b\mid q_A=q_B=x_i+y]\geq 1-12k\varepsilon\quad\text{for all } i.$$

In terms of the measurements and the state $|\psi\rangle$, these conditions are equivalent to

$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{\substack{a,c,a':\\ a_i+c=a'_i\forall i}}\langle\psi|\,M^{F,y}_{a,c,a'}\otimes N^F_a\,|\psi\rangle\geq 1-18\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{\substack{a,c,a':\\ a_i+c=a'_i\forall i}}\langle\psi|\,M^{F,y}_{a,c,a'}\otimes N^y_c\,|\psi\rangle\geq 1-18\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{\substack{a,c,a':\\ a_i+c=a'_i\forall i}}\langle\psi|\,M^{F,y}_{a,c,a'}\otimes N^{F+y}_{a'}\,|\psi\rangle\geq 1-18\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{\substack{a,c,a':\\ a_i+c=a'_i\forall i}}\langle\psi|\,M^{F,y}_{a,c,a'}\otimes \mathbb{1}_B\,|\psi\rangle\geq 1-6\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{a\in\mathbb{F}_2^k}\langle\psi|\,N^{F+y}_a\otimes N^{x_i+y}_{a_i}\,|\psi\rangle\geq 1-6k\varepsilon\quad\text{for all } i$$
$$\mathop{\mathbb{E}}_{F\sim D}\sum_{a\in\mathbb{F}_2^k}\langle\psi|\,N^F_a\otimes N^F_a\,|\psi\rangle\geq 1-12\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{a\in\mathbb{F}_2}\langle\psi|\,N^{x_i+y}_a\otimes N^{x_i+y}_a\,|\psi\rangle\geq 1-12k\varepsilon\quad\text{for all } i.$$

We define binary observables

$$M^{x_i|F,y}=\sum_{a,c,a'}(-1)^{a_i}M^{F,y}_{a,c,a'}\qquad M^{y|F,y}=\sum_{a,c,a'}(-1)^{c}M^{F,y}_{a,c,a'}\qquad M^{x_i+y|F,y}=\sum_{a,c,a'}(-1)^{a'_i}M^{F,y}_{a,c,a'}$$
$$N^{x_i|F}=\sum_b(-1)^{b_i}N^F_b\qquad N^y=N^y_0-N^y_1\qquad N^{x_i+y|F+y}=\sum_b(-1)^{b_i}N^{F+y}_b.$$

We can prove

$$\begin{aligned}
&\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}} \langle\psi| M^{x_i|F,y}\otimes N^{x_i|F}|\psi\rangle \\
&= \mathop{\mathbb{E}}_{x\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\Big[\Pr[a_i=b_i\wedge\Omega \mid q_A=(F,y),q_B=F] \\
&\qquad -(\Pr[a_i\neq b_i \mid q_A=(F,y),q_B=F]-\Pr[a_i=b_i\wedge\overline{\Omega} \mid q_A=(F,y),q_B=F])\Big] \\
&\geq \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\Big[\Pr[a_i=b_i\wedge\Omega \mid q_A=(F,y),q_B=F] \\
&\qquad\qquad -(1-\Pr[a_i=b_i\wedge\Omega \mid q_A=(F,y),q_B=F])\Big] \\
&= \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\Big[2\Pr[a_i=b_i\wedge\Omega \mid q_A=(F,y),q_B=F]-1\Big] \\
&\geq \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\Big[2\Pr[a_1=b_1\wedge\ldots\wedge a_k=b_k\wedge\Omega \mid q_A=(F,y),q_B=F]-1\Big] \\
&\geq 1-36\varepsilon,
\end{aligned}$$

which implies that $\mathbb{E}_{F\sim D}\mathbb{E}_{y\in D_{\mathsf{Unif}}}\|M^{x_i|F,y}\otimes \mathbb{1}_B|\psi\rangle-\mathbb{1}_A\otimes N^{x_i|F}|\psi\rangle\|^2\leq 72\varepsilon$ by expanding the vector norm. Similarly, from the two other checks of **Subtest 1**,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|M^{y|F,y}\otimes \mathbb{1}_B|\psi\rangle-\mathbb{1}_A\otimes N^{y}|\psi\rangle\|^2\leq 72\varepsilon \\
&\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|M^{x_i+y|F,y}\otimes \mathbb{1}_B|\psi\rangle-\mathbb{1}_A\otimes N^{x_i+y|F+y}|\psi\rangle\|^2\leq 72\varepsilon.
\end{aligned}$$

Applying a similar argument to the probability of the event Ω, we can also show

$$\begin{aligned}
&\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi| M^{x_i|F,y}M^{y|F,y}M^{x_i+y|F,y}\otimes\mathbb{1}_B|\psi\rangle \\
&= \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{a,c,a'}(-1)^{a_i+c+a_i'}\langle\psi| M^{F,y}_{a,c,a'}\otimes\mathbb{1}_B|\psi\rangle \\
&= \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}2\Pr[a_i+c=a_i' \mid q_A=(F,y)]-1 \\
&\geq \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}2\Pr[\Omega \mid q_A=(F,y)]-1\geq 1-12\varepsilon.
\end{aligned}$$

Next, we would like to replace $M^{x_i|F,y}$ by $N^{x_i|F}$, $M^{y|F,y}$ by N^y and $M^{x_i+y|F,y}$ by $N^{x_i+y|F+y}$ and show

$$\Big|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\,\mathbb{1}_A\otimes N^{x_i+y|F+y}N^yN^{x_i|F}|\psi\rangle-1\Big|\leq 38\sqrt{\varepsilon}. \tag{27}$$

