Quantum Merlin-Arthur and Proofs Without Relative Phase

Roozbeh Bassirian
University of Chicago, IL, USA

Bill Fefferman
University of Chicago, IL, USA

Kunal Marwaha
University of Chicago, IL, USA

Abstract

We study a variant of QMA where quantum proofs have no relative phase (i.e. non-negative amplitudes, up to a global phase). If only completeness is modified, this class is equal to QMA [21]; but if both completeness and soundness are modified, the class (named QMA^+ by Jeronimo and Wu [24]) can be much more powerful. We show that QMA^+ with some constant gap is equal to NEXP, yet QMA^+ with some *other* constant gap is equal to QMA. One interpretation is that Merlin's ability to "deceive" originates from *relative phase* at least as much as from *entanglement*, since $\mathsf{QMA}(2) \subseteq \mathsf{NEXP}$.

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases quantum complexity, QMA(2), PCPs

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.9

Related Version *Full Version*: https://arxiv.org/abs/2306.13247

Funding *Roozbeh Bassirian*: AFOSR (award number FA9550-21-1-0008); NSF under Grant CCF-2044923 (CAREER); U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers; DOE QuantISED grant DE-SC0020360.
Bill Fefferman: AFOSR (award number FA9550-21-1-0008); NSF under Grant CCF-2044923 (CAREER); U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers; DOE QuantISED grant DE-SC0020360.
Kunal Marwaha: NSF Graduate Research Fellowship Program under Grant No. DGE-1746045.

Acknowledgements Thanks to Zachary Remscrim for collaborating on early stages of this project. Thanks to Noam Lifshitz, Dor Minzer, and Kevin Pratt for answering questions about algebraic constructions of expanders. Thanks to Srinivasan Arunachalam, Fernando Granha Jeronimo, Supartha Podder, and Pei Wu for comments on a draft of this manuscript.

1 Introduction

The strangeness of quantum states has at least two fundamental sources: *entanglement*, the source of "spooky action at a distance"; and *relative phase*, which allows for destructive interference. We use complexity theory to probe these sources of strangeness. Extending the main result of [24], we find that QMA^+ (QMA where quantum proofs have no relative phase) is as powerful as NEXP.

A $\mathsf{QMA}_{c,s}$ protocol is a verification task for a quantum computer (termed "Arthur") when interacting with a dishonest but all-powerful machine (termed "Merlin"). If the statement is true ("completeness"), Merlin sends a quantum state ("proof") that truthfully convinces Arthur. If the statement is false ("soundness"), Merlin will send any quantum state possible to deceive Arthur. A valid protocol distinguishes these cases, succeeding with probability at least c in completeness and at most $s < c$ in soundness. Canonically, QMA is the class of all valid $\mathsf{QMA}_{2/3,1/3}$ protocols.

One could potentially reduce the power of QMA by restricting Merlin's proof in completeness. Surprisingly, many restrictions of this type *do not reduce the power of the class*. For example, this is true even if the quantum state is a *subset state* (with no relative phase nor relative non-zero amplitude) [21]. The reason behind this is *promise gap amplification*: there exist techniques to increase the gap $c - s$ to $1 - 2^{-p(n)}$ for any polynomial $p(n)$. As a result, a subset state with polynomially small overlap with the best completeness proof succeeds. This argument generalizes to any set of states that form an ϵ_n-covering of all n-qubit quantum states, where ϵ_n is at least inverse polynomial in n.

By contrast, restricting Merlin's proof in soundness seems to increase the power of this complexity class, since this reduces Merlin's ability to "deceive". For example, if Merlin must send a quantum proof *without relative phase*, Arthur can ask about its *sparsity* (ℓ_1 norm). When a state has no relative phase, a low overlap with $|+\rangle^{\otimes n}$ actually implies it is sparse, as opposed to a state with destructively interfering phases (i.e. any other Hadamard basis vector).

One popular variant of QMA restricts Merlin's *entanglement* over a fixed barrier; it is named QMA(2) (as if there are two unentangled Merlins each sending a quantum proof [26]). This complexity class may seem more powerful than QMA, but despite much study [8, 2, 11, 33, 22, 20, 35], little is known except the trivial bounds $\mathsf{QMA} \subseteq \mathsf{QMA}(2) \subseteq \mathsf{NEXP}$.

What happens if one restricts both *entanglement* and *relative phase*? [24] define $\mathsf{QMA}^+_{c,s}$ and $\mathsf{QMA}^+(2)_{c,s}$, where quantum proofs are required to have no relative phase (non-negative amplitudes, up to a global phase) *in both cases*.[1] Surprisingly, [24] show the existence of constants $1 > c > s > 0$ such that $\mathsf{QMA}^+(2)_{c,s} = \mathsf{NEXP}$, crucially including a protocol to estimate the *sparsity* of a quantum proof. This hints perhaps at a route to prove $\mathsf{QMA}(2) = \mathsf{NEXP}$, since there are other constants $1 > c' > s' > 0$ where $\mathsf{QMA}^+(2)_{c',s'} = \mathsf{QMA}(2)$.[2]

In this work, we show that restricting *relative phase* alone gives the power of NEXP; i.e., there exist constants $1 > c > s > 0$ where $\mathsf{QMA}^+_{c,s} = \mathsf{NEXP}$. Note that assuming $\mathsf{EXP} \neq \mathsf{NEXP}$, this implies QMA^+ cannot be amplified, since as before, there are other constants $1 > c' > s' > 0$ where $\mathsf{QMA}^+_{c',s'} = \mathsf{QMA} \subseteq \mathsf{EXP}$. As a result, techniques to prove $\mathsf{QMA}(2) = \mathsf{QMA}^+(2) = \mathsf{NEXP}$ must crucially use the unentanglement promise inherent in QMA(2). See Figure 1 and Figure 2 for a pictorial description.

