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

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

YANGJING DONG, Computer Science, Nanjing University, Nanjing, China
HONGHAO FU, CSAIL, Massachusetts Institute of Technology, Cambridge, United States and Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada
ANAND NATARAJAN, CSAIL, Massachusetts Institute of Technology, Cambridge, United States
MINGLONG QIN, Computer Science, Nanjing University, Nanjing, China
HAOCHEN XU, Institute of Software, Chinese Academy of Sciences, Beijing, China
PENGHUI YAO, Computer Science, Nanjing University, Nanjing, China

The class MIP* of quantum multiprover interactive proof systems with entanglement is much more powerful than its classical counterpart MIP [8, 31, 32]: 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 [53, 54] 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* [poly, $O(1)$], where the verifier sends polynomially many bits to the provers and the provers send back constantly many bits. We show that 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) [53]. We also show that this collapse in power is due to 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* [poly, 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 that 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.

CCS Concepts: • **Theory of computation** → **Interactive proof systems**; **Quantum complexity theory**; **Complexity classes**; Pseudorandomness and derandomization.

Additional Key Words and Phrases: Boolean analysis

*A preliminary version has been accepted by the 39th Computational Complexity Conference (CCC 2024).

Authors' Contact Information: Yangjing Dong, Computer Science, Nanjing University, Nanjing, Jiangsu, China; e-mail: dongmassimo@gmail.com; Honghao Fu, CSAIL, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States and Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Quebec, Canada; e-mail: honghaof@mit.edu; Anand Natarajan, CSAIL, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States; e-mail: anandn@mit.edu; Minglong Qin, Computer Science, Nanjing University, Nanjing, Jiangsu, China; e-mail: mlqin@smail.nju.edu.cn; Haochen Xu, Institute of Software, Chinese Academy of Sciences, Beijing, Beijing, China; e-mail: xuhc@ios.ac.cn; Penghui Yao, Computer Science, Nanjing University, Nanjing, Jiangsu, China; e-mail: phyao1985@gmail.com.

ACM 1557-735X/2025/8-ART
https://doi.org/10.1145/3760771

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 [28, 30, 35, 36, 61, 62] leading to the seminal result MIP* = RE [31, 32] 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 provers 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 for illustration, while 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 [14]. In this paper, we are able to handle a more general noise model, see Section 2.1.) 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 [3, 5], 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 [53]. They investigated two-player nonlocal games[1] when the states shared between the players are arbitrarily many copies of a maximally entangled state

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

(MES) with an arbitrarily small but *constant* amount of noise on each copy, which is termed as *noisy MES* in their paper. The noise will cause the quantum maximal correlation, as defined in Definition 2.1, to be less than 1, and the marginal state to be a completely mixed state. For instance, applying a depolarizing channel to an MES results in a noisy MES. 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 [54]. 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. [31, 32, 56], 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 [53, 54]. Moreover, [53, 54] 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 [53], which is much more powerful than MIP $[\text{poly}, \text{poly}] = \text{NEXP}$. 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 $[\text{poly}, \text{poly}] = \text{MIP}\,[\text{poly}, O(1)] = \text{NEXP}$ [8, 42][2]. Our main result, stated in the language of nonlocal games, is the following.

Theorem 1.1 (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 1.1 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 1.1 shows that noisy 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 1.2. $\text{MIP}^*[\text{poly}, O(1), \psi] = \text{NEXP}$*, where ψ is a noisy MES.*

Intuitively, Theorem 1.1 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 [26] 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 [19, 36], while the classical XOR games are NP-hard, and the classical unique games are conjectured to be NP-hard as well [37]. Thus introducing noisy quantum states doesn't introduce any quantum effect to the hardness at all.

[2] MIP $[\text{poly}, \text{poly}] = \text{NEXP}$ was proved in [8]. MIP $[\text{poly}, O(1)] = \text{NEXP}$ can be proved using a scaled-up version of PCP theorem [42].

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 1.3 (Theorem 6.11)**.** RE *is equal to* MIP*$[\mathrm{poly}, O(1), |\mathrm{EPR}\rangle]$ *with completeness* 1 *and constant soundness.*

To put this in context, the original work [31, 32] 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. Recently, Natarajan and Zhang [48] proved, by repeatedly applying the "question reduction" technique from [31], that it is still RE-complete if the question length is $O(1)$ and the answer length is $\mathrm{polylog}(n)$. Here, we achieve constant *answer* length by combining a tightened version of the previous answer reduction technique with a new answer reduction transformation, obtained by instantiating the error-correcting code-based scheme of [47] with the Hadamard code. We also show how to alternately achieve constant answer length by iterative application of the tightened standard answer reduction, similarly to how [48] obtained constant question length.

Theorems 1.1 and 1.3 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.[3] Alternately, using the power of 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 MIP* = RE, our result puts some limitations on the design of self-tests robust against noise. More specifically, our result 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 [22].

1.1 Proof Overview

The harder part is to show that there is an NP-algorithm for this problem. To illustrate our algorithm, we adapt the framework of Fourier analysis on matrix spaces. This framework was initiated in [43, 65] and views the set of n-qubit operators as a Hilbert space obtained by tensoring n copies of 2-dimensional Hilbert spaces. Furthermore, we extend the results in the analysis of Boolean functions [50] to such a space. Readers may refer to [53] for a thorough treatment.

1.1.1 Approximating the Values of Noisy Games is NP*-Complete.* Given a nonlocal game sharing arbitrary copies of a noisy MES ψ, Qin and Yao [53] 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 [53]. 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 Fourier expansion of P be

[3]Note that quantum states from which MES can be obtained through local operations without any communication are considered equivalent to MES in this model. This is because two non-communicating provers can transform any such state to an MES.

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

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 [53], 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

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

so we denote the new strategy by $S^{(\text{Smooth})}$. Using the argument in [53], the probability of winning the game with this new strategy changes at most slightly, i.e.

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

Let τ be a small constant independent of n. Since the degree of $P^{(\text{Smooth})}$ 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 [53] to replace all the non-identity Pauli bases in the registers with low influence by Gaussian variables while maintaining the strategy value. Let

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

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 $S^{(\text{Apprx})}$, then

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

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 [53], 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 Gaussian 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 be positive semidefinite and thus do 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 $\mathrm{Tr}\left(\mathcal{P}_i \otimes \mathcal{P}_j\right)\psi = \delta_{i,j} c_i$, where $c_0 = 1$ and $c_1 = c_2 = c_3 = \rho$. Thus for any degree-d operators A, B, we have

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

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

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

Then $\mathrm{Tr}\,\zeta(P)$ 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 [53]. 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 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 [40] to obtain a derandomized invariance principle, which only uses a logarithmic number of independent bits to simulate these variables[4]. 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 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 MIP$^*[\mathrm{poly}, O(1)]$. To show hardness of MIP$^*[\mathrm{poly}, O(1)]$, we start from the known result MIP$^*[\mathrm{poly}, \mathrm{poly}]$ = RE [31], and apply *answer reduction* transformations 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 [47] and [31]. Intuitively, the idea of answer reduction is to ask the two provers in

[4]An alternate approach is using Gaussian variables and derandomizing the Gaussian variables as in [33], which discretizes the Gaussian variables via the Box-Muller transformation and further derandomizes the discrete random variables.

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 (PCPP) 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 [31], the code that was used was the Reed-Muller code, which has superconstant alphabet size. Moreover, the formulation in [31] was for the setting of reducing the answer length from exponential to polynomial, and in fact the specific theorem shown there is incapable of reducing the answer length below polylog(n). Our first contribution is to improve the parameters of this answer reduction transformation to make sure that in each application it can reduce answer size exponentially and can be recursively applied to reduce answer size below $\log(n)$.

To go all the way down to $O(1)$-sized answers, we combine this Reed-Muller-based answer reduction with a new answer reduction theorem based on 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" [29, 46]. Moreover, the answer-reduction protocol in [47] 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 PCPP 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 $O(\log(n))$ to $O(1)$ directly.

1.2 Technical Contributions

1.2.1 Invariance Principle and Derandomized Invariance Principle for Matrix Functions. The invariance principle [44] 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 on 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[5]. 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 [44] is via Lindeberg's hybrid argument, which is also a classic method to prove the Central Limit Theorem.

In [53], 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

[5]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.

are complicated and difficult to analyze in general [58]. Qin and Yao [53] 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 [60], 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 [25, 40, 52] 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 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 [40] 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 [9, 24, 39, 49]. 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 [9, 24, 49], 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 [47], 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 [13, 46]. Secondly, the answer reduction procedure in [47] 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 [47] by querying a large constant number of "dummy coordinates" from the provers. It is worth mentioning that the answer reduction procedure from [47] is different from the procedure used in [31]; 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 addition to this new answer reduction based on the Hadamard code, we also required a tightened version of the Reed-Muller-based answer reduction of [31], as noted above. This is because, due to the low rate of the Hadamard code, we must first reduce the answer length to $O(\log n)$ before applying our new answer reduction. However, the answer reduction as stated in [31] can never reduce the answer size to smaller than $\text{polylog}(n)$, because the reduced answer size depends poly-logarithmically on the verification time, which can never be smaller than $\text{poly}(n)$ since the verifier must read the entire input. Our improvement is based on the observation that the verifier's verification process can be broken into two phases. In the first phase, a predicate of the answers is calculated, and in the second phase, the predicate is applied to the answers. We observe that the new answer size only depends poly-logarithmically on the size of the Boolean circuit implementing the predicate, which can be much smaller than the total runtime of the verifier when the answers are short. This observation is standard in the classical PCP literature, but was not necessary for [31] since they were not concerned with obtaining sub-polynomial answer length.

Using this observation, we show that in each application of the answer reduction transformation, both the answer size and the predicate size are reduced exponentially, which allows us to apply it recursively to reduce answer size to below $O(\log n)$, at which point the Hadamard-based answer reduction can take us to constant answer size. We remark that it is also possible to achieve constant answer size by iteratively applying the improved answer reduction. The analysis of this is slightly less clean, but we sketch it at the end of the proof of Theorem 6.11.

1.3 Discussions and Open Problems

Our result characterizes the effect of depolarizing 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 [21]. 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.

Our proof techniques can be applied to the depolarizing noise but not the bit-flipping noise, phase-flipping noise, or phase-damping noise. This is because those types of noise do not reduce the quantum maximal correlation.

Similarly, our techniques cannot be applied to the amplitude-damping noise because under this noise the marginal state is not completely mixed. Hence, the effect of these noise channels on MIP* is not clear. On the other hand, if Alice and Bob start with tilted EPR pairs, for example, caused by some unitary noise, they can produce maximally entangled states via local operations, which is called entanglement concentration in literature [12]. Then they can execute the MIP* protocol for RE.

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 [15]. 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 [55]. 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 [17] showed that no quantum algorithm could achieve any speed-up in the unstructured search problem either. For a more detailed survey of the effect of noise on quantum query algorithms, we refer to [17, Section 3].

In recent years, the study of noise has focused on its effect on quantum circuits. In the circuit model, the study is about how noise accumulates in quantum circuits where each gate is subject to some noise. Now we know that noise effectively truncates a quantum circuit to a logarithmic depth [41]. In our case, only the entangled states are subject to noise, and there is no accumulation of noise in the measurements. Our results show that the noise still limits the effective width of the circuit, but do not say anything about the effective depth, which means that in our setting the prover could perform quantum circuits with arbitrary depths.

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. The current answer reduction techniques do not work when the provers can only share noisy MES. Hence, we ask: Does the vanishing phenomenon for computational advantages occur for general MIP* protocols?
(2) 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.
(3) Classical invariance principle serves as a pivotal tool in the analysis of Boolean functions, which has found applications in designing various areas including pseudorandom generators and counting algorithms [7, 25, 34, 51, 52]. Analysis on matrix spaces and the space of super-operators, a.k.a, Pauli analysis [45] is receiving increasing attention[6, 10, 18, 38, 45, 57, 63]. Will our invariance principle lead to new applications?
(4) 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 [9] and references therein). Multiple studies have been devoted to designing efficient algorithms for positivity testing [9, 24, 49]. Will our algorithm of positivity testing find new applications?

Acknowledgment

We thank the anonymous reviewers for their helpful comments. 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. Y.D., M.Q. and P.Y. were supported by National Natural Science Foundation of China (Grant No. 62332009, 12347104), Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0302901), NSFC/RGC Joint Research Scheme (Grant no. 12461160276) and Natural Science Foundation of Jiangsu Province (No. BK20243060). H.F. was supported by

the US National Science Foundation QLCI program (grant OMA-2016245). H.X. was supported by Beijing Nova Program (Grant No. 20220484128 and 20240484652).

2 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.

2.1 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.

The notion of *quantum maximal correlations* introduced by Beigi [11] is crucial to our analysis.

Definition 2.1 (Quantum maximal correlation). [11] 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 quantum 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 2.2. [11] 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 2.3. 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 quantum 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 2.4. [53, 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 2.5. Fact 2.4 indicates the quantum 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 2.6. [53, 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

$$\mathrm{Tr}\left(\left(\mathcal{A}_i \otimes \mathcal{B}_j\right)\psi_{AB}\right) = \begin{cases} c_i & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $c_0 = 1 \geq c_1 = \rho\left(\psi_{AB}\right) \geq c_2 \geq \ldots c_{m^2-1} \geq 0$ and $\rho\left(\psi_{AB}\right)$ is defined in Definition 2.1.

2.2 Matrix analysis

2.2.1 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

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

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

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 $M \in \mathcal{H}_m$, $|||M|||_2^2 = \langle M, M\rangle$.

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 $\mathcal{B}_0 = \mathbb{1}_m$, which is an $m \times m$ identity matrix.

Fact 2.7. [53, 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 2.8. 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:

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

where $\mathbb{1}_m$ is in the i'th quantum system, and the partial trace Tr_i is defined as the operator $\mathbb{1} \otimes \mathrm{Tr}$, with the trace operator Tr acting on the i'th quantum system.

(3) The total influence is defined by

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

Fact 2.9. [53, 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 2.7).

Fact 2.10 (Parseval's identity)**.** For any $P \in \mathcal{H}_m^{\otimes n}$,

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

Definition 2.11. 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 2.12. [53, 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

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

2.2.2 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 $\mathbb{E}\,[XY] = \rho$.

We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is in $L^2\,(\mathbb{R}, \gamma_n)$ if

$$\int_{\mathbb{R}^n} f(x)^2 \gamma_n\,(\mathrm{d}x) < \infty.$$

We equip $L^2\,(\mathbb{R}, \gamma_n)$ with an inner product

$$\langle f, g \rangle_{\gamma_n} = \mathop{\mathbb{E}}_{x \sim \gamma_n}\,[f(x)g(x)]\,.$$

Given $f \in L^2\,(\mathbb{R}, \gamma_n)$, the 2-norm of f is defined to be

$$\|f\|_2 = \sqrt{\langle f, f \rangle_{\gamma_n}}.$$

The set of *Hermite polynomials* forms an orthonormal basis in $L^2(\mathbb{R}, \gamma_1)$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\gamma_1}$. The Hermite polynomials $H_r : \mathbb{R} \to \mathbb{R}$ for $r \in \mathbb{Z}_{\geq 0}$ are defined as

$$H_0(x) = 1; H_1(x) = x; H_r(x) = \frac{(-1)^r}{\sqrt{r!}} e^{x^2/2} \frac{d^r}{dx^r} e^{-x^2/2}.$$

For any $\sigma \in (\sigma_1, \ldots, \sigma_n) \in \mathbb{Z}^n_{\geq 0}$, define $H_\sigma : \mathbb{R}^n \to \mathbb{R}$ as

$$H_\sigma(x) = \prod_{i=1}^{n} H_{\sigma_i}(x_i).$$

The set $\left\{H_\sigma : \sigma \in \mathbb{Z}^n_{\geq 0}\right\}$ forms an orthonormal basis in $L^2(\mathbb{R}, \gamma_n)$. Every function $f \in L^2(\mathbb{R}, \gamma_n)$ has an *Hermite expansion* as

$$f(x) = \sum_{\sigma \in \mathbb{Z}^n_{\geq 0}} \widehat{f}(\sigma) \cdot H_\sigma(x),$$

where $\widehat{f}(\sigma)$'s are the *Hermite coefficients* of f, which can be obtained by $\widehat{f}(\sigma) = \langle H_\sigma, f \rangle_{\gamma_n}$. The degree of f is defined to be

$$\deg(f) = \max\left\{\sum_{i=1}^{n} \sigma_i : \widehat{f}(\sigma) \neq 0\right\}.$$

We say $f \in L^2(\mathbb{R}, \gamma_n)$ is *multilinear* if $\widehat{f}(\sigma) = 0$ for $\sigma \notin \{0, 1\}^n$.

Now we give the definition of random matrix.

Definition 2.13. Given $h, n, m \in \mathbb{Z}_{>0}$, we say $P(\mathbf{g})$ is a random matrix if it can be expressed as

$$P(\mathbf{g}) = \sum_{\sigma \in [m^2]^h_{\geq 0}} p_\sigma(\mathbf{g}) \mathcal{B}_\sigma, \tag{6}$$

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 $\sigma \in [m^2]^h_{\geq 0}$ 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 $p_\sigma \in L^2(\mathbb{R}, \gamma_n)$ for all $\sigma \in [m^2]^h_{\geq 0}$.

We define the degree of random operators:

Definition 2.14. 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

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

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

2.2.3 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 [20] for a detailed treatment.

Definition 2.15. 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 notations, 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 C^n to be the space of functions continuously differentiable n times.

Definition 2.16. Let $\lambda_0, \ldots, \lambda_n \in \mathbb{R}$ and let $f \in C^n$. 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 2.17. [60, Theorem 5.3.2] [58, Theorem 6.1] Given $m, n \in \mathbb{Z}_{>0}$, $P, Q \in \mathcal{H}_m$. Suppose that P has a spectral decomposition

$$P = \sum_{i=1}^{m} \lambda_i \Pi_i, \tag{7}$$

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 $f \in C^n$. Then

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

The following is one of the main results in the theory of multilinear operator integrals [60].