In the first step

$$\begin{aligned}
&\Big|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi| M^{x_i|F,y}M^{y|F,y}(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})|\psi\rangle\Big| \\
&\leq \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|M^{y|F,y}M^{x_i|F,y}\otimes\mathbb{1}_B|\psi\rangle\|\cdot\|(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})|\psi\rangle\| \\
&\qquad\text{(Cauchy-Schwarz)} \\
&= \mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})|\psi\rangle\| \\
&\leq \sqrt{\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})|\psi\rangle\|^2}\quad\text{(Jensen)} \\
&\leq 6\sqrt{2\varepsilon}.
\end{aligned}$$

Similarly,

$$|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, M^{x_i|F,y}\otimes N^{x_i+y|F,y}\cdot(M^{y|F,y}\otimes \mathbb{1}_B-\mathbb{1}_A\otimes N^y)\,|\psi\rangle|\le 6\sqrt{2\varepsilon}$$
$$|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, \mathbb{1}_A\otimes N^{x_i+y|F+y}N^y\cdot(M^{x_i|F,y}\otimes \mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i|F})\,|\psi\rangle|\le 6\sqrt{2\varepsilon}.$$

Hence

$$|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, \mathbb{1}_A\otimes N^{x_i+y|F+y}N^yN^{x_i|F}\,|\psi\rangle-1|\le 18\sqrt{2\varepsilon}+12\varepsilon\le 38\sqrt{\varepsilon}.$$

On the other hand, from **Subtest 2**, we have that for all $i\in[k]$

$$\begin{aligned}&\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, N^{x_i+y|F+y}\otimes N^{x_i+y}\,|\psi\rangle\\ &=2\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\Pr[a_i=b\mid q_A=F+y,q_B=x_i+y]-1\ge 1-12k\varepsilon,\end{aligned}$$

which implies that

$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|(N^{x_i+y|F+y}\otimes \mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y})\,|\psi\rangle\|^2\le 24k\varepsilon.$$

From **Subtest 3**, with similar reasoning we know

$$\mathop{\mathbb{E}}_{F\sim D}\|(N^{x_i|F}\otimes \mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i|F})\,|\psi\rangle\|^2\le 48\varepsilon$$
$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|(N^{x_i+y}\otimes \mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y})\,|\psi\rangle\|^2\le 48k\varepsilon\quad\text{for all } i.$$

Then

$$\begin{aligned}&\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, \mathbb{1}_A\otimes N^{x_i+y|F+y}N^yN^{x_i|F}\,|\psi\rangle\\ &\approx_{\sqrt{24k\varepsilon}}\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, N^{x_i+y}\otimes N^yN^{x_i|F}\,|\psi\rangle\\ &\approx_{\sqrt{48\varepsilon}}\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, N^{x_i+y}N^{x_i|F}\otimes N^y\,|\psi\rangle\\ &\approx_{\sqrt{48k\varepsilon}}\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, N^{x_i|F}\otimes N^{x_i+y}N^y\,|\psi\rangle\end{aligned}$$

Hence Equation (27) implies that

$$|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, N^{x_i|F}\otimes N^{x_i+y}N^y\,|\psi\rangle-1|\le(45+12\sqrt{k})\sqrt{\varepsilon}.\tag{28}$$

Let $C_1:=45+12\sqrt{k}$. Let $\tilde{N}_u=\frac{1}{2^n}\sum_{y\in\mathbb{F}_2^n}(-1)^{u\cdot y}N^y$ and $G_u=(\tilde{N}_u)^2$. Since each N^y is a binary observable, $\{G_u\}$ is a POVM. It can be checked that $N^y=\sum_{u\in\mathbb{F}_2^n}(-1)^{u\cdot y}\tilde{N}_u$. Averaging over $F\sim D$, the consistency between $\left\{N_0^{x_i|F},N_1^{x_i|F}\right\}$ and $\left\{\sum_{u:u\cdot x_i=0}G_u,\sum_{u:u\cdot x_i=1}G_u\right\}$ is

$$\begin{aligned}&\mathop{\mathbb{E}}_{F\sim D}\frac{1}{2}(1+\langle\psi|\sum_u(-1)^{u\cdot x_i}N^{x_i|F}\otimes G_u\,|\psi\rangle)\\ &=\frac{1}{2}+\frac{1}{2}\langle\psi|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y,z\in D_{\mathsf{Unif}}}\sum_u(-1)^{u\cdot(x_i+y+z)}N^{x_i|F}\otimes N^yN^z\,|\psi\rangle\\ &=\frac{1}{2}+\frac{1}{2}\langle\psi|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{z\in D_{\mathsf{Unif}}}N^{x_i|F}\otimes N^{x_i+z}N^z\,|\psi\rangle\approx_{\frac{C_1}{2}\sqrt{\varepsilon}}1,\end{aligned}$$

which follows Equation (28). We consider the Naimark's dialation of $\{G_u\}$ on $\mathcal{H}\otimes\mathcal{H}'$ denoted by $\left\{\hat{G}_u\right\}$, which is a projective measurement. There exists $|aux\rangle \in \mathcal{H}'$ such that averaging over $F\sim D$, the consistency between $\left\{N_0^{x_i|F}\otimes \mathbb{1}_{\mathcal{H}'}, N_1^{x_i|F}\otimes \mathbb{1}_{\mathcal{H}'}\right\}$ and $\left\{\sum_{u:u\cdot x_i=0}\hat{G}_u, \sum_{u:u\cdot x_i=1}\hat{G}_u\right\}$ with respect to $|\psi'\rangle = |\psi\rangle\otimes|aux\rangle\otimes|aux\rangle$ is