1.1 Techniques

Our primary technical contribution is to show a QMA^+ protocol for a NEXP-complete problem. This directly extends the work of Jeronimo and Wu [24], who show a $\mathsf{QMA}^+(2)$ protocol for a NEXP-complete problem. As in [24], we study constraint satisfaction problems (CSPs) with *constant* gap. In $(1, \delta)$-GapCSP, either *all* constraints can be satisfied, or at most a δ fraction of constraints can be satisfied. These problems are known to be NP-hard or NEXP-hard (depending on the problem size) using the PCP theorem [6, 5, 23].

Before proving QMA^+ with some constant gap equals NEXP, we prove $\mathsf{QMA}^+_{\log}$ (with some other constant gap) equals NP. This choice (also taken by [24]) is pedagogical: it allows us to explain the protocol without worrying about input encoding size, since QMA^+ has a polynomial amount of space and verifier runtime. Here, we consider $(1, \delta)$-GapCSP with polynomially many variables and clauses; the quantum proof must certify that there is a satisfying assignment to all clauses.

---

[1] As noted before, restricting the state in completeness may not change the complexity class, but restricting the state in soundness can make the class more powerful, since the latter limits Merlin's adversarial strategies.

[2] This is because every state has constant overlap with some state without relative phase. See also Proposition 29.

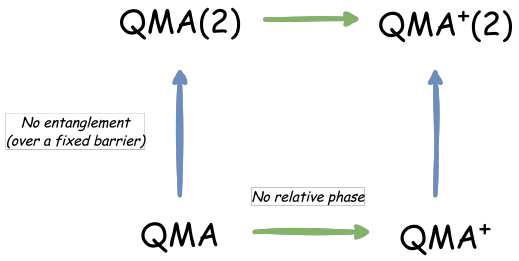

Figure 1 Relationship between QMA^+ and $\mathsf{QMA}(2)$. [24] show that for some constants $1 > c > s > 0$, $\mathsf{QMA}^+(2)_{c,s} = \mathsf{NEXP}$. We show that the same is true for QMA^+. Restricting relative phase does not restrict entanglement across a fixed barrier: for example, consider the GHZ state $|0\rangle^{\otimes n} + |1\rangle^{\otimes n}$, or more generally states where the Schmidt vectors have no relative phase.

The $\mathsf{QMA}^+_{\log}(2)$ protocol of [24] crucially relies on an estimate of *sparsity* (ℓ_1 norm) of a quantum state without relative phase. The overlap of a m-qubit quantum state without relative phase $|\psi\rangle$ with $|+\rangle^{\otimes m}$ is exactly the value $2^{-m/2} \cdot \| |\psi\rangle \|_1$. With multiple quantum proofs $|\psi_1\rangle \otimes \ldots |\psi_k\rangle$, one can *estimate* the sparsity by repeating this "sparsity test" on each $|\psi_i\rangle$, and using a swap test to ensure that all $|\psi_i\rangle$ are approximately equal. Interestingly, no other part of their protocol requires the *no relative phase* assumption.[3]

In $\mathsf{QMA}^+_{\log}$, we have a single quantum proof, so we cannot use this test to estimate sparsity. Instead, we design a similar test that directly enforces a *rigidity* property of the proof.[4] The required form is $\frac{1}{\sqrt{R}} \sum_{j \in [R]} |j\rangle |\vec{v}_j\rangle$, where the second register is constant-sized. Using a "sparsity test" over the second register, Arthur ensures that the second register has one $\vec{v}_j$ per j; but using the *complement* of a "sparsity test" over the whole proof, Arthur ensures the overall state maximizes ℓ_1 norm. States of the required form are optimal for this combination of tests. We make use of the *no relative phase* property in Lemmas 6 and 8.

Now we can describe our protocol. For each constraint $j \in R$, Arthur asks for the values $\vec{v}_j$ associated with the variables involved in constraint j. The protocol either enforces *rigidity* of the quantum proof, or verifies the *constraints* of the CSP. Note that we need two kinds of constraint checks: the values $\vec{v}_j$ must *satisfy* constraint j, and the value of a variable must be *consistent* across the constraints it participates in. For states with the rigidity property, checking *satisfiability* is simple: measure in the computational basis and verify the measured constraint $j, \vec{v}_j$. States with the rigidity property will succeed with probability equal to the satisfying fraction of the CSP assignment.