Fact 2.18. [60, Theorem 5.3.12] Given $m, n \in \mathbb{Z}_{>0}$, $P, Q \in \mathcal{H}_m$. Let $f \in C^n$. 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

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

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

2.2.4 The distance from PSD matrices. Define the function $\zeta : \mathbb{R} \to \mathbb{R}$ as follows.

$$\zeta(x) = \begin{cases} x^2 & \text{if } x \leq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

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

Fact 2.19. [53, Lemma 9.1] Given an integer $m > 0$, $M \in \mathcal{H}_m$, $\mathrm{Pos} = \{X \in \mathcal{H}_m : X \geq 0\}$, let

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

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

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

Fact 2.20. [53, Lemma 10.4] For any Hermitian matrices P and Q, it holds that

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

We will need to mollify[6] ζ to get a smooth function:

Fact 2.21. [44, Lemma 3.21] Given $\lambda > 0$, there exists a C^∞ function ζ_λ satisfying

(1) $\|\zeta_\lambda - \zeta\|_\infty \le 2\lambda^2$,
(2) For any integer $n \ge 2$, there exists a constant B_n independent of λ such that

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

2.3 k-wise uniform hash functions and random variables

Definition 2.22. 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 2.23. A random vector $\mathbf{z} \in [p]^n$ is k-wise uniform if for any $y_1, \ldots, y_k \in [p]$ and distinct $x_1, \ldots, x_k \in [n]$:

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

Lemma 2.24. *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., [16]). 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 [59]. 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 [64], 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 2.25. *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 $\mathcal{F}$. □

2.4 Nonlocal games and MIP* protocols

Two-player one-round MIP* protocols are also nonlocal games. We follow the notations of [31] for nonlocal games.

Definition 2.26 (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

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

- $\mathcal{X}$ 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 2.27 (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 $x \in \mathcal{X}$, $A^x = \left\{A^x_a \mid a \in \mathcal{A}\right\}$ is a POVM over $\mathcal{H}_A$, and
- B is a set $\{B^y\}$ such that for every $y \in \mathcal{Y}$, $B^y = \left\{B^y_b \mid b \in \mathcal{B}\right\}$ is a POVM over $\mathcal{H}_B$.

Definition 2.28 (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

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

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 [32, Theorem 5.1].

Definition 2.29. 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 2.30 (Lemma 5.7 in [31])**.** *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.

2.5 Lemmas for the answer reduction of MIP*

This section introduces several lemmas to prove the hardness of MIP*[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

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

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

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

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

Lemma 2.31 (Fact 4.13 of [47]). *Let $\{A_a^x\}$ and $\{B_a^x\}$ be POVM measurements. If $A_a^x \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes B_a^x$, then $A_a^x \otimes \mathbb{1} \approx_{2\delta} \mathbb{1} \otimes B_a^x$.*

Lemma 2.32. *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

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

This proof is deferred to Appendix B.

Lemma 2.33 (Fact 4.14 of [47]). *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 $A_a^x \otimes \mathbb{1} \approx_{\sqrt{\delta}} \mathbb{1} \otimes B_a^x$.*

Lemma 2.34 (Proposition 4.26 of [32]). *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 2.35 (Proposition 4.28 of [32]). *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 $i \in [k]$. Then*

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

Lemma 2.36 (Fact 4.33 of [47]). *Let $k \geq 0$ be a constant. Let $\left\{A_{a_1,\ldots,a_k}^x\right\}$ be a projective measurement. For $1 \leq j \leq k$, let $\left\{(B_j)_{a_j}^x\right\}$ be a projective measurement, and suppose that*

$$A_{a_j}^x \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes (B_j)_{a_j}^x.$$

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

$$J_{a_1,\ldots,a_k}^x = (B_k)_{a_k}^x \ldots (B_2)_{a_2}^x (B_1)_{a_1}^x (B_2)_{a_2}^x \ldots (B_k)_{a_k}^x.$$

Then

$$A_{a_1,\ldots,a_k}^x \otimes \mathbb{1} \approx_{(2k-1)^2\delta} \mathbb{1} \otimes J_{a_1,\ldots,a_k}^x.$$

This proof is also deferred to Appendix B.

Lemma 2.37 (Fact 4.35 of [47]). *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 $\left\{(G_i)_g^x \mid g \in \mathcal{G}_i\right\}$ 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 $\left\{A_{a_1,\ldots,a_k}^{x,y_1,\ldots,y_k}\right\}$ be a projective measurement with outcomes $a_i \in \mathcal{R}_i$. For each $1 \leq i \leq k$, suppose that

$$A_{a_i}^{x,y_1,\ldots,y_k} \otimes \mathbb{1} \approx_\delta \mathbb{1} \otimes (G_i)_{[g_i(y_i)=a_i]}^x \tag{9}$$

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

Also suppose that

$$A_{a_i}^{x,y_1,\ldots,y_k} \otimes \mathbb{1} \simeq_\delta \mathbb{1} \otimes A_{a_i}^{x,y_1,\ldots,y_k}. \tag{11}$$

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

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

Then

$$A_{a_1,\ldots,a_k}^{x,y_1,\ldots,y_k} \otimes \mathbb{1} \approx_{O(\exp(k)(\delta^{1/4^{k-1}}+\varepsilon^{1/(2\cdot 4^{k-2})}))} \mathbb{1} \otimes J_{[g_1(y_1),\ldots,g_k(y_k)=a_1,\ldots,a_k]}^x.$$

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

3 Invariance principle for matrix spaces

This section we will prove an invariance principle for general functions on matrix spaces. Hypercontractivity is crucial in the proofs of many invariance principles [7, 25, 27, 44, 53]. We also need to establish a new hypercontractive inequality before proving the invariance principle.

3.1 Hypercontractivity

In this subsection, we adopt the concept of orthonormal ensembles as introduced in [44].

Definition 3.1. Given $m, n \in \mathbb{Z}_{>0}$, a collection of n real random variables $\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ are orthonormal if $\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_j\right] = \delta_{i,j}$. 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 $i \in [n]$, $\mathbf{x}_i = \left\{\mathbf{x}_{i,0} = 1, \mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,m-1}\right\}$ is an m-orthonormal ensemble.

Definition 3.2. Given $m, n \in \mathbb{Z}_{>0}$, $\tau \in [m]_{\geq 0}^n$ 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

$$Q(\mathbf{x}) = \sum_{\tau \in [m]_{\geq 0}^n} \widehat{Q}(\tau)\, \mathbf{x}_\tau,$$

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

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

Definition 3.3. For $1 \leq r < \infty$, let $\mathbf{y}$ be a random variable with $\mathbb{E}[|\mathbf{y}|^r] < \infty$. Define

$$\|\mathbf{y}\|_r = \left(\mathbb{E}[|\mathbf{y}|^r]\right)^{1/r}.$$

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

$$\left\|\left(T_\eta Q\right)(\mathbf{x})\right\|_q \leq \|Q(\mathbf{x})\|_p.$$

Fact 3.4. [44, Remark 3.10] If $\mathbf{x}$ is (p, q, η)-hypercontractive, then it is (p, q, η')-hypercontractive for any $0 < \eta' \leq \eta$.

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 [44] for an extensive treatment on hypercontractive ensembles.

We need the following lemma for technical reasons.

Lemma 3.5. *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. Let $q_i = T_\eta p_i$. Then

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

□

We then introduce the noise operator Γ_γ for random matrices, which is a hybrid of T_γ in Definition 3.2 and Δ_γ in Definition 2.11.

Definition 3.6. 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 $\sigma \in \left[m^2\right]^h_{\geq 0}$, the noise operator Γ_γ is defined to be

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

The lemma below follows directly from Definition 3.2 and Fact 2.12.

Lemma 3.7. *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* $\sigma \in \left[m^2\right]_{\geq 0}^h$*, suppose that for all* $\sigma \in \left[m^2\right]_{\geq 0}^h$*,* p_σ *has an expansion*

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

It holds that

$$\Gamma_\gamma\left(P(\mathbf{x})\right) = \sum_{\sigma \in [m^2]_{\geq 0}^h} \sum_{\tau \in [m^2]_{\geq 0}^n} \gamma^{|\sigma|+|\tau|}\widehat{p_\sigma}(\tau)\mathbf{x}_\tau \mathcal{B}_\sigma. \tag{12}$$

We need a hypercontractivity inequality for Hermitian matrices.

Fact 3.8. [53, 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

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

where $\Delta_\gamma(\cdot)$ is defined in Definition 2.11.

The main result in this subsection is stated below.

Theorem 3.9 (Hypercontractivity for random matrices)**.** *Given* $h, n, m \in \mathbb{Z}_{>0}$, $m \geq 2$, $0 < \eta < 1$, $0 \leq \gamma \leq \min\left\{\eta, (9m)^{-1/4}\right\}$*, a* $(2, 4, \eta)$*-hypercontractive* (m^2, n) *ensemble* $\mathbf{x}$ *and a random matrix*

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

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

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

where Γ_γ *is defined in Definition 3.6.*

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

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

Using Fact 3.8,

$$\mathbb{E}\left[\left\|\left\|\Delta_\gamma\left(Q(\mathbf{x})\right)\right\|\right\|_4^4\right] \leq \mathbb{E}\left[\left\|\left\|Q(\mathbf{x})\right\|\right\|_2^4\right]. \tag{13}$$

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

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

where the inequality follows from Fact 3.4 and Lemma 3.5. We conclude the result by combining it with Eq. (13). □

The following is an application of Theorem 3.9.

Theorem 3.10. *Given $h, n, m, d \in \mathbb{Z}_{>0}$, $m \geq 2$, $0 < \eta < 1$, a $(2, 4, \eta)$-hypercontractive (m^2, n) ensemble $\mathbf{x}$, and a random matrix*

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

where $\{\mathcal{B}_i\}_{i=0}^{m^2-1}$ is a standard orthonormal basis and for all $\sigma \in [m^2]^h_{\geq 0}$ and p_σ is a real multilinear polynomial satisfying $\deg(p_\sigma) + |\sigma| \leq d$, it holds that

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

Proof. 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 3.7 and Theorem 3.9,

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

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

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

Therefore,

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

□

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 C^4.

Theorem 3.11. *Given $0 < \tau, \eta < 1$, $d, h, m, n \in \mathbb{Z}_{>0}$, $H \subseteq [n]$ of size $|H| = h$, $\xi \in C^3$ satisfying $\|\xi^{(3)}\|_\infty \leq B$ 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 $\operatorname{Inf}_i(P) \leq \tau$ 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*

$$\left|m^{-n}\mathrm{Tr}\,\xi(P) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\,\xi\left(P^H(\mathbf{x})\right)\right]\right| \leq 2c_3 B\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d$$

for some absolute constant c_3.

Proof. 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.

$$\mathcal{X}_\sigma^{(i)} = \mathbf{x}_{\sigma_{\leq i}}\cdot\mathcal{B}_{\sigma_{>i}} \text{ for } \sigma\in[m^2]^n_{\geq 0}; \tag{14}$$

$$P^{(i)}(\mathbf{x}) = \sum_{\sigma\in[m^2]^n_{\geq 0}}\widehat{P}(\sigma)\,\mathcal{X}_\sigma^{(i)}, \tag{15}$$

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

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

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,

$$\mathrm{Tr}\,\xi(\mathbf{A}) = m\cdot\mathrm{Tr}\,\xi(\mathbf{C}). \tag{16}$$

From Fact 2.18 and then Eq. (16),

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

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

Claim 3.12. 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 2.18, there exists a universal constant $c_3 > 0$ such that

$$\begin{aligned}
&\left|\mathbb{E}\left[m^{i+1-n}\mathrm{Tr}\,\xi\left(P^{(i+1)}\left(\mathbf{x}\right)\right) - m^{i-n}\mathrm{Tr}\,\xi\left(P^{(i)}\left(\mathbf{x}\right)\right)\right]\right| \\
&\leq c_3 B\left(\mathbb{E}\left[|||\mathbf{B}|||_3^3\right] + \mathbb{E}\left[|||\mathbf{D}|||_3^3\right]\right) \\
&\leq c_3 B\left(\mathbb{E}\left[|||\mathbf{B}|||_2|||\mathbf{B}|||_4^2\right] + \mathbb{E}\left[|||\mathbf{D}|||_2|||\mathbf{D}|||_4^2\right]\right) \quad \text{(Hölder's)} \\
&\leq c_3 B\left(\left(\mathbb{E}\left[|||\mathbf{B}|||_2^2\right]\mathbb{E}\left[|||\mathbf{B}|||_4^4\right]\right)^{1/2} + \left(\mathbb{E}\left[|||\mathbf{D}|||_2^2\right]\mathbb{E}\left[|||\mathbf{D}|||_4^4\right]\right)^{1/2}\right) \quad \text{(Cauchy-Schwartz)} \\
&\leq c_3 B\theta^d\left(\left(\mathbb{E}\left[|||\mathbf{B}|||_2^2\right]\right)^{3/2} + \left(\mathbb{E}\left[|||\mathbf{D}|||_2^2\right]\right)^{3/2}\right) \quad \text{(Theorem 3.10)},
\end{aligned}$$

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

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

Therefore,

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

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

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

□

It remains to prove Claim 3.12.

Proof of Claim 3.12. 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

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

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

By Fact 2.17, we have

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

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

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

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

By Fact 2.17, we have

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

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

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Tr} D^2\xi\left(\mathbf{C}\right)\left[\mathbf{D}\right]\right] \\
=& \sum_{j,k,\ell\in[m']} \mathbb{E}\left[\xi^{[2]}\left(\mathbf{a}_j,\mathbf{a}_k,\mathbf{a}_\ell\right)\operatorname{Tr}\left(\mathbf{\Pi}_j\mathbf{D}\mathbf{\Pi}_k\mathbf{D}\mathbf{\Pi}_\ell\right)\right]
\end{aligned}$$

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

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

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 Section 2.2.4, which has a mollifier $\zeta_\lambda(\cdot)$ guaranteed by Fact 2.21.

Lemma 3.13. *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 $\mathrm{Inf}_i(P) \le \tau$ 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 $\sum_{\sigma\neq 0}\widehat{P}(\sigma)^2 \le 1$, we have

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

for some universal constants C.

Proof. Let $\lambda > 0$ be determined later, and ζ_λ be defined as in Fact 2.21. By Theorem 3.11 and Fact 2.21,

$$\left|m^{-n}\mathrm{Tr}\,\zeta_\lambda(P) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\,\zeta_\lambda\left(P^H(\mathbf{x})\right)\right]\right| \le 2c_3B_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d/\lambda,$$

where c_3, B_3 are universal constants. By Fact 2.21 we also have

$$|m^{-n}\mathrm{Tr}\,\zeta(P) - m^{-n}\mathrm{Tr}\,\zeta_\lambda(P)| \le 2\lambda^2$$

and

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

By the triangle inequality, we have

$$\left|m^{-n}\mathrm{Tr}\,\zeta(P) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(P^H(\mathbf{x})\right)\right]\right| \le 4\lambda^2 + 2c_3B_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d/\lambda.$$

Choosing $\lambda = \left(2c_3B_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d/8\right)^{1/3}$, we have

$$\left|m^{-n}\mathrm{Tr}\,\zeta(P) - m^{-h}\,\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(P^H(\mathbf{x})\right)\right]\right| \le 3\left(2c_3B_3\max\left\{9m, 1/\eta^4\right\}^d\sqrt{\tau}d\right)^{2/3}.$$

Let $C = 3(2c_3B_3)^{2/3}$, we conclude the result. □

Remark 3.14. 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 [27, 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 3.11, 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[52] and the references therein). We adopt the pseudorandom generator (PRG) introduced in [40]. 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

$$G\left(f, z^1, \ldots, z^p\right) = x, \text{ where } x_i = z_i^{f(i)} \text{ for } i \in [n]. \tag{17}$$

We define the influence of a random variable in a random matrix using the notation VarInf $(\cdot)$ to distinguish from the notation for the influence of a register in Definition 2.8.

Definition 3.15. Given $m, n, p \in \mathbb{Z}_{>0}$, let $P(\mathbf{b}) = \sum_{S\subseteq[n]} \mathbf{b}_S P_S$ be a random matrix with $\mathbf{b}$ drawn uniformly from $\{\pm 1\}^n$, where $P_S \in \mathcal{H}_m$ and $\mathbf{b}_S = \prod_{i\in S} \mathbf{b}_i$ for all $S \subseteq [n]$. Then the influence of i'th coordinate of $\mathbf{b}$ is defined to be

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

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

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

The following is the main theorem in this section.

Theorem 3.16 (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 $\mathrm{VarInf}_i\left(P(\mathbf{b})\right) \leq \tau$ 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 Eq. (17). Then we have

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\mathrm{Tr}\,\zeta\left(P(\mathbf{b})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x_f}}\left[\mathrm{Tr}\,\zeta\left(\mathbf{P}(\mathbf{x_f})\right)\right]\right| \leq C_2\sqrt{(9m)^d d\tau},$$

where $\mathbf{f}$ is drawn uniformly from $\mathcal{F}$ and C_2 is a universal constant.

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

Theorem 3.17 (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 $\mathbf{b} \sim_{\mathrm{u}} \{-1,1\}^n$, $\mathbb{E}_{\mathbf{b}}\left[\|\|P(\mathbf{b})\|\|_2^2\right] \le 1$, $\mathbf{b}_S = \prod_{i\in S} \mathbf{b}_i$ *and* $P_S \in \mathcal{H}_m^{\otimes h}$, *they satisfy that* $|S| + \deg(P_S) \le d$ *and* $\mathrm{VarInf}_i(P(\mathbf{b})) \le \tau$ *for all* $i \in [n]$.

Let p *be the smallest power of* 2 *satisfying* $p \ge d/\tau$; $\mathcal{F} = \{f : [n] \to [p]\}$ *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 Eq.* (17). *Then for any* $\xi \in C^4$ *with* $\|\xi^{(4)}\|_\infty \le C_0$ *where* C_0 *is a constant, it holds that*

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\mathrm{Tr}\,\xi(P(\mathbf{b}))] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x}_\mathbf{f}}[\mathrm{Tr}\,\xi(\mathbf{P}(\mathbf{x}_\mathbf{f}))]\right| \le 4C_1C_0(9m)^d d\tau,$$

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

Assuming Theorem 3.17, Theorem 3.16 is straightforward:

Proof of Theorem 3.16. Let $\lambda > 0$ be determined later and let ζ_λ be defined as in Fact 2.21. By Theorem 3.17 and Fact 2.21,

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

where C_1, B_4 are universal constants. By Fact 2.21 we also have

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

and

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

By the triangle inequality, we have

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

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

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

Let $C_2 = 8\sqrt{C_1B_4}$, we conclude the result. □

Remark 3.18. It is also possible to generalize Theorem 3.16 to Lipschitz continuous functions using the argument in [27, Lemma 3.5].

Lemma 3.19. *Given* $d, n \in \mathbb{Z}_{>0}$, *and a random matrix*

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

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

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

Proof.

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

□

The following lemma is crucial to our proof. The proof follows closely to the proof of [40, Lemma 5.4].

Lemma 3.20. *Given $d, n, p \in \mathbb{Z}_{>0}$, and a random matrix*

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

satisfying $\mathbb{E}_{\mathbf{b}}\left[|||P(\mathbf{b})|||_2^2\right] \le 1$, 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 $\mathbf{f} \sim_{\mathrm{u}} \mathcal{F}$,

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

Proof. Fix $j \in [p]$ and for $1 \le i \le n$, 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,

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

Thus

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

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

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

The lemma follows by using Lemma 3.19 and summing all $j \in [p]$. □

We are ready to prove Theorem 3.17.