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{F\sim D}\sum_{a=0,1}\langle\psi'|\left(N_a^{x_i|F}\otimes\mathbb{1}_{\mathcal{H}'}\right)\otimes\Big(\sum_{u:u\cdot x_i=a}\hat{G}_u\Big)|\psi'\rangle\\
&=\mathop{\mathbb{E}}_{F\sim D}\sum_{a=0,1}\langle\psi|\,N_a^{x_i|F}\otimes\left(\sum_{u:u\cdot x_i=a}(\mathbb{1}\otimes\langle aux|)\hat{G}_u(\mathbb{1}\otimes|aux\rangle)\right)|\psi\rangle\\
&=\mathop{\mathbb{E}}_{F\sim D}\sum_{a=0,1}\langle\psi|\,N_a^{x_i|F}\otimes\left(\sum_{u:u\cdot x_i=a}G_u\right)|\psi\rangle\\
&\approx_{C_1/2\sqrt{\varepsilon}}1.
\end{aligned}
$$

Since both $\left\{N_a^{x_i|F}\otimes\mathbb{1}_{\mathcal{H}'}\right\}$ and $\left\{\sum_{u:u\cdot x_i=a}\hat{G}_u\right\}$ are projective measurements, their consistency implies that

$$
\mathop{\mathbb{E}}_{F\sim D}\sum_{d=0,1}\|N_d^{x_i|F}\otimes\mathbb{1}_{\mathcal{H}'}\,|\psi'\rangle-\sum_{u:u\cdot x_i=d}\hat{G}_u\,|\psi'\rangle\|^2\le C_1\sqrt{\varepsilon}.
$$

Next, notice that

$$
N_a^F=N_{a_k}^{x_k|F}\dots N_{a_1}^{x_1|F}\text{ and }\sum_{\substack{u:u\cdot x_i=a_i\\ \forall i\in[k]}}\hat{G}_u=\left(\sum_{u:u\cdot x_k=a_k}\hat{G}_u\right)\dots\left(\sum_{u:u\cdot x_1=a_1}\hat{G}_u\right)\dots\left(\sum_{u:u\cdot x_k=a_k}\hat{G}_u\right)
$$

Then by Lemma 74

$$
\mathop{\mathbb{E}}_{F\sim D}\sum_{a\in\mathbb{F}_2^k}\|N_a^F\otimes\mathbb{1}_{\mathcal{H}'}\otimes\mathbb{1}_B\,|\tilde{\psi}\rangle-\mathbb{1}_A\otimes\sum_{\substack{u:u\cdot x_i=a_i\\ \forall i\in[k]}}\hat{G}_u\,|\tilde{\psi}\rangle\|^2\le(2k-1)^2C_1\sqrt{\varepsilon},
$$

which completes the proof. ◀

The answer reduced verifier V^{AR}

Setup Flip two unbiased coins $\boldsymbol{b},\boldsymbol{c}\sim\{0,1\}$. Sample questions $(\boldsymbol{x}_0,\boldsymbol{x}_1)\sim\mathsf{Alg}_Q(\mathsf{input})$. Sample a view $\boldsymbol{I}_0,\boldsymbol{I}_1,\boldsymbol{J}\sim V_{\mathrm{PCPP}}(\mathsf{input},\boldsymbol{x}_0,\boldsymbol{x}_1)$. Set $\boldsymbol{J}'=\mu_{\ell_2}(\boldsymbol{J})$. Randomly select $\boldsymbol{I}_0',\boldsymbol{I}_1'\subseteq[2^{\ell_1}]$ and $\boldsymbol{J}''\subseteq[2^{\ell_2}]$ such that $|\boldsymbol{I}_0'|=|\boldsymbol{I}_1'|=|\boldsymbol{J}''|=\kappa$, which is a sufficiently large constant. Details about how to choose κ can be found in the proof below. Set $\boldsymbol{T}_0=\boldsymbol{I}_0\cup\boldsymbol{I}_0'$, $\boldsymbol{T}_1=\boldsymbol{I}_1\cup\boldsymbol{I}_1'$ and $\boldsymbol{U}=\boldsymbol{J}'\cup\boldsymbol{J}''$.

With probability 1/10 each, perform one of the following ten tests.

Verify : Distribute the questions as follows:

- Player $\boldsymbol{b}$: give $(\boldsymbol{x}_0,\boldsymbol{x}_1),\boldsymbol{T}_0,\boldsymbol{T}_1,\boldsymbol{U}$; receive $\boldsymbol{a}_0,\boldsymbol{a}_1,\boldsymbol{a}_2$.

Accept if $V_{\mathrm{PCPP}}(\mathsf{input},\boldsymbol{x}_0,\boldsymbol{x}_1)$ accepts on $\boldsymbol{a}_0|_{\boldsymbol{I}_0}$, $\boldsymbol{a}_1|_{\boldsymbol{I}_1}$ and $\boldsymbol{a}_2|_{\boldsymbol{J}'}$.

Cross check:

Consistency test: Distribute the questions as follows:

- Player $\boldsymbol{b}$: give $(\boldsymbol{x}_0,\boldsymbol{x}_1),\boldsymbol{T}_0,\boldsymbol{T}_1,\boldsymbol{U}$; receive $\boldsymbol{a}_0,\boldsymbol{a}_1,\boldsymbol{a}_2$.