Checking *consistency* is done using a technique called "regularization" from the PCP literature [17]; for each constraint j, we verify that each variable participating in j has the same value in exactly d other constraints for some constant d, in a way that the edges form an *expander graph*. The expansion property guarantees that cheating on this test is as damaging as cheating on the satisfiability test. Jeronimo and Wu [24] use a swap test to implement these new checks, but this requires multiple quantum proofs. We show how to use a Hadamard test (which requires only one quantum proof) to achieve the same result, building on ideas from previous work [7]. Since there exists a δ such that $(1, \delta)$-GapCSP is NP-hard, this completes the proof of $\mathsf{NP} \subseteq \mathsf{QMA}^+_{\log}$ with some constant gap.

[3] Formally, [24] studies states with non-negative amplitudes. Recall that the set of these states, up to global phase, are equivalent to states with *no relative phase*.

[4] Note that we use the intuition of *rigidity* in a more general context, where Arthur's tests, not a non-local game, enforce states of a certain form.

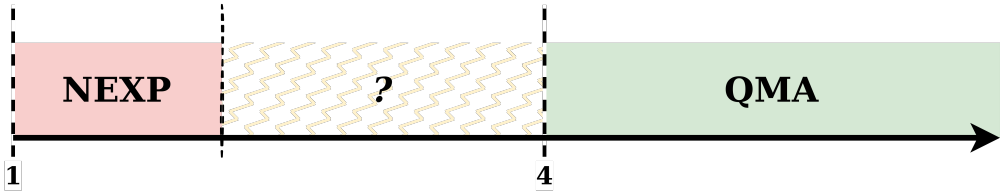

Figure 2 Plot of $\mathsf{QMA}^+_{c,s}$ for increasing $\frac{c}{s}$. By our main result, there exist constants c, s where $1 < \frac{c}{s} < 4$ and $\mathsf{QMA}^+_{c,s} = \mathsf{NEXP}$, but by Corollary 30, $\mathsf{QMA}^+_{c,s} = \mathsf{QMA}$ when $\frac{c}{s} > 4$. Gap amplification of QMA^+ would imply $\mathsf{QMA} = \mathsf{NEXP}$.

When scaling up to QMA^+, one must be careful of how to succinctly encode the input of a NEXP-complete problem. The PCP theorem allows us to choose $(1, \delta)$-GapCSP that is *succinct*, but we need stronger properties. Following the adjustments taken in [24], we choose a PCP system for NEXP that is both *doubly explicit* and *strongly uniform*. *Doubly explicit* means that one can efficiently compute the variables participating in a given constraint *and* the constraints a given variable participates in; using this, we can implement the consistency checks in polynomial time. *Strongly uniform* means that the number of constraints a variable participates in is efficiently computable, and one of a fixed number of possibilities; using this, we only need to build a fixed number of expander graphs during regularization. Recent work also shows how to construct exponentially-sized expander graphs in polynomial time [28, 4]. Once we are through these input encoding difficulties, our protocol is identical to that for NP.

Fundamentally, the *no relative phase* property allows Arthur to verify a number of constraints exponential in the number of qubits. Attempts to do this for QMA(2) gave too small a promise gap [8, 33, 18], too many provers [2, 11, 13], or too much space or time [22, 32]. Jeronimo and Wu [24] show that $\mathsf{QMA}^+(2)$ circumvents this difficulty: using *no relative phase* and *unentanglement*, Arthur enforces the *sparsity* of a quantum proof to solve a NEXP-complete problem. At the center of our work is the insight that *no relative phase* is enough for Arthur to require constant-sized answers to exponentially many questions, solving a NEXP-complete problem with a single polynomial-size quantum proof.

1.2 Related work

The complexity class QMA(2)

The complexity class QMA(2) is known to have promise gap amplification, and to be equal to $\mathsf{QMA}(k)$ for any k at most polynomial in n [22]. It is not obvious how to test for entanglement; even determining whether a polynomially-sized vector is entangled is NP-hard [19]. If there exist efficient approximate "disentanglers" that can create any separable state, then $\mathsf{QMA} = \mathsf{QMA}(2)$; see [2] for some progress. [20] describe quantum variants of the polynomial hierarchy and connect their properties to bounds on QMA(2). It is not even known whether there is a *quantum* oracle separating QMA and QMA(2) [1].

PCPs and expander graphs

Probabilistically checkable proofs (PCPs) show hardness for CSPs with a constant gap [6, 5, 23]. Dinur [17] proves the PCP theorem using a *regularization* step, which adds new constraints associated with the edges of a *regular* expander graph. *Polynomial-time* regularization for NEXP requires an efficient description of exponentially-sized expander graphs. Recent advances in expander graph constructions [28, 4] allow for this type of regularization, first used in [24].