Proof of Theorem 3.17. We prove this by a hybrid argument. Denote $\mathbf{b}^{(0)} = \mathbf{b} = G\left(f, \mathbf{b}, \ldots, \mathbf{b}\right)$. 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

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

By Fact 2.17 and the fact that $\mathbf{z}_j$ is $4d$-wise uniform, we have for $k = 0, 1, 2, 3$,

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

Thus,

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

where the last inequality is from Fact 2.18, 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 3.10 with $\eta \leftarrow 1/\sqrt{3}$ (recall that $\mathbf{b}$ is $(2, 4, 1/\sqrt{3})$-hypercontractive),

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

So we have

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

Summing over $j \in [p]$ and by Lemma 3.20, we have

$$\left|\frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{b}}[\operatorname{Tr}\zeta\left(P(\mathbf{b})\right)] - \frac{1}{m^h}\mathop{\mathbb{E}}_{\mathbf{f},\mathbf{x}_\mathbf{f}}\left[\operatorname{Tr}\zeta\left(\mathbf{P}(\mathbf{x}_\mathbf{f})\right)\right]\right|$$

$$\leq 2C_1C_0(9m)^d\left(\sum_{i=1}^{n}\mathrm{VarInf}_i\left(P(\mathbf{b})\right)^2+\frac{d^2}{p}\right)$$
$$\leq 2C_1C_0(9m)^d\left(\tau\sum_{i=1}^{n}\mathrm{VarInf}_i\left(P(\mathbf{b})\right)+\frac{d^2}{p}\right)$$
$$\leq 4C_1C_0(9m)^d d\tau,$$

where the last inequality is by Lemma 3.19 and $p \geq d/\tau$. □

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 operator or $(\beta+\delta)$-far from all positive semidefinite operators, for error parameters $\beta > \delta > 0$. The input operator is given in the form of a Fourier expansion.

Definition 4.1 (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 $m^{-D}\,\mathrm{Tr}\,\zeta(P) < \beta-\delta$.
- No: if $m^{-D}\,\mathrm{Tr}\,\zeta(P) > \beta+\delta$.

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 4.2. *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 $\mathrm{poly}(D)$.

4.1 Algorithm

The algorithm is shown in Fig. 1, which applies the invariance principle Lemma 3.13 to reduce the dimension of the matrices and then Theorem 3.16 to derandomize, while the distance to positive operators is approximately preserved.

4.2 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

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

Input: Parameters given in Definition 4.1.
Algorithm: Perform the following steps

(1) **Regularization**: Let

$$\tau = \frac{\delta^3}{C' \cdot (9m)^{2d} \cdot d^2}, \tag{18}$$

where $C' = \max\{8C^3, 4C_2^2\}$, with C and C_2 originating from Lemma 3.13 and Theorem 3.16, respectively.

For each i, compute the influence $\mathrm{Inf}_i(P) = \sum_{\sigma:\sigma_i \neq 0} \widehat{P}(\sigma)^2$. Let $H = \{i : \mathrm{Inf}_i(P) > \tau\}$.

(2) **Derandomized invariance principle**: Let p be the smallest power of 2 satisfying $p \geq d/\tau$. Let $n = (m^2 - 1)(D - |H|)$ 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 $(\mathbf{z}^i)$'s be independent across $i \in [p]$. For any $f \in \mathcal{F}$, set $\mathbf{x}_f = G(f, \mathbf{z}^1, \ldots, \mathbf{z}^p)$ as defined in Theorem 3.16. 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 $\mathbf{x}_{f,\sigma_{\bar{H}}} = \prod_{i \notin H} (\mathbf{x}_f)_{(m^2-1)(i-1)+\sigma_i}$ 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

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

Fig. 1. Positivity testing algorithm

(3) By Lemma 2.24 and Corollary 2.25, 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.

4.3 Correctness

Now we proceed to the correctness proof. The first step in our algorithm is to use Lemma 3.13 to reduce the dimension by introducing Rademacher random variables. Let $\mathbf{b} \in \{-1, 1\}^n$ be uniformly distributed. Consider the operator $P^{(1)}$ obtained by replacing the basis outside of H with 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 $\mathcal{B}_{\sigma_H} = \bigotimes_{i \in H} \mathcal{B}_{\sigma_i}$. Recall that $\mathbf{b}$ is $(2, 4, 1/\sqrt{3})$-hypercontractive, so in Lemma 3.13 we have $1/\eta^4 = 9 \leq 9m$. By our choice of τ, the right hand side of the bound in Lemma 3.13 can be upper bounded as

$$C\left((9m)^d \sqrt{\tau} d\right)^{2/3} \leq \delta/2. \tag{20}$$

This implies

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

The second step is derandomization by Theorem 3.16. We define $P^{(2)}$ to be the operator obtained by replacing $\mathbf{b}$ with $\mathbf{x}_{f,\mathbf{z}}$, which is the operator in Eq. (19). Then the right hand side of the bound in Theorem 3.16 can be upper bounded as

$$C_2\sqrt{(9m)^d d\tau} \leq \delta/2. \tag{21}$$

By Theorem 3.16 this implies

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

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.

5 Noisy nonlocal games are NP-complete

Definition 5.1 (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 quantum 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 5.2. *The noisy nonlocal game value problem is* NP*-complete.*

It follows from the two propositions below.

Proposition 5.3. *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 5.4. *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.*

Propositions 5.3 and 5.4 are proved in Sections 5.1 and 5.2 respectively.

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 [53] is $D = \exp(\mathrm{poly}(s), \exp(\mathrm{poly}(t)))$. The follow-up work [54] 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 5.5. *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 quantum maximal correlation $\rho = \rho(\psi_{AB}) < 1$ as defined in Definition 2.1, 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 [53] with several refinements. We include it in Appendix C. [7].

Next we present the algorithm, shown in Fig. 2, which is deterministic provided with a certificate. By Theorem 5.5 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.

Definition 5.6. We summarize the parameters we use for the algorithm in the table below.

- $C_{pt} = 300$.
- $\varepsilon_{rd} = \varepsilon^2/(4t^3)$.
- $\delta = \frac{\varepsilon_{rd}^2}{C_{pt}t(t+1)}$.
- $d = \frac{C_{sm}\log^2\frac{1}{\delta}}{\delta(1-\rho)}$ as in Lemma A.1.
- $s_w = D\log m + \log\left(\frac{2}{\delta}\right)$ as in Lemma D.1.
- D is the polynomial specified in Theorem 5.5 with $\varepsilon \leftarrow \varepsilon/2$.

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) Check 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 test takes time as specified in Theorem 4.2, which is

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

[7] One may wonder why the upper bound in [54] 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 [54], 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.

Input: Parameters in Definition 5.1.

Certificate: Let $\{(\mathcal{A}_i, \mathcal{B}_i)\}_{i=0}^{m^2-1}$ be a pair of standard orthonormal bases satisfying Fact 2.6. 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: Perform the following steps

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

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

where $c_\sigma = c_{\sigma_1}\cdots c_{\sigma_D}$, and $\{c_i\}_{i=0}^{m^2-1}$ is given in Fact 2.6. Reject if

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

(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 tests fails.

(4) Accept.

Fig. 2. Nondeterministic algorithm solving the noisy nonlocal game value problem

By the choices of parameters in Definition 5.6, the overall running time is upper bounded by

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

Completeness. Suppose $\omega^*(G, \psi_{AB}) \geq \beta + \varepsilon$. Then by Theorem 5.5, there exists a strategy $\left(P_a^x, Q_b^y\right)$ that uses D copies of ψ_{AB} with game value $\mathrm{val}_D\left(\left\{P_a^x\right\}, \left\{Q_b^y\right\}\right) \geq \beta + \varepsilon/2$. Let f be the smoothing map in Lemma A.1, 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, $\left\|P_a^{x,(1)}\right\|_2 \le 1$ and $\left\|Q_b^{y,(1)}\right\|_2 \le 1$.
(3) For all x, y, a, b, $\left|\mathrm{Tr}\left(\left(P_a^{x,(1)} \otimes Q_b^{y,(1)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^{x} \otimes Q_b^{y}\right)\psi_{AB}^{\otimes n}\right)\right| \le \delta$.
(4) For all x, y, a, b, $m^{-D}\,\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right) \le \delta$ and $m^{-D}\,\mathrm{Tr}\,\zeta\left(Q_b^{y,(1)}\right) \le \delta$.

We observe that Lemma A.1 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 D.1 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, $\left\|P_a^{x,(2)}\right\|_2 \le 1$ and $\left\|Q_b^{y,(2)}\right\|_2 \le 1$;
(3) For all x, y, a, b, $\left|\mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes Q_b^{y,(2)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^{x,(1)} \otimes Q_b^{y,(1)}\right)\psi_{AB}^{\otimes n}\right)\right| \le \delta$,
(4) For all x, y, a, b, $m^{-D}\,\mathrm{Tr}\,\zeta\left(P_a^{x,(2)}\right) \le 2\delta$ and $m^{-D}\,\mathrm{Tr}\,\zeta\left(Q_b^{y,(2)}\right) \le 2\delta$.

This pseudo-strategy is the certificate. Specifically, by Lemma A.6 the game value is

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

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

Soundness. Suppose that there exists a certificate that passes all tests, 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,
$$\mathrm{val}_D\left(\left\{P_a^{x,(1)}\right\}, \left\{Q_b^{y,(1)}\right\}\right) \ge \beta.$$
- 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 tests, for all x, y, a, b
$$\frac{1}{m^D}\mathrm{Tr}\,\zeta\left(P_a^{x,(1)}\right) \le 6\delta, \text{ and } \frac{1}{m^D}\mathrm{Tr}\,\zeta\left(Q_b^{y,(1)}\right) \le 6\delta.$$

We then apply Lemma A.4 to get a strategy $\left\{P_a^{x,(2)}\right\}$ and $\left\{Q_b^{y,(2)}\right\}$. It holds that for each $x \in \mathcal{X}$

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

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

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

Setup: Flip two unbiased coins $\boldsymbol{c}_1, \boldsymbol{c}_2 \sim \{0,1\}$. Sample questions $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{Alg}_Q(\mathsf{input})$. With probability $1/2$ each, perform one of the following two tests.

Verify: Distribute the questions as follows

- Player $\boldsymbol{c}_1$: give $\boldsymbol{x}$; receive $\boldsymbol{a}$.
- Player $\overline{\boldsymbol{c}_1}$: 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 $\boldsymbol{c}_2 = 0$

- Player $\boldsymbol{c}_1$: give $\boldsymbol{x}$; receive $\boldsymbol{a}$,
- Player $\overline{\boldsymbol{c}_1}$: give $\boldsymbol{x}$; receive $\boldsymbol{b}$,

otherwise

- Player $\boldsymbol{c}_1$: give $\boldsymbol{y}$; receive $\boldsymbol{a}$,
- Player $\overline{\boldsymbol{c}_1}$: give $\boldsymbol{y}$; receive $\boldsymbol{b}$,

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

Fig. 3. The noisy MIP* verifier V^* from an MIP verifier $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$

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

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

Thus there exists a strategy with game value

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

5.2 NP-hardness

In this subsection, we first show that if $L \in \mathrm{MIP}[s,t]$ with perfect completeness and constant soundness, then $L \in \mathrm{MIP}^*[s,t,\psi_{AB}]$ also with perfect completeness and constant soundness.

Theorem 5.7. *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ be an* MIP *protocol for a language L with perfect completeness. Then the verifier V^* described in Fig. 3 is an* MIP* *verifier for L with the following conditions:*

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

Soundsness. *Given* input*, if there is a strategy for V^* with value $1-\epsilon$ where the provers share arbitrarily many copies of a noisy MES, then there is a classical strategy for V with value $1-2\varepsilon-\frac{32\epsilon}{1-\rho}$.*

Proof of Proposition 5.4. Notice that 3-SAT $\in$ MIP[log, 1] [8] with perfect completeness and an arbitrarily small constant soundness. By Theorem 5.7, there exists an MIP*[log, 1, ψ_{AB}] protocol for 3-SAT with perfect completeness and a constant soundness. □

Proof of Theorem 5.7. **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 Alice's question x. Suppose that Alice and Bob share n copies of a noisy m-dimensional MES ψ_{AB}, and that they apply the measurements $\left\{P_a^x\right\}_{a\in\mathcal{A}}$ and $\left\{Q_a^x\right\}_{a\in\mathcal{A}}$, respectively. Hence the probability for the provers to pass the consistency test of x is at least $1-4\epsilon$. It means that

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

Let $\{(\mathcal{A}_i,\mathcal{B}_i)\}_{i=0}^{m^2-1}$ be a pair of standard orthonormal bases satisfying Fact 2.6. Using the Fourier expansions of $P_a^x=\sum_\sigma \widehat{P_a^x}(\sigma)\,\mathcal{A}_\sigma$ and $Q_a^x=\sum_\sigma \widehat{Q_a^x}(\sigma)\,\mathcal{B}_\sigma$, the condition above is equivalent to

$$\mathop{\mathbb{E}}_x \sum_a\sum_\sigma c_\sigma \widehat{P_a^x}(\sigma)\,\widehat{Q_a^x}(\sigma)\geq 1-4\epsilon,$$

where $c_\sigma=c_{\sigma_1}\cdots c_{\sigma_D}$, and $\{c_i\}_{i=0}^{m^2-1}$ is given in Fact 2.6. By the Cauchy-Schwartz inequality,

$$\left(\mathop{\mathbb{E}}_x \sum_a\sum_\sigma c_\sigma \widehat{P_a^x}(\sigma)^2\right)^{1/2}\left(\mathop{\mathbb{E}}_x \sum_a\sum_\sigma c_\sigma \widehat{Q_a^x}(\sigma)^2\right)^{1/2}\geq 1-4\epsilon.$$

Notice that

$$\sum_a\sum_\sigma c_\sigma \widehat{Q_a^x}(\sigma)^2\leq \sum_a\sum_\sigma \widehat{Q_a^x}(\sigma)^2=\sum_a \left\|Q_a^x\right\|_2^2\leq 1$$

for all x, we have

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

On the other hand, we have

$$\begin{aligned}\mathop{\mathbb{E}}_x \sum_a\sum_\sigma c_\sigma \widehat{P_a^x}(\sigma)^2 &\leq \mathop{\mathbb{E}}_x \sum_a\left[\widehat{P_a^x}(0^n)^2+\rho\sum_{\sigma\neq 0^n}\widehat{P_a^x}(\sigma)^2\right]\\ &\leq \mathop{\mathbb{E}}_x \sum_a\left[\widehat{P_a^x}(0^n)^2+\rho\left(\left\|P_a^x\right\|_2^2-\widehat{P_a^x}(0^n)^2\right)\right]\\ &\leq \rho+(1-\rho)\mathop{\mathbb{E}}_x\sum_a \widehat{P_a^x}(0^n)^2.\end{aligned}$$

Therefore,

$$\mathop{\mathbb{E}}_x\sum_a \widehat{P_a^x}(0^n)^2\geq 1-\frac{8\epsilon-16\epsilon^2}{1-\rho}\geq 1-\frac{8\epsilon}{1-\rho}. \tag{23}$$

Note that for all x, $\sum_a \widehat{P_a^x}(0^n) = m^{-n} \sum_a \operatorname{Tr} P_a^x = 1$. For each x, let a_x be the answer that maximizes $\widehat{P_a^x}(0^n)$. Then $\sum_a \widehat{P_a^x}(0^n)^2 \le \widehat{P_{a_x}^x}(0^n) \sum_a \widehat{P_a^x}(0^n) = \widehat{P_{a_x}^x}(0^n)$, and

$$\mathop{\mathbb{E}}_x \widehat{P_{a_x}^x}(0^n) \ge 1 - \frac{8\epsilon}{1-\rho}.$$

Similarly, from the fact that they can pass the consistency test on Bob's questions, we can conclude that the measurements $\{Q_b^y\}$ satisfy the conditions above. In particular, let b_y be the answer that maximizes $\widehat{Q_b^y}(0^n)$. Then

$$\mathop{\mathbb{E}}_y \widehat{Q_{b_y}^y}(0^n) \ge 1 - \frac{8\epsilon}{1-\rho}.$$

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

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

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

$$\begin{aligned}
&2 \mathop{\mathbb{E}}_{x,y} \sum_{\substack{a \ne a_x \text{ or} \\ b \ne b_y}} \operatorname{Tr}\left[\left(P_a^x \otimes Q_b^y\right) \psi_{AB}^{\otimes n} \right] \\
&\le 2 \mathop{\mathbb{E}}_{x,y} \sum_{a \ne a_x, b} \operatorname{Tr}\left[\left(P_a^x \otimes Q_b^y\right) \psi_{AB}^{\otimes n} \right] + 2 \mathop{\mathbb{E}}_{x,y} \sum_{b \ne b_y} \operatorname{Tr}\left[\left(P_{a_x}^x \otimes Q_b^y\right) \psi_{AB}^{\otimes n} \right] \\
&\le 2 \mathop{\mathbb{E}}_{x,y} \left[\sum_{a \ne a_x} \widehat{P_a^x}(0^n) + \sum_{b \ne b_y} \widehat{Q_b^y}(0^n) \right] \\
&\le \frac{32\epsilon}{1-\rho}.
\end{aligned}$$

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 $1 - 2\varepsilon - 32\varepsilon/(1-\rho)$. □

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 first step is to tighten the answer reduction techniques from [31, 48] so that they can reduce the verifier's answer size sequentially from $\mathrm{poly}(n)$ to $\mathrm{polylog}(n)$, and then to $\mathrm{polyloglog}(n)$. The second step is to develop a new answer reduction technique that can reduce the answer size of an MIP* protocol from $\mathrm{polyloglog}(n)$ to $O(1)$ while preserving other parameters of the protocol.

We achieve the second step by modifying the answer reduction technique from [47]. 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 be 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*[poly, $O(1)$] protocol for RE, we successively apply the tightened answer reduction technique, followed by the new technique with the Hadamard code, to the MIP*[poly, poly] protocol for RE [31].

Note that [31] doesn't use the answer reduction technique of [47]. The authors of [31] 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 [31] is more difficult to modify due to its less modular design.

6.1 Tighter answer reduction

For MIP* protocols with short answers, it is useful to separate the part of the verifier that directly acts on the answer bits from the rest. Without loss of generality, we imagine that the decider Alg_A acts in two phases: first, given its internal randomness and the question pair x, y, it computes a Boolean circuit $C_{x,y}^{\mathsf{input}}$, and then it applies $C_{x,y}^{\mathsf{input}}$ to the answers and returns its output. The *verification time* is the total runtime of this process. We define the *decision complexity* denoted by $d_{V,A}(n)$ to be the maximal size of the circuit $C_{x,y}^{\mathsf{input}}$ over all input of size n and question x, y sampled by $\mathsf{Alg}_Q(\mathsf{input})$. This is always at most as large as the verification time, but it can be much smaller.