- Player $\bar{\boldsymbol{b}}$: give $(\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{T}_0, \boldsymbol{T}_1, \boldsymbol{U}$; receive $\boldsymbol{a}'_0, \boldsymbol{a}'_1, \boldsymbol{a}'_2$

Accept if $\boldsymbol{a}_0 = \boldsymbol{a}'_0$, $\boldsymbol{a}_1 = \boldsymbol{a}'_1$ and $\boldsymbol{a}_2 = \boldsymbol{a}'_2$.

Answer cross-check: Distributed the questions as follows:

- Player $\boldsymbol{b}$: give $(\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{T}_0, \boldsymbol{T}_1, \boldsymbol{U}$; receive $\boldsymbol{a}_0, \boldsymbol{a}_1, \boldsymbol{a}_2$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}_{\boldsymbol{c}}, \boldsymbol{T}_{\boldsymbol{c}}$; receive $\boldsymbol{a}'_{\boldsymbol{c}}$

Accept if $\boldsymbol{a}_{\boldsymbol{c}} = \boldsymbol{a}_{\boldsymbol{c}'}$.

Answer consistency check: Distributed the questions as follows:

- Player $\boldsymbol{b}$: give $\boldsymbol{x}_{\boldsymbol{c}}, \boldsymbol{T}_{\boldsymbol{c}}$; receive $\boldsymbol{a}_{\boldsymbol{c}}$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}_{\boldsymbol{c}}, \boldsymbol{T}_{\boldsymbol{c}}$; receive $\boldsymbol{a}'_{\boldsymbol{c}}$

Accept if $\boldsymbol{a}_{\boldsymbol{c}} = \boldsymbol{a}_{\boldsymbol{c}'}$.

Proof cross-check: Distribute the questions as follows:

- Player $\boldsymbol{b}$: give $(\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{T}_0, \boldsymbol{T}_1, \boldsymbol{U}$; receive $\boldsymbol{a}_0, \boldsymbol{a}_1, \boldsymbol{a}_2$.
- Player $\bar{\boldsymbol{b}}$: give $(\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{U}$; receive $\boldsymbol{a}'_2$

Accept if $\boldsymbol{a}_2 = \boldsymbol{a}'_2$.

Code checks :

Answer code check: Sample questions $(\boldsymbol{w}_0, \boldsymbol{w}_1) \sim G_{\ell_1}(\boldsymbol{T}_{\boldsymbol{c}})$. Distributed the questions as follows:

- Player $\boldsymbol{b}$: give $\boldsymbol{x}_{\boldsymbol{c}}, \boldsymbol{w}_0$; receive $\boldsymbol{a}_0$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}_{\boldsymbol{c}}, \boldsymbol{w}_1$; receive $\boldsymbol{a}_1$.

Accept if $G_{\ell_1}(\boldsymbol{T}_{\boldsymbol{c}})$ accepts on $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$.

Proof code check: Sample questions $(\boldsymbol{w}_0, \boldsymbol{w}_1) \sim G_{\ell_2}(\boldsymbol{U})$. Distribute the questions as follows:

- Player $\boldsymbol{b}$: give $(\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{w}_0$; receive $\boldsymbol{a}_0$.
- Player $\bar{\boldsymbol{b}}$: give $(\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{w}_1$; receive $\boldsymbol{a}_1$.

Accept if $G_{\ell_2}(\boldsymbol{U})$ accepts on $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$.

Proof of Theorem 35.

Completeness. This follows the same proof of the completeness part of [39, Theorem 17.10].

Soundness. The constant K_1 depends on the parameter $\kappa = |\boldsymbol{I}'_0|$, so we should set κ to be a sufficiently large constant so that $1 - K_1 - K_2\varepsilon^{1/8}$ is greater than the soundness of V. Operationally, the views are augmented by κ uniformly randomly chosen coordinates. The purpose of this is to drive the distance of the Hadamard code up from $1/2$ to $1 - 1/2^{\kappa}$, which will be needed for Lemma 75.

Suppose input is not in L. Let $(|\psi\rangle, M)$ be a strategy that passes with probability $1 - \varepsilon$. This strategy can pass each **Answer code check** with probability $1 - 10\varepsilon$. Given values c and x_c, write $1 - \varepsilon_{c,x_c}$ for the probability the code check passes conditioned on these values. Then with probability at least $1 - 10\varepsilon^{1/2}$, $\varepsilon_{c,x_c} \leq \varepsilon^{1/2}$. When this occurs, we can apply Proposition 31 to $G_{\ell_1}(\boldsymbol{T}_c)$ where the distribution of $\boldsymbol{T}_c$ is determined by c and x_c. Proposition 31 implies that there exists Hilbert spaces $\mathcal{H}_{x_c}$, $|aux_{x_c}\rangle \in \mathcal{H}_{x_c} \otimes \mathcal{H}_{x_c}$ and projective measurement $\{G_u^{x_c}\}$ on $\mathcal{H}^{x_c}$ such that

$$\mathop{\mathbb{E}}_{T_c \sim D_{x_c}} \sum_{a \in \mathbb{F}_2^k} \left\| (M_a^{x_c, T_c} \otimes \mathbb{1}_{\mathcal{H}_{A,x_c}} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes G^{x_c}_{[w|_{T_c}=a]}) |\psi\rangle \otimes |aux_{x_c}\rangle \right\|^2 \leq O(\sqrt{\varepsilon_{c,x_c}})$$

where we use the fact that k is a constant and $\mathbb{1}_A = \mathbb{1}_{\mathcal{H}_A \otimes \mathcal{H}_{A,x_c}}$ and similar for $\mathbb{1}_B$. When this does not occur, we can still assume such Hilbert spaces and projective measurements so that

$$\mathop{\mathbb{E}}_{T_c\sim D_{x_c}}\sum_{a\in\mathbb{F}_2^k}\|(M_a^{x_c,T_c}\otimes\mathbb{1}_{\mathcal{H}_{A,x_c}}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes G^{x_c}_{[w|_{T_c}=a]})\left|\psi\right\rangle\otimes\left|aux_{x_c}\right\rangle\|^2\leq O(1).$$