Quantum states and relative phase

Up to a global phase, states with non-negative amplitudes are equivalent to states with no relative phase. [24] propose the class QMA^+ and $\mathsf{QMA}^+(2)$, and show $\mathsf{QMA}^+(2) = \mathsf{NEXP}$. Note that by contrast, QMA restricted to states with *real* amplitudes is equal to QMA [31]. Relative phase was recently proposed as a quantum resource [37]. For both QMA and $\mathsf{QMA}(2)$, restricting Merlin in completeness to send a *subset state* does not change the power of the complexity class (i.e., $\mathsf{QMA} = \mathsf{SQMA}$ and $\mathsf{QMA}(2) = \mathsf{SQMA}(2)$). [21] also shows why their proof strategy fails if Merlin is restricted in both completeness and soundness.

Rigidity and games

Rigidity was first formally introduced in the context of non-local games [30], and have been used to prove several complexity class equalities. For example, the CHSH game [14] tests for a maximally entangled state on two qubits [36], and was used to prove $\mathsf{QMIP} = \mathsf{MIP}^*$ [34]. The Mermin-Peres magic square game tests for two copies of a maximally entangled quantum state, and was used to prove $\mathsf{MIP}^* = \mathsf{RE}$ [25]. Rigidity is known to exist in broader contexts, including some (but not all) linear constraint games [15] and monogamy-of-entanglement games [9].

2 Our setup

We restate the definition of $\mathsf{QMA}^+(k)$ from [24]. When the proof length is not specified, it is allowed to be at most any polynomial in input size. We follow the conventions $\mathsf{QMA}^+ := \mathsf{QMA}^+(1)$, $\mathsf{QMA}^+(k) := \bigcup_{c-s=\Omega(1)} \mathsf{QMA}^+(k)_{c,s}$, and $\mathsf{QMA}^+_{\log} := \mathsf{QMA}^+$ with proof length at most $O(\log n)$.

▶ **Definition 1** ($\mathsf{QMA}^+_\ell(k)_{c,s}$). *Let $k : \mathbb{N} \to \mathbb{N}$ and $s, c, \ell : \mathbb{N} \to \mathbb{R}^+$ be polynomial time computable functions. A promise problem $\mathcal{L}_{yes}, \mathcal{L}_{no} \subseteq \{0,1\}^*$ is in $\mathsf{QMA}^+_\ell(k)_{c,s}$ if there exists a BQP verifier V such that for every $n \in \mathbb{N}$ and every $x \in \{0,1\}^n$:*

- ***Completeness:*** *if $x \in \mathcal{L}_{yes}$, then there exist unentangled states $|\psi_1\rangle, \ldots, |\psi_{k(n)}\rangle$, each on at most $\ell(n)$ qubits and with real non-negative amplitudes, s.t.*

$$\mathbf{Pr}[V(x, |\psi_1\rangle \otimes \ldots \otimes |\psi_{k(n)}\rangle) \text{ accepts}] \geq c(n)\,.$$

- ***Soundness:*** *If $x \in \mathcal{L}_{no}$, then for every set of unentangled states $|\psi_1\rangle, \ldots, |\psi_{k(n)}\rangle$, each on at most $\ell(n)$ qubits and with real non-negative amplitudes, we have*

$$\mathbf{Pr}[V(x, |\psi_1\rangle \otimes \ldots \otimes |\psi_{k(n)}\rangle) \text{ accepts}] \leq s(n)\,.$$

We make a few remarks on this complexity class, with extended discussion in Section 5. First, we stress that the restriction to quantum proofs with non-negative amplitudes is *promise-symmetric*, i.e. both in completeness and in soundness. This is unlike, for example, the class SQMA [21]. Although the restriction to *subset states* is stronger than *non-negative amplitudes*,[5] its use *only* in completeness allows for $\mathsf{SQMA} = \mathsf{QMA}$. In fact, our work implies that QMA with a promise-symmetric *subset state* restriction also interpolates from QMA to NEXP, depending on the size of the promise gap.

[5] A *subset state* is a uniform superposition over a subset of all computational basis states. States with non-negative amplitudes are conical combinations of subset states.

We also explain why the promise gap of QMA^+ cannot obviously be amplified. The first strategy one might try is *parallel repetition*: an honest Merlin sends multiple copies of the original proof and Arthur verifies each copy of the original proof. For QMA, entangling the copies in soundness does not help Merlin, since Arthur's protocol is sound for all quantum states. But perhaps unintuitively, it *can* help for QMA^+. (See Fact 28 for a simple example.) This is because partial measurement can *destroy* the restriction on the quantum proof! For example, Arthur's first measurement may introduce relative phase in the rest of the proof. This fact also obstructs more clever amplification strategies for QMA such as the proof-length preserving variant [29].

Furthermore, it is not clear how to upper-bound QMA^+ beyond the trivial NEXP.[6] One technique to upper-bound QMA is to find the optimal proof using a semidefinite program (or a general convex program). This shows that $\mathsf{QMA} \subseteq \mathsf{PSPACE}$ (or $\mathsf{QMA} \subseteq \mathsf{EXP}$ with a convex program). But these arguments do not immediately transfer to QMA^+. Convex optimization over states with non-negative amplitudes is equivalent to optimizing over the *copositive cone* [10]. Even the *weak membership* problem over the copositive cone (deciding if the optimal vector is close to a non-negative vector) is NP-hard in polynomially-sized vector spaces; recall that quantum states are in *exponentially*-sized vector spaces. These are the same reasons that prevent straightforward upper bounds for $\mathsf{QMA}(2)$ [19].