In this section, we will observe that tighter bounds on the answer reduction theorem of [31, 48] can be given in terms of the decider complexity. In particular, the next theorem is a refined version of [48, Theorem 51] as the next theorem can be applied recursively to reduce the answer size below $\log(n)$.

Theorem 6.1. *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ be an MIP* protocol for a language L, with question length $\ell_{V,Q}(n)$, answer length $\ell_{V,A}(n)$, sampling time $t_{V,Q}(n)$, verification time $t_{V,A}(n)$, and decision complexity $d_{V,A}(n)$. Suppose further that V has the following property: for any $\mathsf{input} \in L$, the prover has a real commuting symmetric EPR strategy of value 1. Then there is an answer-reduced verifier $V^{AR} = (\mathsf{Alg}'_Q, \mathsf{Alg}'_A)$ with the following properties:*

Question length. *The new question length is $2\ell_{V,Q}(n) + \mathrm{polylog}(d_{V,A}(n))$.*

Answer length. *The new answer length is $\mathrm{polylog}(d_{V,A}(n))$.*

Sampling time. *The new sampling time is $t_{V,Q}(n) + \mathrm{polylog}(d_{V,A}(n))$.*

Verification time. *The new verification time is $t_{V,Q}(n) + t_{V,A}(n) + \mathrm{poly}(d_{V,A}(n))$.*

Decision complexity. *The new decision complexity is $\mathrm{polylog}(d_{V,A}(n))$.*

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

Soundness. *Given input, if there is a strategy to V^{AR} with value $1 - \epsilon$, then there is a strategy to V with value $1 - \delta(\epsilon, n)$, where $\delta(\epsilon, n) = a((\log d_{V,A}(n))^a \epsilon^b + (\log d_{V,A}(n))^{-100b})$, a is a universal constant such that $a > 0$, and b is a universal constant such that $0 < b < 1$.*

Efficient computability. *There exists an algorithm that takes the description of $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ as input and outputs the description of $V^{AR} = (\mathsf{Alg}'_Q, \mathsf{Alg}'_A)$ in time $O(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A|)$, where $|\mathsf{Alg}_Q|$ and $|\mathsf{Alg}_A|$ denote the sizes of the descriptions of Alg_Q and Alg_A respectively. Moreover, $|\mathsf{Alg}'_Q| = |\mathsf{Alg}_Q| + O(1)$ and $|\mathsf{Alg}'_A| = |\mathsf{Alg}_A| + O(1)$.*

Proof. The proof will follow [48] very closely. The main nontrivial thing to prove is the bound on the decision complexity of V^{AR}.

In more detail, the answer-reduced verifier first apply the Cook-Levin theorem to the circuit $C_{x,y}^{\mathsf{input}}$ computed in the first phase of Alg_A to get a 5SAT instance with size $s = \mathrm{poly}(d_{V,A}(n))$. Following the proof of [48, Theorem

51], we can choose the parameters used in the PCP part of the answer reduction techniques as

$$\begin{aligned} m &= O(\log s) = O(\log(d_{V,A}(n))), \\ m' &= 5m + 5 = O(\log(d_{V,A}(n))), \\ q &= 2^{O(n)} = \mathrm{poly}(d_{V,A}(n)), \\ d &= O(m) = O(\log(d_{V,A}(n))), \end{aligned}$$

so that the proofs of the provers are $m' + 6$ low-degree polynomials on $\mathbb{F}_q^m$ and $\mathbb{F}_q^{m'}$ of individual degree at most d in each variable.

Question size. The question of V^{AR} has at most two questions sampled by V and queries to the low-degree polynomials prepared by the provers. Hence,

$$\ell_{V^{AR},Q}(n) = 2\ell_{V,Q}(n) + O(m' \log q) = 2\ell_{V,Q}(n) + \mathrm{polylog}(d_{V,A}(n)).$$

Answer size. The answer size is at most $O((m' + 6)(d + 1) \log q)$, which is the number of bits to specify the coefficients in $\mathbb{F}_q$ of $m' + 6$ polynomials of degree at most d. Hence,

$$\ell_{V^{AR},A}(n) = O(m' d \log q) = \mathrm{polylog}(d_{V,A}(n)).$$

Sampling time. The verifier will first sample the questions x, y and then sample queries to the low-degree polynomials. Following the proof of [31, Theorem 10.27], we have

$$t_{V^{AR},Q}(n) = t_{V,Q}(n) + \mathrm{poly}(m', \log q) = t_{V,Q}(n) + \mathrm{polylog}(d_{V,A}(n)).$$

Verification time. Following the description of the answer-reduced verifier in [31, Figure 14], we can see that the run time of the answer-reduced verifier is sum of the run time of each step. Step 1 and 2 are consistency checks, so the run time is $O(\ell_{V^{AR},A}(n))$. Step 3 and 4 are low-degree tests, so the run time is at most $\mathrm{poly}(m, d, m', \log q) = \mathrm{polylog}(d_{V,A}(n))$. In Step 5, Alg'_A needs to run the functions L^A and L^B of the original Alg_Q first, which takes time $t_{V,Q}(n)$ and then the PCP verification. The PCP verification takes time $\mathrm{poly}(s) + \mathrm{poly}(s, \log q) + \mathrm{poly}(m', \log q) + \mathrm{poly}(\log q) = \mathrm{poly}(d_{V,A}(n))$ according to the proof of [31, Theorem 10.25]. Hence, overall,

$$t_{V^{AR},A}(n) = t_{V,Q}(n) + t_{V,A}(n) + \mathrm{poly}(d_{V,A}(n)).$$

Decision complexity. The checks performed by V^{AR} in [48] are the same as those in [31], which are specified in [31, Figure 14]. These checks are simple arithmetics over $\mathbb{F}_q$ as explained below.

(1) Step 1 and 2 of V^{AR} as in [31, Figure 14] are consistency checks.
(2) Step 3 and 4 of V^{AR} as in [31, Figure 14] are low-degree test, which is evaluating a univariate polynomial over $\mathbb{F}_q$ of degree at most $m'd$, whose coefficients are specified by the prover, at a point chosen by the verifier. Note that the checks in these 4 steps do not depend on $C^{\mathsf{input}}_{x,y}$.
(3) In Step 5, the checks of the PCP verifier are executed:
   (a) In Step 4 of the PCP verifier as in [31, Figure 13], the values of $\mathcal{F}_{arith}(x, o, w), o_1, \ldots, o_5$ are precomputed, so the verifier only needs to evaluate an individual-degree-1 polynomial on the prover's answers.
   (b) In Step 5 of the PCP veifier as in [31, Figure 13], the values of $\mathrm{zero}(z_i)$ are precomputed, so again the verifier only needs to evaluate an individual-degree-1 polynomial on the prover's answers.

Hence, $d_{V^{AR},A}(n) = \mathrm{poly}(\ell_{V^{AR},A}(n)) = \mathrm{polylog}(d_{V,A}(n))$.

Completeness, soundness and efficient computability follow from the same argument in the proof of [48, Theorem 51]. The soundness error is calculated using the values of q, m, m' and d in our setting. □

We will actually use a parallel-repeated version of this answer reduction, which obtains constant soundness. This is closely modeled on [48, Theorem 52].

Theorem 6.2. *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ be an* MIP* *protocol for a language L, with question length $\ell_{V,Q}$, sampling time $t_{V,Q}$, verification time $t_{V,A}$, and decision complexity $d_{V,A}$. Suppose further that V has the following property: for any* input $\in L$*, the prover has a real commuting symmetric EPR strategy with a value* 1*. Then there exists an efficiently computable function $k(n) = \mathrm{polylog}(d_{V,A}(n))$ and an answer-reduced verifier $V^{AR} = (\mathsf{Alg}'_Q, \mathsf{Alg}'_A)$ such that the following hold:*

Question length. *The new question length is $k(n) \cdot (2\ell_{V,Q}(n) + \mathrm{polylog}(d_{V,A}(n)))$.*
Answer length. *The new answer length is $k(n) \cdot \mathrm{polylog}(d_{V,A}(n))$.*
Sampling time. *The new sampling time is $k(n) \cdot (t_{V,Q}(n) + \mathrm{polylog}(d_{V,A}(n)))$.*
Verification time. *The new verification time is $k(n) \cdot (t_{V,Q}(n) + t_{V,A}(n) + \mathrm{poly}(d_{V,A}(n)))$.*
Decision complexity. *The new decision complexity is $k(n) \cdot \mathrm{polylog}(d_{V,A}(n))$.*
Completeness. *If* input $\in L$*, there is a value-*1 *strategy for V^{AR}.*
Soundness. *Given* input*, if the value of V is at most $1/2$, then the value of V^{AR} on* input *is also at most $1/2$.*
Efficient computability. *There exists an algorithm that takes the description of $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ as input and outputs the description of $V^{AR} = (\mathsf{Alg}'_Q, \mathsf{Alg}'_A)$ in time $O(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A|)$. Moreover, $|\mathsf{Alg}'_Q| = |\mathsf{Alg}_Q| + O(1)$ and $|\mathsf{Alg}'_A| = |\mathsf{Alg}_A| + O(1)$.*

The choice of $k(n)$ follows the same reasoning in the proof of [48, Theorem 52]. The only difference is that here s, the size of the SAT formula, is $\mathrm{poly}(d_{V,A}(n))$, instead of $\mathrm{poly}(t_{V,A}(n), |\mathsf{Alg}_A|)$, so $k(n) = \mathrm{polylog}(s) = \mathrm{polylog}(d_{V,A}(n))$.

We note that, unlike in the answer reduction theorems of [31], we do not keep track of the *level* of the sampler in our answer reduction theorems. To recall, the level is a measure of the complexity of the sampler distribution when expressed in terms of a "conditional linear" function of a uniform random seed. The level is important in the context of [31] because it affects the complexity of the question reduction procedure, in which the sampling of questions is delegated to the provers. Thus, in that paper, it was important to keep track of the level to make sure that it remains bounded by a universal constant, so that question reduction can be recursively applied. In our setting, we will never apply question reduction directly, so we do not need to track the level. It can be checked that the answer reduction theorems as stated here hold for all levels, and all of the asymptotic bounds are independent of the number of levels.

Remark 6.3. The important conclusion from Theorem 6.2 is that answer reduction shrinks the decision complexity of a protocol. Looking ahead, this will help us when we repeatedly apply answer reduction. One might ask how the decision complexity behaves under *question reduction*, which is the other component of the compression procedure in [31]. It turns out that question reduction will in general always "reset" the decision complexity to be the same as the decider runtime prior to question reduction. This is because question reduction delegates sampling the questions to the provers, so the "new" decider, after question reduction has been applied, must wait for the "new" answers in order to simulate the computation of the old decider. Thus, even if the old decider could do most of its work as precomputation before the answers were received, the new decider may not be able to do any precomputation, so the decision complexity can only be bounded by the runtime of the old decider.

6.2 Subset tester for the Hadamard code

To use the [47] 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 [47, Section 16], shown in Fig. 4, 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 6.4. 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,

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 $\overline{\boldsymbol{b}}$: 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 $\overline{\boldsymbol{b}}$: 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 $\overline{\boldsymbol{b}}$: 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 $\overline{\boldsymbol{b}}$: 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 $\overline{\boldsymbol{b}}$: 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 $\overline{\boldsymbol{b}}$: give $\boldsymbol{x}_i + \boldsymbol{y}$ for a random i; receive $\boldsymbol{d}$.

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

Fig. 4. Subset tester for the Hadamard code

- 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. Hence the running time of the decoding algorithm is $t_{\mathrm{Dec}}(k) = O(2^k)$. Note that both Enc_k and Dec_k run in time exponential in k.

Proposition 6.5. *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*

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

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 $|\psi'\rangle = |\psi\rangle \otimes |aux\rangle$

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

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

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

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:

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

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

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

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

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

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

$$\begin{aligned}
&\mathop{\mathbb{E}}_{F\sim D} \mathop{\mathbb{E}}_{y\in D_{\text{Unif}}} \sum_{\substack{a,c,a': \\ a_i+c=a'_i \forall i}} \langle\psi| M_{a,c,a'}^{F,y} \otimes N_a^F |\psi\rangle \geq 1 - 18\varepsilon \\
&\mathop{\mathbb{E}}_{F\sim D} \mathop{\mathbb{E}}_{y\in D_{\text{Unif}}} \sum_{\substack{a,c,a': \\ a_i+c=a'_i \forall i}} \langle\psi| M_{a,c,a'}^{F,y} \otimes N_c^y |\psi\rangle \geq 1 - 18\varepsilon \\
&\mathop{\mathbb{E}}_{F\sim D} \mathop{\mathbb{E}}_{y\in D_{\text{Unif}}} \sum_{\substack{a,c,a': \\ a_i+c=a'_i \forall i}} \langle\psi| M_{a,c,a'}^{F,y} \otimes N_{a'}^{F+y} |\psi\rangle \geq 1 - 18\varepsilon
\end{aligned}$$

$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{\substack{a,c,a':\\ a_i+c=a'_i\forall i}}\langle\psi|\, M^{F,y}_{a,c,a'}\otimes \mathbb{1}_B\,|\psi\rangle \geq 1-6\varepsilon$$

$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{a\in\mathbb{F}_2^k}\langle\psi|\, N^{F+y}_{a}\otimes N^{x_i+y}_{a_i}\,|\psi\rangle \geq 1-6k\varepsilon \quad \text{for all } i$$

$$\mathop{\mathbb{E}}_{F\sim D}\sum_{a\in\mathbb{F}_2^k}\langle\psi|\, N^{F}_{a}\otimes N^{F}_{a}\,|\psi\rangle \geq 1-12\varepsilon$$

$$\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\sum_{a\in\mathbb{F}_2}\langle\psi|\, N^{x_i+y}_{a}\otimes N^{x_i+y}_{a}\,|\psi\rangle \geq 1-12k\varepsilon \quad \text{for all } i.$$

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

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

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**,

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

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

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

$$|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\,\mathbb{1}_A\otimes N^{x_i+y|F+y}N^yN^{x_i|F}\,|\psi\rangle-1|\le 38\sqrt{\varepsilon}. \tag{24}$$

In the first step

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\,M^{x_i|F,y}M^{y|F,y}(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})\,|\psi\rangle|\\
&\le\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|M^{y|F,y}M^{x_i|F,y}\otimes\mathbb{1}_B\,|\psi\rangle\|\cdot\|(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})\,|\psi\rangle\| \quad \text{(Cauchy-Schwarz)}\\
&=\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})\,|\psi\rangle\|\\
&\le\sqrt{\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|(M^{x_i+y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y|F+y})\,|\psi\rangle\|^2} \quad \text{(Jensen)}\\
&\le 6\sqrt{2\varepsilon}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\,M^{x_i|F,y}\otimes N^{x_i+y|F,y}\cdot(M^{y|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^y)\,|\psi\rangle|\le 6\sqrt{2\varepsilon}\\
&|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\,\mathbb{1}_A\otimes N^{x_i+y|F+y}N^y\cdot(M^{x_i|F,y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i|F})\,|\psi\rangle|\le 6\sqrt{2\varepsilon}.
\end{aligned}$$

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

$$\begin{aligned}
&\mathop{\mathbb{E}}_{F\sim D}\|(N^{x_i|F}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i|F})\,|\psi\rangle\|^2\le 48\varepsilon\\
&\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\|(N^{x_i+y}\otimes\mathbb{1}_B-\mathbb{1}_A\otimes N^{x_i+y})\,|\psi\rangle\|^2\le 48k\varepsilon\quad\text{for all } i.
\end{aligned}$$

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 Eq. (24) implies that

$$|\mathop{\mathbb{E}}_{F\sim D}\mathop{\mathbb{E}}_{y\in D_{\mathsf{Unif}}}\langle\psi|\, N^{x_i|F}\otimes N^{x_i+y}N^y\,|\psi\rangle - 1| \le (45+12\sqrt{k})\sqrt{\varepsilon}. \tag{25}$$

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 $\left\{\sum_{u:u\cdot x_i=0}G_u, \sum_{u:u\cdot x_i=1}G_u\right\}$, where $N_c^{x_i|F} = \sum_{b:b_i=c}N_b^F$ for $c=0,1$, 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 Eq. (25). 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

$$\begin{aligned}
&\mathop{\mathbb{E}}_{F\sim D}\sum_{a=0,1}\langle\psi'|\,(N_a^{x_i|F}\otimes\mathbb{1}_{\mathcal{H}'})\otimes(\sum_{u:u\cdot x_i=a}\hat{G}_u)\,|\psi'\rangle\\
&=\mathop{\mathbb{E}}_{F\sim D}\sum_{a=0,1}\langle\psi|\,N_a^{x_i|F}\otimes\left(\sum_{u:u\cdot x_i=a}(\mathbb{1}\otimes\langle aux|)\hat{G}_u(\mathbb{1}\otimes|aux\rangle)\right)|\psi\rangle\\
&=\mathop{\mathbb{E}}_{F\sim D}\sum_{a=0,1}\langle\psi|\,N_a^{x_i|F}\otimes\left(\sum_{u:u\cdot x_i=a}G_u\right)|\psi\rangle\\
&\approx_{C_1/2\sqrt{\varepsilon}}1.
\end{aligned}$$

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 2.36

$$\mathop{\mathbb{E}}_{F\sim D}\sum_{a\in\mathbb{F}_2^k}\|N_a^F\otimes\mathbb{1}_{\mathcal{H}'}\otimes\mathbb{1}_B\left|\tilde{\psi}\right\rangle - \mathbb{1}_A\otimes\sum_{\substack{u:u\cdot x_i=a_i\\ \forall i\in[k]}}\hat{G}_u\left|\tilde{\psi}\right\rangle\|^2 \le (2k-1)^2C_1\sqrt{\varepsilon},$$

which completes the proof. □

6.3 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 [47, Section 17.4] by the Hadamard code. The other key ingredient of Natarajan and Wright's answer reduction is Probabilistically Checkable Proofs of Proximity (PCPP), so we recall its definition and key properties that we will use later.

Definition 6.6 (PCPP)**.** For functions $r, q : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, $t : \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, an (r, q, t)-restricted PCPP verifier is a probabilistic machine that, given a string x (called the *explicit* input) and a number K (in binary) as well as oracle access to an implicit input $y \in \{0,1\}^K$ and to a *proof oracle* $\pi \in \{0,1\}^*$, tosses $r(|x| + K)$ coins, queries the oracles (y, π) for a total of $q(|x| + K)$ symbols, runs in time $t(|x|, K)$, and outputs a Boolean verdict.

For 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 of V on input $(x, |y|)$, oracle access to (y, π), and randomness R.

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, $\Pr_R[V^{y,\pi}(x, |y|; R) = 1] \le s$.