When averaging over c and x_c,

$$\mathop{\mathbb{E}}_{c,x_c}\mathop{\mathbb{E}}_{T_c\sim D_{x_c}}\sum_{a\in\mathbb{F}_2^k}\|(M_a^{x_c,T_c}\otimes\mathbb{1}_{A,x_c}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes G^{x_c}_{[w|_{T_c}=a]})\left|\psi\right\rangle\otimes\left|aux_{x_c}\right\rangle\|^2\leq O(\varepsilon^{1/4}).$$

Passing the **Proof code check** implies that there exists Hilbert spaces $\mathcal{H}_{x_0,x_1}$, states $\left|aux_{x_0,x_1}\right\rangle\in\mathcal{H}_{x_0,x_1}\otimes\mathcal{H}_{x_0,x_1}$ and projective measurements $\{H_w^{x_0,x_1}\}$ on $\mathcal{H}\otimes\mathcal{H}_{x_0,x_1}$ such that

$$\mathop{\mathbb{E}}_{x_0,x_1}\mathop{\mathbb{E}}_{U\sim D_{(x_0,x_1)}}\sum_{a\in\mathbb{F}_2^k}\|(M_a^{x_0,x_1,U}\otimes\mathbb{1}_{\mathcal{H}_{A,x_0,x_1}}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes H^{x_0,x_1}_{[w|_U=a]})\left|\psi\right\rangle\otimes\left|aux_{x_0,x_1}\right\rangle\|^2\leq O(\varepsilon^{1/4}).$$

The next step is ensuring the G and H measurements act on the same Hilbert space. Let

$$\left|\tilde{\psi}\right\rangle=\left|\psi\right\rangle\otimes(\otimes_x\left|aux_x\right\rangle)\otimes(\otimes_{x_0,x_1}\left|aux_{x_0,x_1}\right\rangle)$$

and

$$\begin{aligned}\tilde{G}_u^{x_c}&=G_u^{x_c}\otimes(\otimes_{x\neq x_c}\mathbb{1}_{\mathcal{H}_x})\otimes(\otimes_{x_0,x_1}\mathbb{1}_{\mathcal{H}_{x_0,x_1}})\\ \tilde{H}_u^{x_0,x_1}&=H^{x_0,x_1}\otimes(\otimes_x\mathbb{1}_{\mathcal{H}_x})\otimes(\otimes_{(z_0,z_1)\neq(x_0,x_1)}\mathbb{1}_{\mathcal{H}_{z_0,z_1}}),\end{aligned}$$

and, let

$$\begin{aligned}N_{a_c}^{x_c,T_c}&=M_{a_c}^{x_c,T_c}\otimes(\otimes_x\mathbb{1}_{\mathcal{H}_x})\otimes(\otimes_{x_0,x_1}\mathbb{1}_{\mathcal{H}_{x_0,x_1}})\\ N^{x_0,x_1,U}&=M_{a_2}^{x_0,x_1,U}\otimes(\otimes_x\mathbb{1}_{\mathcal{H}_x})\otimes(\otimes_{x_0,x_1}\mathbb{1}_{\mathcal{H}_{x_0,x_1}})\\ N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U}&=M_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U}\otimes(\otimes_x\mathbb{1}_{\mathcal{H}_x})\otimes(\otimes_{x_0,x_1}\mathbb{1}_{\mathcal{H}_{x_0,x_1}}).\end{aligned}$$

Note that we omit the permutation of the Hilbert spaces in the definitions above. Then for all x_c

$$\begin{aligned}&\mathop{\mathbb{E}}_{T_c\sim D_{x_c}}\sum_{a\in\mathbb{F}_2^k}\|(N_a^{x_c,T_c}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes\tilde{G}^{x_c}_{[w|_{T_c}=a]})\left|\tilde{\psi}\right\rangle\|^2\\ &=\mathop{\mathbb{E}}_{T_c\sim D_{x_c}}\sum_{a\in\mathbb{F}_2^k}\|(M_a^{x_c,T_c}\otimes\mathbb{1}_{\mathcal{H}_{A,x_c}}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes G^{x_c}_{[w|_{T_c}=a]})\left|\psi\right\rangle\otimes\left|aux_{x_c}\right\rangle\|^2.\end{aligned}$$

Thus

$$\mathop{\mathbb{E}}_{c,x_c}\mathop{\mathbb{E}}_{T_c\sim D_{x_c}}\sum_{a\in\mathbb{F}_2^k}\|(N_a^{x_c,T_c}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes\tilde{G}^{x_c}_{[w|_{T_c}=a]})\left|\tilde{\psi}\right\rangle\|^2\leq O(\varepsilon^{1/4}),\tag{29}$$

and

$$\mathop{\mathbb{E}}_{x_0,x_1}\mathop{\mathbb{E}}_{U\sim D_{(x_0,x_1)}}\sum_{a\in\mathbb{F}_2^k}\|(N_a^{x_0,x_1,U}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes\tilde{H}^{x_0,x_1}_{[w|_U=a]})\left|\tilde{\psi}\right\rangle\|^2\leq O(\varepsilon^{1/4}).\tag{30}$$

Note these relations also hold with the two systems flipped.