3 $\mathsf{QMA}^+_{\log}$ Protocol for NP

We first define the problem we consider:

▶ **Definition 2** (CSP system)**.** *A (N, R, q, Σ)-*CSP system *$\mathcal{C}$ on N variables with values in Σ consists of a set (possibly a multi-set) of R constraints $\{\mathcal{C}_1, \ldots, \mathcal{C}_R\}$ where the arity of each constraint is exactly q.*

▶ **Definition 3** ($\mathsf{Value\ of\ CSP}$)**.** *The* value *of a (N, R, q, Σ)-CSP system $\mathcal{C}$ is the maximum fraction of satisfiable constraints over all possible assignments $\sigma : [N] \to \Sigma$. The value of $\mathcal{C}$ is denoted* $\mathbf{val}(\mathcal{C})$.

▶ **Definition 4** (GapCSP)**.** *The $(1, \delta)$-*GapCSP problem *inputs a CSP system $\mathcal{C}$. The task is to distinguish whether $\mathcal{C}$ is such that (in completeness) $\mathbf{val}(\mathcal{C}) = 1$ or (in soundness) $\mathbf{val}(\mathcal{C}) \leq \delta$.*

Fix the input size n, and consider (N, R, q, Σ)-CSP systems where N and R are polynomials in n. Deciding whether or not these systems are satisfiable is NP-hard. In fact, there exists $\delta < 1$ such that deciding $(1, \delta)$-GapCSP on these CSP systems is NP-hard.

▶ **Theorem 5** ([17])**.** *There exist constants $q > 1$ and $|\Sigma|$ such that $(1, 1/2)$-GapCSP is NP-hard.*

Our goal in this section is to construct a protocol for $(1, \delta)$-GapCSP given any (N, R, q, Σ)-CSP system $\mathcal{C}$ where $N, R = \mathrm{poly}(n)$ and $q, |\Sigma| = O(1)$. Let $\kappa := |\Sigma|^q$. We first outline the protocol. Arthur asks for a quantum state from $\mathbb{C}^R \otimes \mathbb{C}^\kappa$; we call the first register the *constraint register* and the second register the *color register*. A quantum proof has the following form:

$$|\psi\rangle := \sum_{j \in [R], x \in \Sigma^q} a_{j,x} |j\rangle |x\rangle$$

[6] $\mathsf{QMA}^+ \subseteq \mathsf{NEXP}$ by directly simulating the quantum proof and verifier.

For completeness, consider the satisfying assignment of variables (in $[N]$) to values (in Σ). Merlin sends the quantum proof $\frac{1}{\sqrt{R}}\sum_{j\in[R]}|j\rangle|\vec{v}_j\rangle$, where each $\vec{v}_j$ is the (ordered) list of values associated with the variables participating in $\mathcal{C}_j$.

Arthur then applies one of two kinds of tests:

1. *Rigidity tests*: These ensure that the quantum proof is of the form $\frac{1}{\sqrt{R}}\sum_{j\in[R]}|j\rangle|\vec{v}_j\rangle$.
2. *Constraint tests*: These verify that values in the quantum proof satisfy constraints of the CSP system.

Below, we separately describe the rigidity tests and constraint tests. In each, we analyze the success probability in completeness and prove lemmas to study soundness. We then combine the technical statements to prove the result.

3.1 Rigidity tests

Arthur enforces *rigidity* of a quantum proof using two tests. The first test is the *Density test*, which maximizes ℓ_1 norm. Here, we measure the state in the Hadamard basis and *accept* if the outcome is $|+\rangle$.[7] Given $|\psi\rangle$, the success probability of this test is

$$D(|\psi\rangle) = |\langle+|\psi\rangle|^2 = \frac{1}{\kappa R}\left|\sum_{j\in[R],x\in\Sigma^q} a_{j,x}\right|^2 = \frac{1}{\kappa R}(\||\psi\rangle\|_1)^2\,.$$

Recall that if $|\psi\rangle$ is a subset state according to subset S, its sparsity $\||\psi\rangle\|_1$ is exactly $\sqrt{|S|}$. In completeness, the quantum proof is a subset state with R elements, so this test passes with probability $\frac{1}{\kappa}$.