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

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

Definition 6.7. Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ 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 $\ell_{V,Q}(n)$ and answer length $\ell_{V,A}(n)$. We define

$$L_{\mathsf{Enc}} = \left\{ (\mathsf{input}, x_0, x_1, \mathsf{Enc}_{\ell_{V,A}(|\mathsf{input}|)}(y_0), \mathsf{Enc}_{\ell_{V,A}(|\mathsf{input}|)}(y_1)) \mid C^{\mathsf{input}}_{x_0,x_1}(y_0, y_1) = 1 \right\},$$

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 $\mathsf{Dec}_{\ell_{V,A}(n)}$ as pointed out in [47, Proposition 17.7]. Suppose $\gamma \le \eta_H/2 = 1/4$. Then by [47, 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

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

where s is the soundness of V_{PCPP}.

Definition 6.8. We instantiate the answer-reduced MIP* protocol with the following components and notations.

- 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 $\ell_{V,A}(n)$, question sampling time $t_{V,Q}(n)$ and answer verification time $t_{V,A}(n)$.
- Let $G_k(\boldsymbol{T})$ be the subset tester from Section 6.2 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.
- Let L_{Enc} be the language defined in Definition 6.7, which is in $\mathrm{TIME}(T)$ with

$$T(n) = t_{V,A}(n) + t_{\mathrm{Dec}}(\ell_{V,A}(n)).$$

and let V_{PCPP} be its PCPP verifier with $\gamma \leq 1/4$ and constant soundness s. The verification time is

$$t_{\mathrm{PCPP},A}(n) = \mathrm{poly}(n + \ell_{V,Q}(n), \log_2(2^{\ell_{V,A}(n)}), \log_2(T(n))).$$

The sampling time is also upper bounded by the verification time of the PCPP, which includes both the sampling time and answer verification time, so

$$t_{\mathrm{PCPP},Q}(n) = \mathrm{poly}(n + \ell_{V,Q}(n), \log_2(2^{\ell_{V,A}(n)}), \log_2(T(n))).$$

Finally, on inputs of size n, the proof length is

$$\ell_\pi(n) = 2^{r(n)} = T(n) \cdot \mathrm{polylog}(T(n)),$$

where $r(n)$ is the randomness complexity of the PCPP verifier.

- 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 [47, Figure 15], but we include it for completeness.

Theorem 6.9. *Let $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ be an* MIP* *protocol for a language L, with question length $\ell_{V,Q}$, answer length $\ell_{V,A}$, sampling time $t_{V,Q}$ and verification time $t_{V,A}$. Suppose the* PCPP *verifier is chosen so that $\gamma \leq 1/4$. Suppose further that V has the following property: for any* 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 as shown in Section 6.3 is also an* MIP* *verifier for L with the following two conditions:*

Question length. *The new question length is $\ell_{V^{AR},Q}(n) = O(\ell_{V,Q}(n) + \ell_1(n) + \ell_2(n))$.*
Answer length. *The new answer length is $\ell_{V^{AR},A}(n) = O(1)$.*
Sampling time. *The new question sampling time is*

$$t_{V^{AR},Q}(n) = t_{V,Q}(n) + \mathrm{poly}(n + \ell_{V,Q}(n), \ell_1(n), \log_2(T(n))) + O(\ell_1(n) + \ell_2(n)).$$

Verification time. *The new verification time is $t_{V^{AR},A} = \mathrm{poly}(n + \ell_{V,Q}(n), \ell_1(n), \log_2(T(n)))$.*
Completeness. *If* 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 $1 - K_1 - K_2\varepsilon^{1/128}$.*
Efficient computability. *There exists an algorithm that takes the description of $V = (\mathsf{Alg}_Q, \mathsf{Alg}_A)$ as input and outputs the description of $V^{AR} = (\mathsf{Alg}'_Q, \mathsf{Alg}'_A)$ in time $O(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A|)$. Moreover, $|\mathsf{Alg}'_Q| = |\mathsf{Alg}_Q| + O(1)$ and $|\mathsf{Alg}'_A| = |\mathsf{Alg}_A| + O(1)$.*

Remark 6.10. Since $T(n) = O(T_{V,A}(n) + 2^{\ell_{V,A}(n)}) = O(T_{V,A}(n) + 2^{\ell_1(n)})$ and $\ell_2 = T(n)\,\mathrm{polylog}(T(n))$, the new question length, $\ell_{V^{AR},Q}(n)$, and question sampling time, $t_{V^{AR},Q}(n)$, after applying the Hadamard-based answer reduction depend *exponentially* on the answer length of the original protocol.

Proof. **Question length**. The question of V^{AR} consists of a question from V and queries to the encodings of the answers and the PCPP proof. Hence,

$$\ell_{V^{AR},Q}(n) \leq 2\ell_{V,Q}(n) + 2(\kappa + q(n)) \cdot \ell_1(n) + (\kappa + q(n)) \cdot \ell_2(n) = O(\ell_{V,Q}(n) + \ell_1(n) + \ell_2(n)),$$

where $q(n)$ is the query complexity of the PCPP verifier.

Answer length. The prover only needs to answer the queries, so the answer consists of at most $3(\kappa + q(n))$ bits.

The answer reduced verifier V^{AR}

Setup: Flip one unbiased coin $\boldsymbol{b} \sim \{0,1\}$. Sample questions $(\boldsymbol{x}_0, \boldsymbol{x}_1) \sim \mathsf{Alg}_Q(\text{input})$. Sample a view $\boldsymbol{I}_0, \boldsymbol{I}_1, \boldsymbol{J} \sim V_{\text{PCPP}}(\text{input}, \boldsymbol{x}_0, \boldsymbol{x}_1)$. 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 [a].

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 $V_{\text{PCPP}}(\text{input}, \boldsymbol{x}_0, \boldsymbol{x}_1)$ 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: For $c = 0, 1$, 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 $\boldsymbol{x}_c, \boldsymbol{T}_c$; receive $\boldsymbol{a}'_c$

Accept if $\boldsymbol{a}_c = \boldsymbol{a}'_c$.

Answer consistency check: For $c = 0, 1$, distributed the questions as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{x}_c, \boldsymbol{T}_c$; receive $\boldsymbol{a}_c$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}_c, \boldsymbol{T}_c$; receive $\boldsymbol{a}'_c$

Accept if $\boldsymbol{a}_c = \boldsymbol{a}'_c$.

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: For $c = 0, 1$, sample questions $(\boldsymbol{w}_0, \boldsymbol{w}_1) \sim G_{\ell_1}(\boldsymbol{T}_c)$. Distributed the questions as follows:
- Player $\boldsymbol{b}$: give $\boldsymbol{x}_c, \boldsymbol{w}_0$; receive $\boldsymbol{a}_0$.
- Player $\bar{\boldsymbol{b}}$: give $\boldsymbol{x}_c, \boldsymbol{w}_1$; receive $\boldsymbol{a}_1$.

Accept if $G_{\ell_1}(\boldsymbol{T}_c)$ 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 $\boldsymbol{a}_1$.

Fig. 5. The answer reduced verifier V^{AR}.

[a]There are two answer cross-checks, one for $c = 0$ and one for $c = 1$, similarly for the answer consistency checks and answer code checks.

Sampling time. The sampling algorithm of the answer-reduced protocol needs to run the sampling algorithm of the PCPP verifier, which takes time

$$t_{\text{PCPP},Q}(n) = \text{poly}(n + \ell_{V,Q}(n), \log_2(2^{\ell_{V,A}(n)}), \log_2(T(n))).$$

The sampling algorithm must also run Alg_Q, the embedding algorithm $\mu_{\ell_2(n)}$, and sample random indices in the encodings. Hence, the sampling time is

$$\begin{aligned} t_{V^{AR},Q}(n) &\le t_{V,Q}(n) + t_{\text{PCPP},Q}(n) + q(n) \cdot \ell_2(n) + 2\kappa\ell_1(n) + \kappa\ell_2(n) \\ &= t_{V,Q}(n) + \text{poly}(n + \ell_{V,Q}(n), \log_2(2^{\ell_{V,A}(n)}), \log_2(T(n))) + O(\ell_1(n) + \ell_2(n)), \end{aligned}$$

where $q(n) = O(1)$ is the query complexity of the PCPP verifier.

Verification time. Since the verification time of the code checks and consistency tests are $O(1)$,

$$t_{V^{AR},A}(n) \leq t_{\mathrm{PCPP},A}(n) + O(1) = \mathrm{poly}(n + \ell_{V,Q}(n), \ell_1(n), \log_2(T(n))).$$

Efficient computatbility. This follows from the observation that the descriptions of Alg'_Q and Alg'_A contains both Alg_Q and Alg_A respectively with new instructions. Since the new instructions added to Alg_Q and Alg_A are independent of input, Alg_Q and Alg_A, the time to write them down are $O(1)$, and their sizes are also $O(1)$.

Completeness. This follows the same proof of the completeness part of [47, Theorem 17.10]. If an honest prover gets one question x_b, the prover will compute its answer, encode its answer using the Hadamard code, and answer the queries accordingly. If an honest prover gets both questions, the prover will compute its answers, compute a PCPP proof to certify these answers are correct, compute the Hadamard encodings of the answers and the proof, and answer the queries accordingly.

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 $1 - K_1 - K_2\varepsilon^{1/128}$ is greater than the soundness s 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 2.37.

Let $C_I = |\boldsymbol{I}_0| + \kappa = |\boldsymbol{I}_1| + \kappa$ and $C_J = |\boldsymbol{J}| + \kappa$ be two constants. 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 6.5 to $G_{\ell_1}(\boldsymbol{T}_c)$ where the distribution of $\boldsymbol{T}_c$ denoted by D_{x_c} is determined by c and x_c. Proposition 6.5 implies that there exists Hilbert spaces $\mathcal{H}_{x_c}$, $\left|aux_{x_c}\right\rangle \in \mathcal{H}_{x_c} \otimes \mathcal{H}_{x_c}$ and projective measurement $\left\{G_u^{x_c}\right\}$ on $\mathcal{H}^{x_c}$ such that

$$\mathop{\mathbb{E}}_{T_c \sim D_{x_c}} \sum_{a \in \mathbb{F}_2^{C_I}} \left\| (M_a^{x_c,T_c} \otimes \mathbb{1}_{\mathcal{H}_{A,x_c}} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes G^{x_c}_{[w|_{T_c}=a]}) |\psi\rangle \otimes \left|aux_{x_c}\right\rangle \right\|^2 \leq O(C_I^3 \sqrt{\varepsilon_{c,x_c}})$$

where $\mathbb{1}_A = \mathbb{1}_{\mathcal{H}_A \otimes \mathcal{H}_{A,x_c}}$ and similar for $\mathbb{1}_B$. When this does not occur, we can still assume such Hilbert spaces and projective measurements so that

$$\mathop{\mathbb{E}}_{T_c \sim D_{x_c}} \sum_{a \in \mathbb{F}_2^{C_I}} \left\| (M_a^{x_c,T_c} \otimes \mathbb{1}_{\mathcal{H}_{A,x_c}} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes G^{x_c}_{[w|_{T_c}=a]}) |\psi\rangle \otimes \left|aux_{x_c}\right\rangle \right\|^2 \leq O(1).$$

When averaging over c and x_c,

$$\mathop{\mathbb{E}}_{c,x_c} \mathop{\mathbb{E}}_{T_c \sim D_{x_c}} \sum_{a \in \mathbb{F}_2^{C_I}} \left\| (M_a^{x_c,T_c} \otimes \mathbb{1}_{A,x_c} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes G^{x_c}_{[w|_{T_c}=a]}) |\psi\rangle \otimes \left|aux_{x_c}\right\rangle \right\|^2 \leq O(C_I^3 \varepsilon^{1/4}).$$

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 $\left\{H_w^{x_0,x_1}\right\}$ on $\mathcal{H} \otimes \mathcal{H}_{x_0,x_1}$ such that

$$\mathop{\mathbb{E}}_{x_0,x_1} \mathop{\mathbb{E}}_{U \sim D_{(x_0,x_1)}} \sum_{a \in \mathbb{F}_2^{C_J}} \left\| (M_a^{x_0,x_1,U} \otimes \mathbb{1}_{\mathcal{H}_{A,x_0,x_1}} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes H^{x_0,x_1}_{[w|_U=a]}) |\psi\rangle \otimes \left|aux_{x_0,x_1}\right\rangle \right\|^2 \leq O(C_J^3 \varepsilon^{1/4}).$$

The next step is ensuring the G and H measurements act on the same Hilbert space. Let

$$\left|\tilde{\psi}\right\rangle = |\psi\rangle \otimes (\otimes_x |aux_x\rangle) \otimes (\otimes_{x_0,x_1} \left|aux_{x_0,x_1}\right\rangle)$$

and

$$\tilde{G}_u^{x_c} = G_u^{x_c} \otimes (\otimes_{x \neq x_c} \mathbb{1}_{\mathcal{H}_x}) \otimes (\otimes_{x_0,x_1} \mathbb{1}_{\mathcal{H}_{x_0,x_1}})$$

$$\tilde{H}_u^{x_0,x_1} = H^{x_0,x_1} \otimes (\otimes_x \mathbb{1}_{\mathcal{H}_x}) \otimes (\otimes_{(z_0,z_1)\neq(x_0,x_1)} \mathbb{1}_{\mathcal{H}_{z_0,z_1}}),$$

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

$$\begin{aligned}
&\mathop{\mathbb{E}}_{T_c\sim D_{x_c}} \sum_{a\in\mathbb{F}_2^{C_I}} \|(N_a^{x_c,T_c} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes \tilde{G}^{x_c}_{[w|_{T_c}=a]}) \left|\tilde{\psi}\right\rangle\|^2 \\
&= \mathop{\mathbb{E}}_{T_c\sim D_{x_c}} \sum_{a\in\mathbb{F}_2^{C_I}} \|(M_a^{x_c,T_c} \otimes \mathbb{1}_{\mathcal{H}_{A,x_c}} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes G^{x_c}_{[w|_{T_c}=a]}) |\psi\rangle \otimes \left|aux_{x_c}\right\rangle\|^2.
\end{aligned}$$

Thus

$$\mathop{\mathbb{E}}_{c,x_c} \mathop{\mathbb{E}}_{T_c\sim D_{x_c}} \sum_{a\in\mathbb{F}_2^{C_I}} \|(N_a^{x_c,T_c} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes \tilde{G}^{x_c}_{[w|_{T_c}=a]}) \left|\tilde{\psi}\right\rangle\|^2 \leq O(C_I^3 \varepsilon^{1/4}), \tag{26}$$

and

$$\mathop{\mathbb{E}}_{x_0,x_1} \mathop{\mathbb{E}}_{U\sim D_{(x_0,x_1)}} \sum_{a\in\mathbb{F}_2^{C_J}} \|(N_a^{x_0,x_1,U} \otimes \mathbb{1}_B - \mathbb{1}_A \otimes \tilde{H}^{x_0,x_1}_{[w|_U=a]}) \left|\tilde{\psi}\right\rangle\|^2 \leq O(C_J^3 \varepsilon^{1/4}). \tag{27}$$

Note these relations also hold with the two systems flipped.

Similarly, passing the **Cross Checks** implies that

$$N_{a_0}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_0}^{x_0,T_0} \tag{28}$$

$$N_{a_1}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_1}^{x_1,T_1} \tag{29}$$

$$N_{a_0}^{x_0,T_0} \otimes \mathbb{1} \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_0}^{x_0,T_0} \tag{30}$$

$$N_{a_1}^{x_1,T_1} \otimes \mathbb{1} \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_1}^{x_1,T_1} \tag{31}$$

$$N_{a_2}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_2}^{x_0,x_1,U} \tag{32}$$

$$N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(\varepsilon)} \mathbb{1}_A \otimes N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U}, \tag{33}$$

with respect to $\left|\tilde{\psi}\right\rangle$ over the distribution D of x_0, x_1, T_0, T_1, U. Note that here we use the $\approx$ notation introduced at the beginning of Section 2.5 to make the dependence on the distribution implicit. These equations combined with Eqs. (26) and (27) 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 2.37 with respect to $\left|\tilde{\psi}\right\rangle$ and distribution D. 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 2.37. 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 $\eta_H^\kappa = 1/2^\kappa$. Letting $C_0 = \max\{C_I, C_J\}$, Hence we can applying Lemma 2.37 to this POVM with $k = 3$, $\delta := C_0^3 \varepsilon^{1/4}$ and $\varepsilon := 1/2^\kappa$, and get that

$$N_{a_0,a_1,a_2}^{x_0,x_1,T_0,T_1,U} \otimes \mathbb{1}_B \approx_{O(C_0^{3/16}\varepsilon^{1/64} + \frac{1}{2^{\kappa/8}})} \mathbb{1}_A \otimes \Lambda^{x_0,x_1}_{[u_0|_{T_0},u_1|_{T_1},w|_U=a_0,a_1,a_2]} \tag{34}$$

with respect to $\left|\tilde{\psi}\right\rangle$ and D, 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 (34) and Lemma 2.32 implies that $\left\{\Lambda^{x_0,x_1}_{u_0,u_1,w}\right\}$ can be used to pass the verify test with probability $1-10\varepsilon-O(C_0^{1/8}\varepsilon^{1/128}+\frac{1}{2^{\kappa/16}})$ where we upper bound $C_0^{3/16}$ by $C_0^{1/4}$. 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 $\left\{\Lambda^{x_0,x_1}_{u_0,u_1}:=\sum_w\Lambda^{x_0,x_1}_{u_0,u_1,w}\right\}$ Let

$$p:=\mathop{\mathbb{E}}_{x_0,x_1}\sum_{u_0,u_1:V(\mathsf{input},x_0,x_1,u_0,u_1)=1}\left\langle\tilde{\psi}\right|\mathbb{1}_A\otimes\Lambda^{x_0,x_1}_{u_0,u_1}\left|\tilde{\psi}\right\rangle,$$

which is the probability that measuring with $\Lambda^{x_0,x_1}_{u_0,u_1}$ gives answers u_0 and u_1 accepted by the verifier V when the questions are x_0 and x_1. Then

$$\begin{aligned}
p&=\mathop{\mathbb{E}}_{x_0,x_1}\sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=1}}\sum_w\left\langle\tilde{\psi}\right|\mathbb{1}_A\otimes\Lambda^{x_0,x_1}_{u_0,u_1,w}\left|\tilde{\psi}\right\rangle\\
&\geq\mathop{\mathbb{E}}_{x_0,x_1}\sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=1}}\sum_w\left\langle\tilde{\psi}\right|\mathbb{1}_A\otimes\Lambda^{x_0,x_1}_{u_0,u_1,w}\left|\tilde{\psi}\right\rangle\cdot\Pr_R[V^{u_0,u_1,w}_{\mathrm{PCPP}}(\mathsf{input},x_0,x_1,2\cdot2^{\ell_1};R)=1]\\
&=\Pr[(\left|\tilde{\psi}\right\rangle,\Lambda)\text{ pass }\textbf{verify}\text{ check }]\\
&\quad-\sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=0}}\sum_w\left\langle\tilde{\psi}\right|\mathbb{1}_A\otimes\Lambda^{x_0,x_1}_{u_0,u_1,w}\left|\tilde{\psi}\right\rangle\cdot\Pr_R[V^{u_0,u_1,w}_{\mathrm{PCPP}}(\mathsf{input},x_0,x_1,2\cdot2^{\ell_1};R)=1]\\
&\geq1-10\varepsilon-O(C_0^{1/8}\varepsilon^{1/128}+\frac{1}{2^{\kappa/16}})\\
&\quad-\sum_{\substack{u_0,u_1:\\ V(\mathsf{input},x_0,x_1,u_0,u_1)=0}}\sum_w\left\langle\tilde{\psi}\right|\mathbb{1}_A\otimes\Lambda^{x_0,x_1}_{u_0,u_1,w}\left|\tilde{\psi}\right\rangle\cdot\Pr_R[V^{u_0,u_1,w}_{\mathrm{PCPP}}(\mathsf{input},x_0,x_1,2\cdot2^{\ell_1};R)=1]\\
&\geq1-10\varepsilon-O(C_0^{1/8}\varepsilon^{1/128}+\frac{1}{2^{\kappa/16}})-(1-p)s,
\end{aligned}$$

where s is the soundness of V_{PCPP}. In the derivation above, $\Pr_R[V^{u_0,u_1,w}_{\mathrm{PCPP}}(\mathsf{input},x_0,x_1,2\cdot2^{\ell_1};R)=1]$ 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 [47, Proposition 17.8]. Hence

$$p\geq\frac{1-10\varepsilon-O(C_0^{1/8}\varepsilon^{1/128}+\frac{1}{2^{\kappa/16}})-s}{1-s}=1-\frac{10\varepsilon+O(C_0^{1/8}\varepsilon^{1/128}+\frac{1}{2^{\kappa/16}})}{1-s}.$$