Passing the **Cross Checks** implies that

$$N_{a_0}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_0}^{x_0,T_0} \tag{31}$$
$$N_{a_1}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_1}^{x_1,T_1} \tag{32}$$
$$N_{a_0}^{x_0,T_0} \otimes \mathbb{1} \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_0}^{x_0,T_0} \tag{33}$$
$$N_{a_1}^{x_1,T_1} \otimes \mathbb{1} \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_1}^{x_1,T_1} \tag{34}$$
$$N_{a_2}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_2}^{x_0,x_1,U} \tag{35}$$
$$N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U}, \tag{36}$$

with respect to $\left|\tilde{\psi}\right\rangle$. These equations combined with Equations (29) and (30) imply the measurements $\left\{N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U}\right\}$, $\left\{\tilde{G}_u^x\right\}$ and $\left\{\tilde{H}_w^{x_0,x_1}\right\}$ satisfy conditions of Lemma 75 with respect to $\left|\tilde{\psi}\right\rangle$. Let

$$\left\{\Lambda_{u_0,u_1,w}^{x_0,x_1} := \tilde{G}_{u_0}^{x_0} \cdot \tilde{G}_{u_1}^{x_1} \cdot \tilde{H}_w^{x_0,x_1} \cdot \tilde{G}_{u_1}^{x_1} \cdot \tilde{G}_{u_0}^{x_0}\right\}$$

be a POVM constructed following Lemma 75. Recall that T_c and U has κ independent coordinates, so two different codewords agree on T_c or U with a probability at most $\eta_H^\kappa = 1/2^\kappa$. Hence we can applying Lemma 75 to this POVM with $k = 3$, $\delta := \varepsilon^{1/4}$ and $\varepsilon := 1/2^\kappa$, and get that

$$N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon^{1/48}+1/2^{\kappa/6})} \mathbb{1}_A \otimes \Lambda_{[u_0|_{T_0},u_1|_{T_1},w|_U=a_0,a_1,a_2]}^{x_0,x_1} \tag{37}$$

with respect to $\left|\tilde{\psi}\right\rangle$, where $[u_0|_{T_0}, u_1|_{T_1}, w|_U = a_0, a_1, a_2]$ means that $\mathsf{Enc}_{\ell_1}(u_0)|_{T_0} = a_0$ and etc.. Passing **Verify** with a probability at least $1 - 10\varepsilon$ along with Equation (37) and Lemma 70 implies that $\left\{\Lambda_{u_0,u_1,w}^{x_0,x_1}\right\}$ can be used to pass the verify test with probability $1 - 10\varepsilon - O(\varepsilon^{1/96} + (1/2)^{\kappa/6})$. The player would measure $\mathbb{1}_A \otimes \Lambda$ on $\left|\tilde{\psi}\right\rangle$ and return the local views of the measurement outcomes according to the questions.

Consider the measurements $\left\{\Lambda_{u_0,u_1}^{x_0,x_1} := \sum_w \Lambda_{u_0,u_1,w}^{x_0,x_1}\right\}$ Let

$$p := \mathop{\mathbb{E}}_{x_0,x_1} \sum_{u_0,u_1:V(\mathsf{input},x_0,x_1,u_0,u_1)=1} \left\langle\tilde{\psi}\right| \mathbb{1}_A \otimes \Lambda_{u_0,u_1}^{x_0,x_1} \left|\tilde{\psi}\right\rangle,$$

which is the probability that measuring with $\Lambda_{u_0,u_1}^{x_0,x_1}$ gives answers u_0 and u_1 accepted by the verifier V when the questions are x_0 and x_1. Then

$$\begin{aligned}
p &= \mathop{\mathbb{E}}_{x_0,x_1} \sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=1}} \sum_w \left\langle\tilde{\psi}\right| \mathbb{1}_A \otimes \Lambda_{u_0,u_1,w}^{x_0,x_1} \left|\tilde{\psi}\right\rangle \\
&\geq \mathop{\mathbb{E}}_{x_0,x_1} \sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=1}} \sum_w \left\langle\tilde{\psi}\right| \mathbb{1}_A \otimes \Lambda_{u_0,u_1,w}^{x_0,x_1} \left|\tilde{\psi}\right\rangle \cdot \Pr_R[V_{\mathrm{PCPP}}^{u_0,u_1,w}(\mathsf{input}, x_0, x_1, 2 \cdot 2^{\ell_1}; R) = 1] \\
&= \Pr[(\left|\tilde{\psi}\right\rangle, \Lambda) \text{ pass } \textbf{verify} \text{ check }] \\
&\quad - \sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=0}} \sum_w \left\langle\tilde{\psi}\right| \mathbb{1}_A \otimes \Lambda_{u_0,u_1,w}^{x_0,x_1} \left|\tilde{\psi}\right\rangle \cdot \Pr_R[V_{\mathrm{PCPP}}^{u_0,u_1,w}(\mathsf{input}, x_0, x_1, 2 \cdot 2^{\ell_1}; R) = 1] \\
&\geq 1 - 10\varepsilon - O(\varepsilon^{1/96} + (1/2)^{\kappa/6}) \\
&\quad - \sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=0}} \sum_w \left\langle\tilde{\psi}\right| \mathbb{1}_A \otimes \Lambda_{u_0,u_1,w}^{x_0,x_1} \left|\tilde{\psi}\right\rangle \cdot \Pr_R[V_{\mathrm{PCPP}}^{u_0,u_1,w}(\mathsf{input}, x_0, x_1, 2 \cdot 2^{\ell_1}; R) = 1] \\
&\geq 1 - 10\varepsilon - O(\varepsilon^{1/8} + 1/2^\kappa) - (1-p)s,
\end{aligned}$$