The second test is the *Validity test*, which minimizes ℓ_1 norm *only* on the second register. Here, we measure the color register in the Hadamard basis, and *reject* if the outcome is $|+\rangle$. Given $|\psi\rangle$, the success probability of this test is

$$V(|\psi\rangle) = 1 - \langle+|\operatorname{Tr}_R(|\psi\rangle\langle\psi|)|+\rangle = 1 - \frac{1}{\kappa}\sum_{j\in[R]}\left|\sum_{x\in\Sigma^q} a_{j,x}\right|^2\,,$$

where Tr_R is partial trace over the constraint register. In completeness, recall that the proof has the form $\frac{1}{\sqrt{R}}\sum_{j\in[R]}|j\rangle|\vec{v}_j\rangle$, so the success probability is

$$1 - \langle+|\left(\frac{1}{R}\sum_{j\in[R]}|\vec{v}_j\rangle\langle\vec{v}_j|\right)|+\rangle = 1 - \frac{1}{\kappa}\,.$$

In fact, no quantum state without relative phase can pass the *Validity test* with a higher probability:

▶ **Lemma 6.** *Suppose $|\psi\rangle$ has no relative phase. Then $V(|\psi\rangle) \le 1 - \frac{1}{\kappa}$.*

Proof. The success probability $V(|\psi\rangle)$ is

$$\begin{aligned}1 - \frac{1}{\kappa}\sum_{j\in[R]}(\sum_{x\in\Sigma^q} a_{j,x})^2 &= 1 - \frac{1}{\kappa}(\sum_{j\in[R]}\sum_{x,y\in\Sigma^q} a_{j,x}a_{j,y})\\ &= 1 - \frac{1}{\kappa}(1 + \sum_{j\in[R]}\sum_{x,y\in\Sigma^q;x\neq y} a_{j,x}a_{j,y}) \le 1 - \frac{1}{\kappa}\,,\end{aligned}$$

where the second equality follows from $\sum_{j,x} a_{j,x}^2 = 1$ and the inequality holds since $a_{j,x} \ge 0$. ◀

[7] For simplicity, we denote the uniform superposition over all standard basis states by $|+\rangle$. The dimension is clear from the context.

It turns out that is impossible to score high on both the *Validity test* and the *Density test*. We use this to enforce the rigidity property of $|\psi\rangle$.

▶ **Lemma 7.** $D(|\psi\rangle) + V(|\psi\rangle) \leq 1$.

Proof. By Cauchy-Schwarz,

$$D(|\psi\rangle) = \frac{1}{\kappa R}\left|\sum_{j\in[R]}\sum_{x\in\Sigma^q} a_{j,x}\right|^2 \leq \frac{1}{\kappa}\sum_{j\in[R]}\left|\sum_{x\in\Sigma^q} a_{j,x}\right|^2 = 1 - V(|\psi\rangle)\,. \qquad \blacktriangleleft$$

Why does this help with rigidity? Suppose Arthur inputs a quantum proof (without relative phase) $|\psi\rangle$ and runs *Density test* with probability p_1 and *Validity test* with probability p_2. Suppose also that $p_2 > p_1$. Then the expected success probability is $p_1 D(|\psi\rangle) + p_2 V(|\psi\rangle) \leq p_1 + (p_2 - p_1)V(|\psi\rangle) \leq p_1 + (p_2 - p_1)(1 - \frac{1}{\kappa})$. Note that this upper bound is achieved in completeness, and for any state of the form $\frac{1}{\sqrt{R}}\sum_{j\in R}|j\rangle|\vec{v}_j\rangle$. We show that quantum proofs must have this form to reach the upper bound.

One requirement to get close to the upper bound is near-optimal success probability on *Validity test*. We prove that any quantum proof that has this property must be close to a state that assigns one color to each constraint.

▶ **Lemma 8.** *Given* $|\psi\rangle = \sum_{j,x} a_{j,x}|j\rangle|x\rangle$ *with no relative phase (i.e.* $a_{j,x} \geq 0$*), let*

$$\gamma := \max_{\nu:[R]\to\Sigma^q} \sum_{j\in[R]} a_{j,\nu(j)}^2$$

be associated with maximizing function σ*, and let* $|\phi\rangle := \frac{1}{\sqrt{\gamma}}\sum_j a_{j,\sigma(j)}|j\rangle|\sigma(j)\rangle$*. Fix any* $d \geq 0$*. If* $V(|\psi\rangle) = 1 - \frac{1+d}{\kappa}$*, then* $|\langle\psi|\phi\rangle|^2 \geq 1 - d$*.*

Proof. Note that for all $x \in \Sigma^q$, $a_{j,\sigma(j)} \geq a_{j,x}$; otherwise, σ is not maximizing. Using the proof of Lemma 6,

$$d = \sum_{j\in[R]}\sum_{x,y\in\Sigma^q;x\neq y} a_{j,x}a_{j,y} \geq \sum_{j\in[R]}\sum_{y\in\Sigma^q;y\neq\sigma(j)} a_{j,\sigma(j)}a_{j,y} \geq \sum_{j\in[R]}\sum_{y\in\Sigma^q;y\neq\sigma(j)} a_{j,y}^2\,.$$

So then,

$$\gamma = \sum_{j\in[R]} a_{j,\sigma(j)}^2 = \Big(\sum_{j\in[R]}\sum_{x\in\Sigma^q} a_{j,x}^2\Big) - \Big(\sum_{j\in[R]}\sum_{x\in\Sigma^q;x\neq\sigma(j)} a_{j,x}^2\Big) \geq 1 - d\,.$$

So $|\langle\psi|\phi\rangle|^2 = (\frac{1}{\sqrt{\gamma}}\sum_j a_{j,\sigma(j)}^2)^2 = \gamma \geq 1 - d$. $\blacktriangleleft$