In the end, we use $(\left\{\tilde{G}^x_u\right\},\left|\tilde{\psi}\right\rangle)$ as a strategy for V. Applying Lemma 2.35 to Eqs. (26), (30) and (31), we get that

$$\tilde{G}^{x_0}_{u|_{T_0}=a}\otimes\mathbb{1}\approx_{O(C_0^3\varepsilon^{1/4})}\mathbb{1}\otimes\tilde{G}^{x_0}_{u|_{T_0}=a}$$

with respect to the distribution of x_0 and the distribution of T_0 determined by x_0 on the state $\left|\tilde{\psi}\right\rangle$. Since $\left\{\tilde{G}^{x_0}_u\right\}$ is a projective measurement, we know

$$\mathop{\mathbb{E}}_{x_0}\mathop{\mathbb{E}}_{T_0\sim D_{x_0}}\sum_a\left\langle\tilde{\psi}\right|\tilde{G}^{x_0}_{u|_{T_0}=a}\otimes\tilde{G}^{x_0}_{u|_{T_0}=a}\left|\tilde{\psi}\right\rangle\geq1-O(C_0^3\varepsilon^{1/4}).$$

On the other hand

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_0}\mathop{\mathbb{E}}_{T_0\sim D_{x_0}}\sum_a \left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{u|_{T_0}=a}\otimes\tilde{G}^{x_0}_{u|_{T_0}=a}\left|\tilde{\psi}\right\rangle\\
&=\mathop{\mathbb{E}}_{x_0}\sum_u \left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{u}\otimes\tilde{G}^{x_0}_{u}\left|\tilde{\psi}\right\rangle + \mathop{\mathbb{E}}_{x_0}\mathop{\mathbb{E}}_{T_0\sim D_{x_0}}\sum_{u\neq u':u|_{T_0}=u'|_{T_0}} \left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{u}\otimes\tilde{G}^{x_0}_{u'}\left|\tilde{\psi}\right\rangle\\
&=\mathop{\mathbb{E}}_{x_0}\sum_u \left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{u}\otimes\tilde{G}^{x_0}_{u}\left|\tilde{\psi}\right\rangle + \mathop{\mathbb{E}}_{x_0}\mathop{\mathbb{E}}_{T_0\sim D_{x_0}}\sum_{u\neq u'} \mathbb{1}[u|_{T_0}=u'|_{T_0}]\left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{u}\otimes\tilde{G}^{x_0}_{u'}\left|\tilde{\psi}\right\rangle.
\end{aligned}$$

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

$$\mathop{\mathbb{E}}_{x_0}\sum_u \left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{u}\otimes\tilde{G}^{x_0}_{u}\left|\tilde{\psi}\right\rangle \geq 1-1/2^\kappa - O(C_0^3\varepsilon^{1/4}).$$

Again, because $\{\tilde{G}^{x_0}_u\}$ is a projective measurement

$$\mathop{\mathbb{E}}_{x_0}\sum_u \|(\tilde{G}^{x_0}_u\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}^{x_0}_u)\left|\tilde{\psi}\right\rangle\|^2 \leq \frac{1}{2^{\kappa-1}} + O(C_0^3\varepsilon^{1/4}).$$

Let $S(x_0,x_1)=\{(a_0,a_1)\mid V(x_0,x_1,a_0,a_1)=1\}$. We can calculate

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{a_0}\otimes\tilde{G}^{x_1}_{a_1}\left|\tilde{\psi}\right\rangle - \left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{a_0}\otimes\tilde{G}^{x_1}_{a_1}\tilde{G}^{x_0}_{a_0}\left|\tilde{\psi}\right\rangle|\\
&\leq \sqrt{\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\|\tilde{G}^{x_0}_{a_0}\otimes\tilde{G}^{x_1}_{a_1}\left|\tilde{\psi}\right\rangle\|^2}\\
&\cdot\sqrt{\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\left\langle\tilde{\psi}\right|(\tilde{G}^{x_0}_{a_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}^{x_0}_{a_0})(\mathbb{1}\otimes\tilde{G}^{x_1}_{a_1})(\tilde{G}^{x_0}_{a_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}^{x_0}_{a_0})\left|\tilde{\psi}\right\rangle}\\
&\leq 1\cdot\sqrt{\mathop{\mathbb{E}}_{x_0}\sum_{a_0}\|(\tilde{G}^{x_0}_{a_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}^{x_0}_{a_0})\left|\tilde{\psi}\right\rangle\|^2}\\
&\leq O(\frac{1}{2^{\kappa/2}}+C_0^{3/2}\varepsilon^{1/8}),
\end{aligned}$$

and

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\left\langle\tilde{\psi}\right| \mathbb{1}\otimes\tilde{G}^{x_0}_{a_0}\tilde{G}^{x_1}_{a_1}\tilde{G}^{x_0}_{a_0}\left|\tilde{\psi}\right\rangle - \left\langle\tilde{\psi}\right| \tilde{G}^{x_0}_{a_0}\otimes\tilde{G}^{x_1}_{a_1}\tilde{G}^{x_0}_{a_0}\left|\tilde{\psi}\right\rangle|\\
&\leq \sqrt{\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\|\mathbb{1}\otimes\tilde{G}^{x_1}_{a_1}\tilde{G}^{x_0}_{a_0}\left|\tilde{\psi}\right\rangle\|^2}\\
&\cdot\sqrt{\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\left\langle\tilde{\psi}\right|(\tilde{G}^{x_0}_{a_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}^{x_0}_{a_0})(\mathbb{1}\otimes\tilde{G}^{x_1}_{a_1})(\tilde{G}^{x_0}_{a_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}^{x_0}_{a_0})\left|\tilde{\psi}\right\rangle}\\
&\leq 1\cdot\sqrt{\mathop{\mathbb{E}}_{x_0}\sum_{a_0}\|(\tilde{G}^{x_0}_{a_0}\otimes\mathbb{1}-\mathbb{1}\otimes\tilde{G}^{x_0}_{a_0})\left|\tilde{\psi}\right\rangle\|^2}\\
&\leq O(\frac{1}{2^{\kappa/2}}+C_0^{3/2}\varepsilon^{1/8}).
\end{aligned}$$

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,

$$|\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\left\langle\tilde{\psi}\right|(\tilde{G}_{a_0}^{x_0}\otimes\tilde{G}_{a_1}^{x_1}-\mathbb{1}\otimes\Lambda_{a_0,a_1}^{x_0,x_1})\left|\tilde{\psi}\right\rangle| \le O(\frac{1}{2^{\kappa/2}}+C_0^{3/2}\varepsilon^{1/8}).$$

On the other hand, we have shown

$$\mathop{\mathbb{E}}_{x_0,x_1}\sum_{(a_0,a_1)\in S}\left\langle\tilde{\psi}\right|\mathbb{1}\otimes\Lambda_{a_0,a_1}^{x_0,x_1})\left|\tilde{\psi}\right\rangle = p \ge 1-O(C_0^{1/8}\varepsilon^{1/128}+\frac{1}{2^{\kappa/16}}).$$

Since the big-O notations in the derivations above hide constants that are independent of κ, the winning probability of the strategy $(\{\tilde{G}_u^x\}, \left|\tilde{\psi}\right\rangle)$ for V is at least $1-\frac{C_1}{2^{\kappa/16}}-C_2C_0^{3/2}\varepsilon^{1/128}$ for some constants C_1 and C_2. Hence, $K_1 = \frac{C_1}{2^{\kappa/16}}$ and $K_2 = C_2C_0^{3/2}$ in the soundness statement. We need to pick κ large enough such that $1-\frac{C_1}{2^{\kappa/16}} > s$, then we can solve for

$$\varepsilon_0 = \left[\frac{1-\frac{C_1}{2^{\kappa/16}}-s}{C_2C_0^{3/2}}\right]^{128},$$

such that $1-\varepsilon_0$ is the soundness of V^{AR}. □

6.4 Putting everything together

Theorem 6.11. RE *is contained in* MIP*[poly, $O(1)$] *with completeness* 1 *and a constant soundness.*

Proof. We know that there is an MIP*[poly, poly] protocol $V = (\mathrm{Alg}_Q, \mathrm{Alg}_A)$ for any language in RE. To prove the theorem, we construct an MIP*[poly, $O(1)$] protocol for the same language, by applying several answer reduction transformations to V. Specifically, we apply two iterations of an answer reduction scheme based on the low-degree test over finite fields to reduce the answer size to $O(\mathrm{poly}\log\log(n))$, followed by one iteration of answer reduction based on the Hadamard code over $\mathbb{F}_2$ to further reduce the answer size to $O(1)$. The effect of each step is summarized in the following table.

	Question size	**Answer size**	**Sampling time**	**Verification time**	**Decision complexity**
Original protocol	poly(n)	poly(n)	poly(n)	poly(n)	poly(n)
After Step 1 (Theorem 6.2)	poly(n)	polylog(n)	poly(n)	poly(n)	polylog(n)
After Step 2 (Theorem 6.2)	poly(n)	polyloglog(n)	poly(n)	poly(n)	polyloglog(n)
After Step 3 (Theorem 6.9)	poly(n)	$O(1)$	poly(n)	poly(n)	O(1)

Table 1. Effect of each step of the proof

Step 1. We parallel repeat and oracularize V to ensure its soundness is at most 1/2, and apply the answer reduction technique summarized in Theorem 6.2. Since parallel repetition and oracularization only introduce constant overhead in the question length, answer length, sampling time and verification time, we still use V to denote the oracularized protocol. Denote the answer-reduced protocol by $V_1 = (\mathrm{Alg}_{Q_1}, \mathrm{Alg}_{A_1})$. Since $d_{V,A}(n) = \mathrm{poly}(n)$ and $k(n) = \mathrm{polylog}(d_{V,A}(n)) = \mathrm{polylog}(n)$, by Theorem 6.2, the new question length is

$$\ell_{V_1,Q}(n) = \mathrm{polylog}(d_{V,A}(n))\cdot(2\ell_{V,Q}(n)+\mathrm{polylog}(d_{V,A}(n))) = \mathrm{poly}(n).$$

The new answer length and decision complexity are

$$\ell_{V_1,A}(n) = d_{V_1,A}(n) = \text{polylog}(d_{V,A}(n)) \cdot \text{polylog}(d_{V,A}(n)) = \text{polylog}(n).$$

The new sampling time is

$$t_{V_1,Q}(n) = \text{polylog}(d_{V,A}(n)) \cdot (t_{V,Q}(n) + \text{polylog}(d_{V,A}(n))) = \text{poly}(n).$$

The new verification time is

$$t_{V_1,A}(n) = \text{polylog}(d_{V,A}(n)) \cdot (t_{V,Q}(n) + t_{V,A}(n) + \text{polylog}(d_{V,A}(n))) = \text{poly}(n).$$

Lastly, V has completeness 1 and soundness at most $1/2$, so is V_1.

Step 2. We apply Theorem 6.2 again and denote the new protocol by $V_2 = (\text{Alg}_{Q_2}, \text{Alg}_{A_2})$. Since $d_{V_1,A}(n) = \text{polylog}(n)$ and $k(n) = \text{polylog}(d_{V_1,A}(n)) = \text{polyloglog}(n)$, by Theorem 6.2, the new question length is

$$\ell_{V_2,Q}(n) = \text{polylog}(d_{V_1,A}(n)) \cdot (2\ell_{V_1,Q}(n) + \text{polylog}(d_{V_1,A}(n))) = \text{poly}(n).$$

The new answer length and decision complexity are

$$\ell_{V_2,A}(n) = d_{V_2,A}(n) = \text{polylog}(d_{V_1,A}(n)) \cdot \text{polylog}(d_{V_1,A}(n)) = \text{polyloglog}(n).$$

The new sampling time is

$$t_{V_2,Q}(n) = \text{polylog}(d_{V_1,A}(n)) \cdot (t_{V_1,Q}(n) + \text{polylog}(d_{V_1,A}(n))) = \text{poly}(n).$$

The new verification time is

$$t_{V_2,A}(n) = \text{polylog}(d_{V_1,A}(n)) \cdot (t_{V_1,Q}(n) + t_{V_1,A}(n) + \text{polylog}(d_{V_1,A}(n))) = \text{poly}(n).$$

Moreover, V_2 has perfect completeness and soundness at most $1/2$.

Step 3. The protocol V_2 is oracularized again. Again, since oracularization only introduces constant overhead in the question length, answer length, sampling time and verification time, we still use V_2 to denote the oracularized protocol. Define the language L_{Enc} as in Definition 6.7 for V_2. We can calculate the parameters in Definition 6.8. First, $L_{\mathsf{Enc}} \in \text{DTIME}(T(n))$ where

$$T(n) = t_{V_2,A}(n) + 2^{\ell_{V_2,A}(n)} = \text{poly}(n),$$

where the decoder $\text{Dec}_{\ell_{V_2,A}(n)}$ takes $O(2^{\text{polyloglog}(n)})$ time. Moreover, the query lengths are

$$\ell_1(n) = \ell_{V_2,A}(n) = \text{polyloglog}(n) \qquad \ell_2(n) = \ell_\pi(n) = T(n)\log_2(T(n)) = \text{poly}(n).$$

Then the PCPP question sampling and verification times are

$$t_{\text{PCPP},Q}(n) = t_{V_2,Q}(n) + \text{poly}(n + \ell_{V_2,Q}(n), \ell_{V_2,A}(n), \log_2(T(n))) + O(\ell_1(n) + \ell_2(n)) = \text{poly}(n),$$
$$t_{\text{PCPP},A}(n) = \text{poly}(n + \ell_{V_2,Q}(n), \ell_{V_2,A}(n), \log_2(T(n))) = \text{poly}(n).$$

Next, we apply the answer reduction technique in Theorem 6.9 to get verifier V^{AR}. By Theorem 6.9,

$$\ell_{V^{AR},Q}(n) = O(\ell_{V_2,Q}(n) + \ell_1(n) + \ell_2(n)) = \text{poly}(n),$$
$$\ell_{V^{AR},A}(n) = O(1),$$
$$t_{V^{AR},Q}(n) = O(t_{V_2,Q}(n) + \ell_1(n) + \ell_2(n) + t_{\text{PCPP},Q}(n)) = \text{poly}(n),$$
$$t_{V^{AR},A}(n) = O(t_{\text{PCPP},A}(n)) = \text{poly}(n).$$

Moreover, it is easy to see that the new decision complexity is $d_{V^{AR},A}(n) = O(1)$. The perfect completeness and constant soundness of V^{AR} follow from Theorem 6.9.

Alternatively, we can apply the answer reduction technique of Theorem 6.2 iteratively until the answer size is constant. The proof follows the same line of argument in the proof of [48, Theorem 54], so we only sketch the

proof here. The sampler Alg_Q and decider Alg_A both start by calculating the description of $V_0 = (\mathsf{Alg}_{Q_0}, \mathsf{Alg}_{A_0})$, which is an MIP*[poly, poly] protocol for RE, then repeatedly applying the answer reduction procedure from Theorem 6.2 followed by parallel repetition and oracularization to calculate the description of V_{i+1} from the description of V_i for $i \geq 0$ until V_m has answer size $O(1)$. Then Alg_Q executes Alg_{Q_m} to sample the questions. When the answers are returned, the decider Alg_A executes Alg_{A_m} to check the answers.

Following the same analysis, we can get $m = O(\log\log\log(\ell_{V,A}(n)))$. Besides the $O(1)$ answer size and decision complexity, the question size, sampling time, and verification time of V_m are:

Question size. The question size follows the recursive relation

$$\ell_{V_{i+1},Q}(n) = \mathrm{polylog}(d_{V_i,A}(n))(2\ell_{V_i,Q}(n) + \mathrm{polylog}(d_{V_i,A}(n))).$$

Since $d_{V_{i+1},A}(n) = \mathrm{polylog}(d_{V_i,A}(n))$, we can bound

$$\begin{aligned}
\ell_{V_m,Q}(n) &= \mathrm{polylog}(d_{V_{m-1},A}(n))(2\ell_{V_{m-1},Q}(n) + \mathrm{polylog}(d_{V_{m-1},A}(n))) \\
&= 2^m \cdot \prod_{i=0}^{m-1}[\mathrm{polylog}(d_{V_i,A}(n))] \cdot \ell_{V_0,Q}(n) + \sum_{i=0}^{m-1} 2^{m-1-i} \prod_{j=i}^{m-1}[\mathrm{polylog}(d_{V_j,A}(n))]\,\mathrm{polylog}(d_{V_i,A}(n)) \\
&\leq 2^m\,\mathrm{polylog}(n)\ell_{V_0,Q}(n) + 2^m m\,\mathrm{polylog}(n)\,\mathrm{polylog}(d_{V_0,A}(n))) \\
&= O(\mathrm{polylog}(n)\,\mathrm{poly}(n) + \mathrm{polylog}(n)) = O(\mathrm{poly}(n)),
\end{aligned}$$

where we upper bound $\prod_{i=0}^{m-1}[\mathrm{polylog}(d_{V_i,A}(n))] = \mathrm{polylog}(n)\,\mathrm{polyloglog}(n)\ldots O(1)$ by $\mathrm{polylog}(n)\cdot\mathrm{polyloglog}(n)^m = \mathrm{polylog}(n)$.