where s is the soundness of V_{PCPP}. In the derivation above, $\Pr_R[V_{\mathrm{PCPP}}^{u_0,u_1,w}(\mathsf{input}, x_0, x_1, 2 \cdot 2^{\ell_1}; R) = 1]$ is the probability that V_{PCPP} accepts input. For any x_0, x_1, u_0, u_1 not accepted by V, this probability is below s by [39, Proposition 17.8]. Hence

$$p \geq \frac{1 - 10\varepsilon - O(\varepsilon^{1/96} + (1/2)^{\kappa/6}) - s}{1 - s} = 1 - \frac{10\varepsilon + O(\varepsilon^{1/96} + (1/2)^{\kappa/6})}{1 - s}.$$

In the end, we use $(\{\tilde{G}_u^x\}, |\tilde{\psi}\rangle)$ as a strategy for V. Applying Lemma 73 to Equations (29), (33), and (34), we get that

$$\tilde{G}^{x_0}_{u|_{T_0}=a} \otimes \mathbb{1} \approx_{O(\varepsilon^{1/4})} \mathbb{1} \otimes \tilde{G}^{x_0}_{u|_{T_0}=a}$$

with respect to the distribution of x_0 and the distribution of T_0 determined by x_0 on the state $|\tilde{\psi}\rangle$. Since $\{\tilde{G}_u^{x_0}\}$ is a projective measurement, we know

$$\mathop{\mathbb{E}}_{x_0} \mathop{\mathbb{E}}_{T_0 \sim D_{x_0}} \sum_a \langle\tilde{\psi}| \tilde{G}^{x_0}_{u|_{T_0}=a} \otimes \tilde{G}^{x_0}_{u|_{T_0}=a} |\tilde{\psi}\rangle \geq 1 - O(\varepsilon^{1/4}).$$

On the other hand

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_0} \mathop{\mathbb{E}}_{T_0 \sim D_{x_0}} \sum_a \langle\tilde{\psi}| \tilde{G}^{x_0}_{u|_{T_0}=a} \otimes \tilde{G}^{x_0}_{u|_{T_0}=a} |\tilde{\psi}\rangle \\
&= \mathop{\mathbb{E}}_{x_0} \sum_u \langle\tilde{\psi}| \tilde{G}^{x_0}_u \otimes \tilde{G}^{x_0}_u |\tilde{\psi}\rangle + \mathop{\mathbb{E}}_{x_0} \mathop{\mathbb{E}}_{T_0 \sim D_{x_0}} \sum_{u \neq u' : u|_{T_0} = u'|_{T_0}} \langle\tilde{\psi}| \tilde{G}^{x_0}_u \otimes \tilde{G}^{x_0}_{u'} |\tilde{\psi}\rangle \\
&= \mathop{\mathbb{E}}_{x_0} \sum_u \langle\tilde{\psi}| \tilde{G}^{x_0}_u \otimes \tilde{G}^{x_0}_u |\tilde{\psi}\rangle + \mathop{\mathbb{E}}_{x_0} \mathop{\mathbb{E}}_{T_0 \sim D_{x_0}} \sum_{u \neq u'} \mathbb{1}[u|_{T_0} = u'|_{T_0}] \langle\tilde{\psi}| \tilde{G}^{x_0}_u \otimes \tilde{G}^{x_0}_{u'} |\tilde{\psi}\rangle .
\end{aligned}$$

Since for all x_0 and $u \neq u'$, $\mathbb{E}_{T_0 \sim D_{x_0}} \mathbb{1}[u|_{T_0} = u'|_{T_0}] \leq 1/2^\kappa$, we know

$$\mathop{\mathbb{E}}_{x_0} \sum_u \langle\tilde{\psi}| \tilde{G}^{x_0}_u \otimes \tilde{G}^{x_0}_u |\tilde{\psi}\rangle \geq 1 - 1/2^\kappa - O(\varepsilon^{1/4}).$$

Again, because $\{\tilde{G}_u^{x_0}\}$ is a projective measurement

$$\mathop{\mathbb{E}}_{x_0} \sum_u \|(\tilde{G}^{x_0}_u \otimes \mathbb{1} - \mathbb{1} \otimes \tilde{G}^{x_0}_u) |\tilde{\psi}\rangle\|^2 \leq \frac{1}{2^{\kappa-1}} + O(\varepsilon^{1/4}).$$

Let $S(x_0, x_1) = \{(a_0, a_1) \mid V(x_0, x_1, a_0, a_1) = 1\}$. We can calculate