Another requirement to get close to the upper bound is near-optimal success probability on *Density test*, up to Lemma 7. Consider any quantum proof that passes *Validity test* with probability close to $1 - \frac{1}{\kappa}$ and *Density test* with probability close to $\frac{1}{\kappa}$; we show it must be close to a state of the form $\frac{1}{\sqrt{R}}\sum_{j\in R}|j\rangle|\vec{v}_j\rangle$. Now we can prove the soundness of the rigidity test by relying on the following fact:

▶ **Fact 9.** *Let* $0 \leq \Pi \leq \mathbb{I}$ *be a positive semi-definite matrix, and let* $|\psi_1\rangle$ *and* $|\psi_2\rangle$ *be quantum states such that* $|\langle\psi_1|\psi_2\rangle|^2 \geq 1 - d$*. Then* $|\langle\psi_1|\Pi|\psi_1\rangle - \langle\psi_2|\Pi|\psi_2\rangle| \leq \sqrt{d}$*.*

Proof. The quantity $|\langle\psi_1|\Pi|\psi_1\rangle - \langle\psi_2|\Pi|\psi_2\rangle| = |\mathrm{Tr}(\Pi(|\psi_1\rangle\langle\psi_1| - |\psi_2\rangle\langle\psi_2|))|$ is upper-bounded by the trace distance of $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$, which has value $\sqrt{1 - |\langle\psi_1|\psi_2\rangle|^2} \leq \sqrt{d}$. $\blacktriangleleft$

▶ **Lemma 10** (Rigidity lemma). *Let $d_2 \geq d_1 \geq 0$ be small constants. Suppose $|\psi\rangle$, $|\phi\rangle$, and σ are defined as in Lemma 8, and $|\chi\rangle$ is defined as*

$$|\chi\rangle := \frac{1}{\sqrt{R}} \sum_{j \in [R]} |j\rangle |\sigma(j)\rangle .$$

If $D(|\psi\rangle) = \frac{1}{\kappa} - d_1$ and $V(|\psi\rangle) = 1 - \frac{1}{\kappa} - d_2$, then $|\langle\chi|\psi\rangle|^2 \geq 1 - \kappa d_1 - (\kappa+1)\sqrt{\kappa \cdot d_2}$.

Proof. By Lemma 8, we know that $|\langle\psi|\phi\rangle|^2 \geq 1 - \kappa \cdot d_2$. So by Fact 9, for any quantum state $|\mu\rangle$, $\left||\langle\mu|\phi\rangle|^2 - |\langle\mu|\psi\rangle|^2\right| \leq \sqrt{\kappa \cdot d_2}$. We use this in two places. First, when $|\mu\rangle = |+\rangle$. Since $D(|\psi\rangle) = |\langle+|\psi\rangle|^2 = \frac{1}{\kappa} - d_1$, we have $|\langle+|\phi\rangle|^2 \geq \frac{1}{\kappa} - d_1 - \sqrt{\kappa \cdot d_2}$ by triangle inequality. Second, when $|\mu\rangle = |\chi\rangle$. Notice that

$$|\langle\chi|\phi\rangle|^2 = \kappa\, |\langle+|\phi\rangle|^2 \geq 1 - \kappa(d_1 + \sqrt{\kappa \cdot d_2}) .$$

Again by triangle inequality,

$$|\langle\chi|\psi\rangle|^2 \geq 1 - \kappa(d_1 + \sqrt{\kappa \cdot d_2}) - \sqrt{\kappa \cdot d_2} . \qquad \blacktriangleleft$$

Intuitively, Lemma 10 allows us to *tune* the probability of each test in the NP protocol. As we explain in the analysis (Section 3.3), if the probabilities of running *Validity test* and *Density test* are much higher than that for constraint tests, then if d_1 or d_2 is large, these two tests catch a "deceptive" quantum proof in soundness. This allows constraint tests to focus on the case of small d_1 and d_2; i.e. nearly *rigid* quantum proofs.

3.2 Constraint tests

We analyze the constraint tests on *rigid* quantum proofs, i.e. states of the form $|\psi\rangle = \frac{1}{\sqrt{R}} \sum_{j \in [R} |j\rangle |\vec{v}_j\rangle$. The verifier needs to check two properties:

- **(i)** (*satisfiability*) For all $j \in [R]$, the assignment $\vec{v}_j$ satisfies $\mathcal{C}_j$.
- **(ii)** (*consistency*) Each variable is assigned the same value when participating in different constraints.

One may ask why we even need to check for *consistency*. Couldn't we ask for the assignment of each variable $a : [N] \to \Sigma$, for example as the quantum proof $\frac{1}{\sqrt{N}} \sum_{i \in [N]} |i\rangle |a(i)\rangle$? The problem with this is checking *satisfiability* becomes difficult, since the assigned values are given in superposition.[8]

Instead, with a state $\frac{1}{\sqrt{R}} \sum_{j \in [R} |j\rangle |\vec{v}_j\rangle$, *satisfiability* is easy to verify: measure the first register (observing some $|j\rangle |\vec{v}_j\rangle$), and compute $\mathcal{C}_j(\vec{v}_j)$. Let u_s be the number of unsatisfied constraints. The outcome is 1 with probability $1 - \frac{u_s}{R}$.