Sampling time. It takes m iterations for the sampler to calculate V_m. The $(i+1)$th iteration takes time $O(|\mathsf{Alg}_{Q_i}| + |\mathsf{Alg}_{A_i}|) = O(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A| + i)$. Hence, the total computation time is $O(m(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A|) + m^2)$. The running time of Alg_{Q_m} follows the relation

$$\begin{aligned}
t_{V_m,Q}(n) &= \mathrm{polylog}(d_{V_{m-1},A}(n)) \cdot (t_{V_{m-1},Q}(n) + \mathrm{polylog}(d_{V_{m-1},A}(n))) \\
&= \ldots \\
&= \prod_{i=0}^{m-1}[\mathrm{polylog}(d_{V_i,A}(n))] \cdot t_{V_0,Q}(n) + \sum_{i=0}^{m-1}\prod_{j=i}^{m-1}[\mathrm{polylog}(d_{V_j,A}(n))] \cdot \mathrm{polylog}(d_{V_i,A}(n)) \\
&\leq \mathrm{polylog}(n)t_{V_0,Q}(n) + m\,\mathrm{polylog}(n)\,\mathrm{polylog}(d_{V_0,A}(n)) = \mathrm{poly}(n).
\end{aligned}$$

Hence, the total sampling time of V_m is $O(m(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A|) + m^2 + \mathrm{poly}(n)) = \mathrm{poly}(n)$.

Verification time. Similar to the previous case, the time to calculate V_m is $O(m(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A|) + m^2)$. The verification time of V_m is

$$\begin{aligned}
t_{V_m,A}(n) &= \mathrm{polylog}(d_{V_{m-1},A}(n))(t_{V_{m-1},Q}(n) + t_{V_{m-1},A}(n) + \mathrm{polylog}(d_{V_{m-1},A}(n))) \\
&= \ldots \\
&= \prod_{i=0}^{m-1}[\mathrm{polylog}(d_{V_i,A}(n))] \cdot (t_{V_0,Q}(n) + t_{V_0,A}(n)) + \sum_{i=0}^{m-1}\prod_{j=i}^{m-1}[\mathrm{polylog}(d_{V_j,A}(n))] \cdot \mathrm{polylog}(d_{V_i,A}(n)) \\
&\leq \mathrm{polylog}(n)(t_{V_0,Q}(n) + t_{V_0,A}(n)) + m\,\mathrm{polylog}(n)\,\mathrm{polylog}(d_{V_0,A}(n)) = \mathrm{poly}(n).
\end{aligned}$$

Hence the total verification time is $O(m(|\mathsf{Alg}_Q| + |\mathsf{Alg}_A|) + m^2 + \mathrm{poly}(n)) = \mathrm{poly}(n)$. Lastly, the protocol V_m has completeness 1 and soundness at most $1/2$.

□

Remark 6.12. We have just shown how to use our new answer reduction techniques to get very small answer sizes, without a large overhead in question length. A natural question that arises is whether this can be applied to the protocol of [48], which has *constant* question length but polylogarithmic answer length, in order to obtain a protocol with total communication that scales as $O(\text{polyloglog}(n))$. This would contradict the lower bound in [48], which shows that RE (or indeed EEXP) cannot be decided by MIP* protocols with total communication smaller than $\log(n)$. Indeed, our tighter answer reduction fails to give such a result when applied to the constant-question-size protocol of [48]. This is because of the phenomenon described in Remark 6.3: an application of question reduction resets the decision complexity to be $\text{poly}(n)$, so in particular, the protocol from that work has decision complexity $\text{poly}(n)$, and applying answer reduction to it would blow up both the question size and answer size to $\text{polylog}(n)$.

References

[1] Dorit Aharonov and Michael Ben-Or. 2008. Fault-Tolerant Quantum Computation with Constant Error Rate. *SIAM J. Comput.* 38, 4 (2008), 1207.

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

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

[4] Rotem Arnon-Friedman, Zvika Brakerski, and Thomas Vidick. 2023. Computational Entanglement Theory. *arXiv preprint arXiv:2310.02783* (2023).

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

[6] Srinivasan Arunachalam, Arkopal Dutt, Francisco Escudero Gutiérrez, and Carlos Palazuelos. 2024. Learning Low-Degree Quantum Objects. In *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 297)*, Karl Bringmann, Martin Grohe, Gabriele Puppis, and Ola Svensson (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 13:1–13:19. doi:10.4230/LIPIcs.ICALP.2024.13

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

[8] László Babai, Lance Fortnow, and Carsten Lund. 1991. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity* 1, 1 (01 Mar 1991), 3–40. doi:10.1007/BF01200056

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

[10] Zongbo Bao and Penghui Yao. 2023. On Testing and Learning Quantum Junta Channels. In *Proceedings of Thirty Sixth Conference on Learning Theory (Proceedings of Machine Learning Research, Vol. 195)*, Gergely Neu and Lorenzo Rosasco (Eds.). PMLR, 1064–1094. https://proceedings.mlr.press/v195/bao23b.html

[11] Salman Beigi. 2013. A new quantum data processing inequality. *J. Math. Phys.* 54, 8 (2013), 082202. arXiv:https://doi.org/10.1063/1.4818985 doi:10.1063/1.4818985

[12] Charles H Bennett, Herbert J Bernstein, Sandu Popescu, and Benjamin Schumacher. 1996. Concentrating partial entanglement by local operations. *Physical Review A* 53, 4 (1996), 2046.

[13] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. 1993. Self-testing/correcting with applications to numerical problems. *J. Comput. System Sci.* 47, 3 (1993), 549–595.

[14] 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. 2023. Logical quantum processor based on reconfigurable atom arrays. *Nature* (2023), 1–3.

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

[16] J. Lawrence Carter and Mark N. Wegman. 1977. Universal Classes of Hash Functions (Extended Abstract). In *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing* (Boulder, Colorado, USA) *(STOC 1977)*. Association for Computing Machinery, New York, NY, USA, 106–112. doi:10.1145/800105.803400

[17] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. 2023. The complexity of NISQ. *Nature Communications* 14, 1 (26 Sep 2023), 6001. doi:10.1038/s41467-023-41217-6

[18] Thomas Chen, Shivam Nadimpalli, and Henry Yuen. 2023. Testing and Learning Quantum Juntas Nearly Optimally. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 1163–1185. arXiv:https://epubs.siam.org/doi/pdf/10.1137/1.9781611977554.ch43 doi:10.1137/1.9781611977554.ch43

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

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

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

[22] Honghao Fu. 2022. Constant-sized correlations are sufficient to self-test maximally entangled states with unbounded dimension. *Quantum* 6 (Jan. 2022), 614. doi:10.22331/q-2022-01-03-614

[23] Badih Ghazi, Pritish Kamath, and Prasad Raghavendra. 2018. Dimension Reduction for Polynomials over Gaussian Space and Applications. In *Proceedings of the 33rd Computational Complexity Conference* (San Diego, California) *(CCC '18)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Germany, Article 28, 37 pages. doi:10.4230/LIPIcs.CCC.2018.28

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

[25] Prahladh Harsha, Adam Klivans, and Raghu Meka. 2013. An Invariance Principle for Polytopes. *J. ACM* 59, 6, Article 29 (Jan 2013), 25 pages. doi:10.1145/2395116.2395118

[26] Johan Håstad. 2001. Some Optimal Inapproximability Results. *J. ACM* 48, 4 (Jul 2001), 798–859. doi:10.1145/502090.502098

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

[28] Tsuyoshi Ito, Hirotada Kobayashi, and Keiji Matsumoto. 2009. 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)*. IEEE Computer Society, Washington, DC, USA, 217–228. doi:10.1109/CCC.2009.22

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

[30] Zhengfeng Ji. 2017. Compression of Quantum Multi-prover Interactive Proofs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* (Montreal, Canada) *(STOC 2017)*. ACM, New York, NY, USA, 289–302. doi:10.1145/3055399.3055441

[31] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. 2020. MIP* = RE. *arXiv preprint arXiv:2001.04383* (2020).

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

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

[34] Zander Kelley and Raghu Meka. 2022. Random restrictions and PRGs for PTFs in Gaussian space. In *Proceedings of the 37th Computational Complexity Conference* (Philadelphia, Pennsylvania) *(CCC '22)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, DEU, Article 21, 24 pages. doi:10.4230/LIPIcs.CCC.2022.21

[35] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, Ben Toner, and Thomas Vidick. 2011. Entangled Games Are Hard to Approximate. *SIAM J. Comput.* 40, 3 (2011), 848–877. arXiv:https://doi.org/10.1137/090751293 doi:10.1137/090751293

[36] Julia Kempe, Oded Regev, and Ben Toner. 2010. Unique Games with Entangled Provers Are Easy. *SIAM J. Comput.* 39, 7 (2010), 3207–3229. arXiv:https://doi.org/10.1137/090772885 doi:10.1137/090772885

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

[38] Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. 2024. Quantum and Classical Low-Degree Learning via a Dimension-Free Remez Inequality. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 287)*, Venkatesan Guruswami (Ed.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 69:1–69:22. doi:10.4230/LIPIcs.ITCS.2024.69

[39] Robert Krauthgamer and Ori Sasson. 2003. Property Testing of Data Dimensionality. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (Baltimore, Maryland) *(SODA 2003)*. Society for Industrial and Applied Mathematics, USA, 18–27.

[40] Raghu Meka and David Zuckerman. 2010. Pseudorandom Generators for Polynomial Threshold Functions. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing* (Cambridge, Massachusetts, USA) *(STOC 2010)*. Association for Computing Machinery, New York, NY, USA, 427–436. doi:10.1145/1806689.1806749

[41] Antonio Anna Mele, Armando Angrisani, Soumik Ghosh, Sumeet Khatri, Jens Eisert, Daniel Stilck França, and Yihui Quek. 2024. Noise-induced shallow circuits and absence of barren plateaus. *arXiv preprint arXiv:2403.13927* (2024).
[42] Thilo Mie. 2009. Short PCPPs verifiable in polylogarithmic time with $O(1)$ queries. *Annals of Mathematics and Artificial Intelligence* 56, 3 (2009), 313–338. doi:10.1007/s10472-009-9169-y
[43] Ashley Montanaro and Tobias J. Osborne. 2010. Quantum Boolean Functions. *Chicago Journal of Theoretical Computer Science* 2010, 1 (January 2010).
[44] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. 2005. Noise stability of functions with low influences: invariance and optimality. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005)*. IEEE, 21–30.
[45] Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. 2024. On the Pauli Spectrum of QAC0. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (Vancouver, BC, Canada) *(STOC 2024)*. Association for Computing Machinery, New York, NY, USA, 1498–1506. doi:10.1145/3618260.3649662
[46] Anand Natarajan and Thomas Vidick. 2017. A quantum linearity test for robustly verifying entanglement. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2017)*. 1003–1015.
[47] Anand Natarajan and John Wright. 2019. NEEXP is Contained in MIP*. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS 2019)*. IEEE, 510–518. doi:10.1109/FOCS.2019.00039
[48] Anand Natarajan and Tina Zhang. 2023. Quantum Free Games. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (Orlando, FL, USA) *(STOC 2023)*. Association for Computing Machinery, New York, NY, USA, 1603–1616. doi:10.1145/3564246.3585208
[49] Deanna Needell, William Swartworth, and David P. Woodruff. 2022. Testing Positive Semidefiniteness Using Linear Measurements. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS 2022)*. 87–97. doi:10.1109/FOCS54457.2022.00016
[50] Ryan O'Donnell. 2013. *Analysis of Boolean Functions*. Cambridge University Press, Cambridge, UK.
[51] Ryan O'Donnell, Rocco A. Servedio, and Li-Yang Tan. 2020. Fooling Gaussian PTFs via local hyperconcentration. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* (Chicago, IL, USA) *(STOC 2020)*. Association for Computing Machinery, New York, NY, USA, 1170–1183. doi:10.1145/3357713.3384281
[52] Ryan O'Donnell, Rocco A. Servedio, and Li-Yang Tan. 2022. Fooling Polytopes. *J. ACM* 69, 2, Article 9 (Jan 2022), 37 pages. doi:10.1145/3460532
[53] Minglong Qin and Penghui Yao. 2021. Nonlocal games with noisy maximally entangled states are decidable. *SIAM J. Comput.* 50, 6 (2021), 1800–1891.
[54] Minglong Qin and Penghui Yao. 2023. Decidability of Fully Quantum Nonlocal Games with Noisy Maximally Entangled States. In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 261)*, Kousha Etessami, Uriel Feige, and Gabriele Puppis (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 97:1–97:20. doi:10.4230/LIPIcs.ICALP.2023.97
[55] Oded Regev and Liron Schiff. 2008. Impossibility of a quantum speed-up with a faulty oracle. In *International Colloquium on Automata, Languages, and Programming*. Springer, 773–781.
[56] Ben W. Reichardt, Falk Unger, and Umesh Vazirani. 2013. Classical command of quantum systems. *Nature* 496, 7446 (01 Apr 2013), 456–460. doi:10.1038/nature12035
[57] Cambyse Rouzé, Melchior Wirth, and Haonan Zhang. 2024. Quantum Talagrand, KKL and Friedgut's Theorems and the Learnability of Quantum Boolean Functions. *Communications in Mathematical Physics* 405, 4 (09 Apr 2024), 95. doi:10.1007/s00220-024-04981-0
[58] Hristo S. Sendov. 2007. The higher-order derivatives of spectral functions. *Linear Algebra Appl.* 424, 1 (2007), 240–281. doi:10.1016/j.laa.2006.12.013 Special Issue in honor of Roger Horn.
[59] Victor Shoup. 1990. New algorithms for finding irreducible polynomials over finite fields. *Mathematics of computation* 54, 189 (1990), 435–447.
[60] Anna Skripka and Anna Tomskova. 2019. *Multilinear operator integrals*. Springer.
[61] William Slofstra. 2019. THE SET OF QUANTUM CORRELATIONS IS NOT CLOSED. *Forum of Mathematics, Pi* 7 (2019), e1. doi:10.1017/fmp.2018.3
[62] William Slofstra. 2020. Tsirelson's problem and an embedding theorem for groups arising from non-local games. *Journal of the American Mathematical Society* 33 (2020), 1–56. arXiv:https://doi.org/10.1090/jams/929 doi:/10.1090/jams/929
[63] Joseph Slote, Alexander Volberg, and Haonan Zhang. 2024. Noncommutative Bohnenblust–Hille Inequality for qudit systems. https://arxiv.org/abs/2406.08509. arXiv:2406.08509 [math.FA] https://arxiv.org/abs/2406.08509
[64] Salil P. Vadhan. 2012. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science* 7, 1–3 (2012), 1–336. doi:10.1561/0400000010
[65] Guoming Wang. 2011. Property testing of unitary operators. *Phys. Rev. A* 84 (Nov 2011), 052328. Issue 5. doi:10.1103/PhysRevA.84.052328

A Lemmas for Noisy MIP*

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

Lemma A.1. *[53, Lemma 6.1][8] Given parameters $0 \le \rho < 1$, $0 < \delta < 1$, $n, m \in \mathbb{Z}_{>0}$, $m \ge 2$, and an m-dimensional noisy MES ψ_{AB} with the quantum 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 $|||P|||_2 \le 1$ and $|||Q|||_2 \le 1$. 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) $\left\|P^{(1)}\right\|_2 \le 1$ and $\left\|Q^{(1)}\right\|_2 \le 1$.
(3) $\left|\mathrm{Tr}\left(\left(P^{(1)} \otimes Q^{(1)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left((P \otimes Q)\psi_{AB}^{\otimes n}\right)\right| \le \delta$.
(4) $\frac{1}{m^n}\mathrm{Tr}\,\zeta(P^{(1)}) \le \delta$ and $\frac{1}{m^n}\mathrm{Tr}\,\zeta(Q^{(1)}) \le \delta$.
(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 A.2. It is easily verified that for the above lemma, for each $\sigma \in [m^2]^n_{\ge 0}$, we have

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

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 A.3. *[53, Lemma 7.4] Given $0 < \tau < 1$, $d, n, m \in \mathbb{Z}_{>0}$, $m \ge 2$, and a degree-d matrix $P \in \mathcal{H}_m^{\otimes n}$ satisfying $|||P|||_2 \le 1$, there exists a subset $H \subseteq [n]$ of size $h = |H| \le \frac{d}{\tau}$ such that for any $i \notin H$,*

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

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 A.4. *Given $\overrightarrow{X} \in \left(\mathcal{H}_m^{\otimes n}\right)^t$ 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

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

Miscellaneous Lemmas. The following lemmas are used throughout Appendix C.

Fact A.5. [53, 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

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

[8]The statement is slightly different from that in [53, Lemma 6.1]. The difference arises due to our relocation of the truncating step, which was in [53, Lemma 10.5].

Lemma A.6. *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* $(x,y,a,b) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B}$,

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

for some $\delta > 0$. *Then*

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

Proof.