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{x_0,x_1} \sum_{(a_0,a_1) \in S} \langle\tilde{\psi}| \tilde{G}^{x_0}_{a_0} \otimes \tilde{G}^{x_1}_{a_1} |\tilde{\psi}\rangle - \langle\tilde{\psi}| \tilde{G}^{x_0}_{a_0} \otimes \tilde{G}^{x_1}_{a_1} \tilde{G}^{x_0}_{a_0} |\tilde{\psi}\rangle| \\
&\leq \sqrt{\mathop{\mathbb{E}}_{x_0,x_1} \sum_{(a_0,a_1) \in S} \|\tilde{G}^{x_0}_{a_0} \otimes \tilde{G}^{x_1}_{a_1} |\tilde{\psi}\rangle\|^2} \\
&\cdot \sqrt{\mathop{\mathbb{E}}_{x_0,x_1} \sum_{(a_0,a_1) \in S} \langle\tilde{\psi}| (\tilde{G}^{x_0}_{a_0} \otimes \mathbb{1} - \mathbb{1} \otimes \tilde{G}^{x_0}_{a_0})(\mathbb{1} \otimes \tilde{G}^{x_1}_{a_1})(\tilde{G}^{x_0}_{a_0} \otimes \mathbb{1} - \mathbb{1} \otimes \tilde{G}^{x_0}_{a_0}) |\tilde{\psi}\rangle} \\
&\leq 1 \cdot \sqrt{\mathop{\mathbb{E}}_{x_0} \sum_{a_0} \|(\tilde{G}^{x_0}_{a_0} \otimes \mathbb{1} - \mathbb{1} \otimes \tilde{G}^{x_0}_{a_0}) |\tilde{\psi}\rangle\|^2} \\
&\leq O(\frac{1}{2^{\kappa/2}} + \varepsilon^{1/8}),
\end{aligned}$$

and

$$
\begin{aligned}
&|\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\langle\tilde{\psi}|\,\mathbb{1}\otimes\tilde{G}_{a_0}^{x_0}\tilde{G}_{a_1}^{x_1}\tilde{G}_{a_0}^{x_0}\,|\tilde{\psi}\rangle-\langle\tilde{\psi}|\,\tilde{G}_{a_0}^{x_0}\otimes\tilde{G}_{a_1}^{x_1}\tilde{G}_{a_0}^{x_0}\,|\tilde{\psi}\rangle|\\
&\leq\sqrt{\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\|\mathbb{1}\otimes\tilde{G}_{a_1}^{x_1}\tilde{G}_{a_0}^{x_0}\,|\tilde{\psi}\rangle\|^2}\\
&\cdot\sqrt{\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\langle\tilde{\psi}|\,(\tilde{G}_{a_0}^{x_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}_{a_0}^{x_0})(\mathbb{1}\otimes\tilde{G}_{a_1}^{x_1})(\tilde{G}_{a_0}^{x_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}_{a_0}^{x_0})\,|\tilde{\psi}\rangle}\\
&\leq 1\cdot\sqrt{\mathop{\mathbb{E}}_{x_0}\sum_{a_0}\|(\tilde{G}_{a_0}^{x_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}_{a_0}^{x_0})\,|\tilde{\psi}\rangle\|^2}\\
&\leq O(\frac{1}{2^{\kappa/2}}+\varepsilon^{1/8}).
\end{aligned}
$$

Note that $\tilde{G}_{a_0}^{x_0}\tilde{G}_{a_1}^{x_1}\tilde{G}_{a_0}^{x_0}=\Lambda_{a_0,a_1}^{x_0,x_1}$. Therefore,

$$
|\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\langle\tilde{\psi}|\,(\tilde{G}_{a_0}^{x_0}\otimes\tilde{G}_{a_1}^{x_1}-\mathbb{1}\otimes\Lambda_{a_0,a_1}^{x_0,x_1})\,|\tilde{\psi}\rangle|\leq O(\frac{1}{2^{\kappa/2}}+\varepsilon^{1/8}).
$$

On the other hand, we have shown

$$
\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\langle\tilde{\psi}|\,\mathbb{1}\otimes\Lambda_{a_0,a_1}^{x_0,x_1})\,|\tilde{\psi}\rangle=p\geq 1-O(\varepsilon^{1/96}+(1/2)^{\kappa/6}).
$$

Hence, the winning probability of the strategy $(\{\tilde{G}_u^x\},|\tilde{\psi}\rangle)$ is at least $1-\frac{C_1}{2^{\kappa/6}}-C_2\varepsilon^{1/96}$ for some constants C_1 and C_2. Hence, $K_1=\frac{C_1}{2^{\kappa/6}}$ and $K_2=C_2$ in the soundness statement. ◀

Proof of Theorem 36. We first oracularize the MIP* protocol for the Halting problem from [28]. Denote the oracularized verifier by V. For inputs of size n, the verifier's running time for sampling questions and checking answers is $O(\text{poly}(n))$. The sizes of the questions and answers are also $O(\text{poly}(n))$. The oracularized protocol maintains completeness 1 and a constant soundness.

Define the language L_{Enc} as in Definition 33 for V. Then $L_{\mathsf{Enc}}\in\text{DTIME}(2^{\text{poly}(n)})$ because the most costly step of the decider of L_{Enc} is running $\mathsf{Dec}_{\text{poly}(n)}$ which takes $O(2^{\text{poly}(n)})$ time. By Definition 32, the PCPP verifier V_{PCPP} for L_{Enc} has randomness complexity $O(\text{poly}(n))$, query complexity $O(1)$, and verification time $O(\text{poly}(n))$.

Next, we apply the answer reduction technique of this section to V to get verifier V^{AR}. The sampling time of V^{AR} is the sum of the sampling time of V, the sampling time of V_{PCPP}, and the sampling time of the additional constantly many independent coordinates, so it is $O(\text{poly}(n))$. Since the question sizes of V and V_{PCPP} are both $O(\text{poly}(n))$, the question size of V^{AR} is also $O(\text{poly}(n))$. The answers expected by V^{AR} are constantly many bits, so the answer size is $O(1)$. Lastly, the verification time of V^{AR} is determined by the verification time of V_{PCPP}, so it is also $O(\text{poly}(n))$. The completeness and soundness of V^{AR} follow from Theorem 35. Then the theorem statement follows from the Halting problem is RE-complete. ◀