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

□

B Deferred Proofs of Section 2.5

Proof of Lemma 2.32. 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. □

Proof of Lemma 2.36. We start with

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

Because $A^x_{a_k}\otimes\mathbb{1}\approx_\delta \mathbb{1}\otimes(B_k)^x_{a_k}$, To apply Lemma 2.34, 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 $b = (a_1,\dots,a_{k-1})$. Then $\sum_b (C^x_{a,b})^\dagger C^x_{a,b}\le\mathbb{1}$. Hence by Lemma 2.34

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

We can apply Lemma 2.34 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 $b = (a_1,\dots,a_{k-2},a_k)$. 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 2.35

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

which completes the proof. □

Proof of Lemma 2.37, 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

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

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

$$\le\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\|A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes(G_2)^x_{g_2}\,|\psi\rangle\|^2}\cdot$$

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

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

$$\le\sqrt{2\delta}$$

and

$$|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\langle\psi|\,A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes(G_2)^x_{g_2}\,|\psi\rangle-\langle\psi|\,A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes\mathbb{1}\,|\psi\rangle|$$

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

$$\le\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\|A^{x,y_1,y_2}_{a_1,a_2}\,|\psi\rangle\|^2}\cdot$$

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

$$\le\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,a_2}\|A^{x,y_1,y_2}_{a_1,a_2}\,|\psi\rangle\|^2}\cdot$$

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

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

$$\le\sqrt{2\delta},$$

Hence, we focus on proving

$$\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\|\mathbb{1}\otimes\left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}-(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}\right)|\psi\rangle\|^2\le C_1\sqrt{\delta}+C_2\varepsilon \tag{35}$$

for some constants C_1 and C_2, which will imply that

$$|\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\langle\psi|\,A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes(G_2)^x_{g_2}\left((G_1)^x_{[g_1(y_1)=a_1]}(G_2)^x_{g_2}-(G_2)^x_{g_2}(G_1)^x_{[g_1(y_1)=a_1]}\right)|\psi\rangle|$$

$$\le\sqrt{\mathop{\mathbb{E}}_{x,y_1,y_2}\sum_{a_1,g_2}\|A^{x,y_1,y_2}_{a_1,g_2(y_2)}\otimes(G_2)^x_{g_2}\,|\psi\rangle\|^2}\cdot$$

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

and

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

To prove Eq. (35), we start with Eq. (9)

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

for $i=1,2$. Then by Lemma 2.34

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

Chaining the inequalities together using Lemma 2.35 gives

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

Let

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

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

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

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

$$\Delta_4 = |\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,g_2} \langle\psi| \mathbb{1} \otimes (G_1)^x_{[g_1(y_1)=a_1]} (G_2)^x_{g_2} (G_1)^x_{[g_1(y_1)=a_1]} (G_2)^x_{g_2} |\psi\rangle$$

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

First of all

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

By Eq. (10),

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

then Lemma 2.32 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

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

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

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

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

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 2.31 to Eqs. (9) and (10) we get

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

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

Notice that for any $i \in [2]$,

$$\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_i} \langle\psi| A^{x,y_1,y_2}_{a_i} \otimes A^{x,y_1,y_2}_{a_i} |\psi\rangle \ge \mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1,a_2} \langle\psi| A^{x,y_1,y_2}_{a_1,a_2} \otimes A^{x,y_1,y_2}_{a_1,a_2} |\psi\rangle \ge 1 - \delta$$

because $A^{x,y_1,y_2}_{a_1,a_2} \otimes A^{x,y_1,y_2}_{b_1,b_2} \ge 0$ for any a_1, a_2, b_1, b_2. Then Lemma 2.31 also implies that

$$\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_1} \|(A^{x,y_1,y_2}_{a_i} \otimes \mathbb{1} - \mathbb{1} \otimes A^{x,y_1,y_2}_{a_i}) |\psi\rangle\|^2 \le 2\delta.$$

Hence, Lemma 2.35 implies that for all $i \in [2]$.

$$\mathop{\mathbb{E}}_{x,y_1,y_2} \sum_{a_i} \| \Big((G_i)^x_{[g_i(y_i)=a_i]} \otimes \mathbb{1} - \mathbb{1} \otimes (G_i)^x_{[g_i(y_i)=a_i]} \Big) |\psi\rangle \|^2 \leq 18\delta.$$

Also, notice that

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

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,

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

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

$$\mathop{\mathbb{E}}_{x,y_1,\ldots,y_{i+1}} \sum_{a_1,\ldots,a_{i+1}} \langle\psi| A^{x,y_1,\ldots,y_{i+1}}_{a_1,\ldots,a_{i+1}} \otimes J^{x,y_1,\ldots,y_{i+1}}_{[g_1(y_1),\ldots,g_{i+1}(y_{i+1})=a_1,\ldots,a_{i+1}]} |\psi\rangle$$

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

by relating it to

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

So the central step is bounding

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

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

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

and

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

Therefore,

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

and

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

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

C Upper Bound on the Number of Noisy MES's for Nonlocal Games

The proof follows closely to that of [53]. The major difference is that in the proof of [53], 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 [54]. Then our new upper bound below only depends polynomially on the size of the question set whereas the previous one has an exponential dependence.

C.1 Gaussian Dimension Reduction

The following lemma is a simplified version of [54, Lemma 5.13], with the questions and answers being classical. In the proof of Theorem 5.5, 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 C.1, where we preserve the expectation of the ζ function value over the random variable $\mathbf{M}$. In the previous version (Item 2 of [54, 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 C.1. *[54, 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 quantum 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

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

(3)

$$\mathop{\mathbb{E}}_{\mathbf{g}}[\mathrm{Tr}\,\zeta\,(P(\mathbf{g}))] = \mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x}}[\mathrm{Tr}\,\zeta\,(P(\mathbf{M}\tilde{\mathbf{x}}))] \quad \text{and} \quad \mathop{\mathbb{E}}_{\mathbf{h}}[\mathrm{Tr}\,\zeta\,(Q(\mathbf{h}))] = \mathop{\mathbb{E}}_{\mathbf{M},\mathbf{y}}[\mathrm{Tr}\,\zeta\,(Q(\mathbf{M}\tilde{\mathbf{y}}))].$$

(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 $F(M) = \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\mathrm{Tr}\left((P(M\tilde{\mathbf{x}}) \otimes Q(M\tilde{\mathbf{y}}))\, \psi_{AB}^{\otimes h}\right)\right]$. To prove Lemma C.1 item 2, we need the following lemma.

Lemma C.2. *In the setting of Lemma C.1, given $d \in \mathbb{Z}_{>0}$, $\delta > 0$, there exists $n_0 = \frac{d^{O(d)}}{\delta^2}$ such that the following holds: For $\mathbf{M} \sim \gamma_{n\times n_0}$,*

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

We use the following lemma to prove Lemma C.2.

Lemma C.3. *[23, 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*

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

- *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*

if $S \triangle T \triangle S' \triangle T' \neq \emptyset$:

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

if $S \triangle T \triangle S' \triangle T' = \emptyset$:

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

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 C.2. Use Lemma C.3 with parameters d and δ, we have

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{M}}[F(\mathbf{M})] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}\left[\mathrm{Tr}\left((P(\mathbf{g}) \otimes Q(\mathbf{h}))\, \psi_{AB}^{\otimes h}\right)\right]\right| \\
&= \left|\sum_{S,T\subseteq[n]}\left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}\left[\mathbf{u}_S \mathbf{v}_T\right] - \mathop{\mathbb{E}}_{\mathbf{g},\mathbf{h}}\left[\mathbf{g}_S \mathbf{h}_T\right]\right)\mathrm{Tr}\left((P_S \otimes Q_T)\, \psi_{AB}^{\otimes h}\right)\right| \\
&= \left|\sum_{S\subseteq[n]}\left(\mathop{\mathbb{E}}_{\mathbf{M},\mathbf{x},\mathbf{y}}\left[\mathbf{u}_S \mathbf{v}_S\right] - \rho^{|S|}\right)\mathrm{Tr}\left((P_S \otimes Q_S)\, \psi_{AB}^{\otimes h}\right)\right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \delta \sum_{S\subseteq[n]} \left|\mathrm{Tr}\left((P_S \otimes Q_S)\,\psi_{AB}^{\otimes h}\right)\right| \quad \text{(Lemma C.3)}\\
&\leq \delta \sum_{S\subseteq[n]} |||P_S|||_2 |||Q_S|||_2 \quad \text{(Fact A.5)}\\
&\leq \delta \sqrt{\sum_{S\subseteq[n]} |||P_S|||_2^2 \cdot \sum_{S\subseteq[n]} |||Q_S|||_2^2}\\
&= \delta \left(\mathop{\mathbb{E}}_{\mathbf{g}}\left[|||P(\mathbf{g})|||_2^2\right] \mathop{\mathbb{E}}_{\mathbf{g}}\left[|||Q(\mathbf{h})|||_2^2\right]\right)^{1/2} \qquad \leq \delta.
\end{aligned}
$$

Use Lemma C.3 with parameters d and $\delta \leftarrow \delta/9^d$, we have

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

To finish the proof, we will show that,

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

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

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

By the hypercontractivity inequality over the boolean hypercube [50, Page 240]

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

we have

$$
\begin{aligned}
&\sum_{\substack{S,T,S',T'\subseteq[n]\\ S\triangle T\triangle S'\triangle T'=\emptyset}} |||P_S|||_2 |||Q_T|||_2 |||P_{S'}|||_2 |||Q_{T'}|||_2\\
&= \mathop{\mathbb{E}}_{x}\left[f(x)^2 g(x)^2\right]\\
&\leq \sqrt{\mathop{\mathbb{E}}_{x}\left[f(x)^4\right] \mathop{\mathbb{E}}_{x}\left[g(x)^4\right]}
\end{aligned}
$$

$$\begin{aligned}
&\leq 9^d \mathop{\mathbb{E}}_x \left[f(x)^2\right] \mathop{\mathbb{E}}_x \left[g(x)^2\right] \\
&= 9^d \sum_{S\subseteq[n]} |||P_S|||_2^2 \sum_{S\subseteq[n]} |||Q_S|||_2^2 \\
&= 9^d \mathop{\mathbb{E}}_{\mathbf{g}} \left[|||P(\mathbf{g})|||_2^2\right] \mathop{\mathbb{E}}_{\mathbf{g}} \left[|||Q(\mathbf{h})|||_2^2\right] \quad \leq 9^d.
\end{aligned}$$

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

To prove Lemma C.1 Item 1, we need the following lemma whose proof is similar to that of Lemma C.2. We omit the proof here.

Lemma C.4. *In the setting of Lemma C.1, given $d \in \mathbb{Z}_{>0}$, $\delta > 0$, there exists $n_0 = \frac{d^{O(d)}}{\delta^2}$ such that the following holds: For $\mathbf{M} \sim \gamma_{n\times n_0}$,*

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

Proof of Lemma C.1. For item 2, we invoke Lemma C.2 with parameters d and $\delta \leftarrow \delta^3/2$. Using Chebyshev's inequality, we have that for any $\eta > 0$,

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

Using the triangle inequality, we get

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

By Lemma C.4, 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. □

C.2 Upper Bound

We are now ready to prove Theorem 5.5.

Proof of Theorem 5.5. The proof follows that in [53] step by step, except that the Gaussian dimension reduction step in the original proof is replaced by Lemma C.1. 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 A.1 to $P_a^{x,(0)}$ and $Q_b^{y,(0)}$ to get $P_a^{x,(1)}$ and $Q_b^{y,(1)}$, respectively[9]. Note that for all x,y,a,b, $\left\|P_a^{x,(0)}\right\|_2^2 \leq 1$ and $\left\|Q_b^{y,(0)}\right\|_2^2 \leq 1$. Let $d = \frac{C\log^2\frac{1}{\delta}}{\delta(1-\rho)}$, by Lemma A.1 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 A.6 and Lemma A.1 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, $\left\|P_a^{x,(1)}\right\|_2 \leq 1$ and $\left\|Q_b^{y,(1)}\right\|_2 \leq 1$.
  (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 A.3 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 [53, Lemma 10.5] to $P_a^{x,(1)}$, $Q_b^{y,(1)}$ and H, we obtain joint random matrices
  $$\left(P_a^{x,(2)}(\mathbf{g}), Q_b^{y,(2)}(\mathbf{h})\right) \in L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_{2(m^2-1)(n-h)}\right) \times L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_{2(m^2-1)(n-h)}\right),$$
  where $(\mathbf{g},\mathbf{h}) \sim \mathcal{G}_\rho^{\otimes 2(m^2-1)(n-h)}$, such that the following hold.

[9] Specifically, we apply a depolarizing channel Δ_γ for some $\gamma \in (0,1)$ to $P_a^{x,(0)}$ and $Q_b^{y,(0)}$, and then truncate it to be of degree d to get $P_a^{x,(1)}$. Readers may refer to [53] for details.

(1) For any x, y, a, b, $\mathop{\mathbb{E}}_{\mathbf{g}}\left[\left\|P_a^{x,(2)}(\mathbf{g})\right\|_2^2\right] \le 1$ and $\mathop{\mathbb{E}}_{\mathbf{h}}\left[\left\|Q_b^{y,(2)}(\mathbf{h})\right\|_2^2\right] \le 1$.
(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 C.1 to further reduce the number of Gaussian random variables. This is the only part different from the proof in [53].
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 C.1 to $P_a^{x,(2)}(\mathbf{g})$ and $Q_b^{y,(2)}(\mathbf{h})$ with $\delta \leftarrow \delta/(2s^2t^2)$, $d \leftarrow d$, $n \leftarrow 2(m^2-1)(n-h)$, 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 C.1 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 C.1 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, $\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|P_a^{x,(3)}(M\tilde{\mathbf{x}})\right\|_2^2\right] \le 2$ and $\mathop{\mathbb{E}}_{\mathbf{y}}\left[\left\|Q_b^{y,(3)}(M\tilde{\mathbf{y}})\right\|_2^2\right] \le 2$.
(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 $n_0 = \frac{d^{O(d)}s^{12}t^{12}}{\delta^6}$. Therefore, there must exist an M such that all the above four requirements hold. We will use this fixed M throughout the rest of the proof.

- **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 [53, 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 $(x, y, a, b) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$, $\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|P_a^{x,(4)}(\mathbf{x})\right\|_2^2\right] \le 2$ and $\mathop{\mathbb{E}}_{\mathbf{y}}\left[\left\|Q_b^{y,(4)}(\mathbf{y})\right\|_2^2\right] \le 2$.

(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 [53, 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 $P_a^{x,(5)}(\mathbf{x}), Q_b^{y,(5)}(\mathbf{y}) \in L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_{n_0 n_1}\right)$ 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, $\mathop{\mathbb{E}}_{\mathbf{x}}\left[\left\|P_a^{x,(5)}(\mathbf{x})\right\|_2^2\right] \le 2$ and $\mathop{\mathbb{E}}_{\mathbf{y}}\left[\left\|Q_b^{y,(5)}(\mathbf{y})\right\|_2^2\right] \le 2$.

(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 $L^2\left(\mathcal{H}_m^{\otimes h}, \gamma_{n_0 n_1}\right)$ to $\mathcal{H}_m^{\otimes h+n_0 n_1}$.** 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 $n_0 n_1$ quantum registers, so after this step, the number of quantum registers is $h + n_0 n_1$.

For any $(x, y, a, b) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$, applying [53, Lemma 10.11] to $P_a^{x,(5)}(\mathbf{x})$, $Q_b^{y,(5)}(\mathbf{y})$ with $n \leftarrow n_0 n_1$, $h \leftarrow h$, $d \leftarrow 2d$, $\tau \leftarrow \tau$ to get $P_a^{x,(6)}, Q_b^{y,(6)} \in \mathcal{H}_m^{\otimes h+n_0 n_1}$ satisfying the following.

(1) For any x, y, a, b, $\left\|P_a^{x,(6)}\right\|_2^2 \le 2$ and $\left\|Q_b^{y,(6)}\right\|_2^2 \le 2$.

(2) $\mathrm{val}_{h+n_0 n_1}\left(\left\{P_a^{x,(6)}\right\}, \left\{Q_b^{y,(6)}\right\}\right) = \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\mathrm{val}_h\left(\left\{P_a^{x,(5)}(\mathbf{x})\right\}, \left\{Q_b^{y,(5)}(\mathbf{y})\right\}\right)\right]$.

(3) $\sum_{x,a}\mu_A(x)\left|\frac{1}{m^{h+n_0 n_1}}\mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right) - \frac{1}{m^h}\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(P_a^{x,(5)}(\mathbf{x})\right)\right]\right| \le O\left(t\left(9^d m^d\sqrt{\tau}d\right)^{2/3}\right)$ and

$\sum_{y,b}\mu_B(y)\left|\frac{1}{m^{h+n_0 n_1}}\mathrm{Tr}\,\zeta\left(Q_b^{y,(6)}\right) - \frac{1}{m^h}\mathbb{E}\left[\mathrm{Tr}\,\zeta\left(Q_b^{y,(5)}(\mathbf{y})\right)\right]\right| \le O\left(t\left(9^d m^d\sqrt{\tau}d\right)^{2/3}\right)$.

(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 $h + n_0 n_1$. By Lemma A.4 there exist operators $\left\{P_a^{x,(7)}\right\}$ and $\left\{Q_b^{y,(7)}\right\}$ satisfying for all x

$$\sum_a \left\| P_a^{x,(7)} - P_a^{x,(6)} \right\|_2^2 \le \frac{3(t+1)}{m^D} \sum_a \mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right) + 6\sqrt{t}\left(\frac{1}{m^D}\sum_a \mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right)\right)^{1/2}$$
$$\le 10t\left(\frac{1}{m^D}\sum_a \mathrm{Tr}\,\zeta\left(P_a^{x,(6)}\right)\right)^{1/2}. \tag{36}$$

Similarly, for all y, we have

$$\sum_a \left\| Q_b^{y,(7)} - Q_b^{y,(6)} \right\|_2^2 \le 10t\left(\frac{1}{m^D}\sum_b \mathrm{Tr}\,\zeta\left(Q_b^{y,(6)}\right)\right)^{1/2}. \tag{37}$$

Then

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

where in the second last inequality, we use $\left\|P_a^{x,(6)}\right\| \le 2$, $\left\|Q_b^{y,(7)}\right\| \le 1$, and Eqs. (36) and (37). 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

$$\delta = \frac{\epsilon^4}{300t^9}, \tau = \frac{\epsilon^{12}}{t^{27}}\exp\left(-\frac{300t^9\log m}{\epsilon^4(1-\rho)}\log^2\left(\frac{t}{\epsilon}\right)\right) \tag{38}$$

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

$$D = h + n_0 n_1 = \frac{d}{\tau} + \frac{d^{O(d)} s^{12} t^{12}}{\delta^6} \cdot O\left(\frac{d^2}{\tau^2}\right) = O\left(\frac{s^{12} t^{120}}{\epsilon^{48}} \exp\left(\frac{600 t^9 \log m}{\epsilon^4 (1-\rho)} \log^2\left(\frac{t}{\epsilon(1-\rho)}\right)\right)\right),$$

which completes the proof. □

D Truncation

Lemma D.1 (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, $\left|\widehat{P}_a^x(\sigma)\right| \le 1$ and $\left|\widehat{Q}_b^y(\sigma)\right| \le 1$.

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, $\left\|P_a^{x,(2)}\right\|_2 \le 1$ and $\left\|Q_b^{y,(2)}\right\|_2 \le 1$.

(4) For all x, y, a, b, $\left|\mathrm{Tr}\left(\left(P_a^{x,(2)} \otimes Q_b^{y,(2)}\right)\psi_{AB}^{\otimes n}\right) - \mathrm{Tr}\left(\left(P_a^x \otimes Q_b^y\right)\psi_{AB}^{\otimes n}\right)\right| \le \delta$.

(5) For all x, y, a, b,

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

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 $0 \le k_{x,\sigma} < t_{x,\sigma}$ 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 $\widehat{P}_a^{x,(1)}(\sigma) < \widehat{P}_a^x(\sigma) \le 1$, we have $\widehat{P}_a^{x,(1)}(\sigma) \le 1 - \alpha$. So, it can be verified that $\left|\widehat{P}_a^{x,(2)}(\sigma)\right| \le 1$ always holds, which implies that item 3 also holds. To prove the remaining items, we need

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

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,

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

and item 4 follows. Then item 5 follows from Fact 2.20. □

Received 26 July 2024; revised 27 July 2025; accepted 8 August 2025