But this form of quantum proof gives Merlin a new way to "deceive": for a given variable, send different values depending on the constraint! We prevent this by checking for *consistency*, similarly to the pre-processing step of [17] sometimes called *regularization*. As in [24, Section 7], we add "consistency constraints" to the CSP system $\mathcal{C}$ as follows:[9]

- For each variable $i \in [N]$, let V_i represent the constraints that i participates in.

[8] There is a way around this limitation for CSP systems consisting of *unique game constraints*, where each (binary) constraint involving variables i_1, i_2 accepts exactly one $a(i_2)$ for each $a(i_1)$. See [24, Section 6] for more discussion.

[9] Note that since N and R are polynomially-sized, this process is efficient.

- Fix a constant d. For each $i \in [N]$, draw a d-regular graph with vertices V_i that is *expanding*.[10]
- Each edge (j_1, j_2) of each expander V_i represents a "consistency constraint", where we assert that the value of variable i sent with constraint j_1 equals that sent with constraint j_2.

Using *expander* graphs allows us to prevent this kind of "deceptive" Merlin: either the proof fails many of the original constraints, or it fails many "consistency constraints". Let u_e be the number of unsatisfied "consistency constraints" out of $Rq \cdot \frac{d}{2}$:

▷ Claim 11 ([17, Lemma 4.1]). Consider a (N, R, q, Σ)-CSP system $\mathcal{C}$, and apply *regularization*. If $\mathbf{val}(\mathcal{C}) = 1$, then all "consistency constraints" can be simultaneously satisfied. If $\mathbf{val}(\mathcal{C}) \leq \delta$, then the total number of unsatisfied constraints $(u_s + u_e)$ is at least $(1 - \delta)R$.

How do we check these "consistency constraints"? Over the next few paragraphs, we construct a unitary related to permutations on the constraint graph. In completeness, the quantum proof is an eigenvector of this unitary, but in soundness, all *rigid* quantum proofs are detectably far (i.e. using a Hadamard test) from an eigenvector. We study the graph $\widetilde{G}$ with $R \cdot q$ vertices, where each vertex (j, i) corresponds to a clause j and a variable i that participates in clause j. Let $\widetilde{G}$ be the union of all consistency edges created during regularization, i.e. (j_1, j_2) for variable i becomes the edge $((j_1, i), (j_2, i))$. Note that $\widetilde{G}$ contains a copy of each expander graph, so it is d-regular.

We now choose d permutations. It is a classical fact that the adjacency matrix of any d-regular graph can always be decomposed to d permutations. Let $\pi_1, \ldots, \pi_d$ be the decomposition of $\widetilde{G}$; recall that these are permutations on $V(\widetilde{G})$ where $|V(\widetilde{G})| = R \cdot q$. For each $k \in [d]$, we identify π_k with a permutation on $[R] \times [N]$, where any $(j, i) \in [R] \times [N]$ that is not a vertex of $\widetilde{G}$ (i.e. variable i does not participate in constraint j) is mapped to itself.[11] Note that this map always preserves the variable $i \in [N]$; without loss of generality, we also identify π_k with its restriction $[R] \times [N] \to [R]$. From here on, we use this last definition of π_k, which maps constraint j_1 that variable i participates in to another constraint j_2 that variable i participates in, and identity otherwise.

Now consider a *rigid* quantum proof, i.e. of the form $|\psi\rangle = \frac{1}{\sqrt{R}} \sum_{j \in [R]} |j\rangle |\vec{v}_j\rangle$. Since there are a polynomial number of variables and constraints, we can efficiently transform $|\psi\rangle$ to $|\phi'\rangle$, where

$$|\phi'\rangle := \frac{1}{\sqrt{q \cdot R}} \sum_{j \in [R]} \sum_{i \in \mathcal{C}_j} |j\rangle |\vec{v}_j\rangle |i\rangle |v_j(i)\rangle \ .$$

Here, $i \in \mathcal{C}_j$ are the variables participating in $\mathcal{C}_j$, and $v_j(i)$ is the value of this variable according to $\vec{v}_j$.

We now would like to construct a unitary on $|\phi'\rangle$ that maps $|j\rangle |i\rangle |value\rangle$ to $|j'\rangle |i\rangle |value\rangle$ for some other constraint j' that i participates in. In completeness, this unitary would leave the state unchanged. Notice that from the perspective of such a unitary, the second register containing $|\vec{v}_j\rangle$ is "junk". Fortunately, we can measure out the second register in the Hadamard basis, and reject if the outcome is *not* $|+\rangle$. All *rigid* states will observe outcome $|+\rangle$ with probability $\frac{1}{\kappa}$; one can see this by writing the second register in the Hadamard basis.

[10] For technical reasons of Claim 11, we require that the Cheeger constant is at least 2.
[11] These permutations (and their inverses) are all efficient because N and R are polynomially-sized.

Suppose the observed outcome is $|+\rangle$; let us call the postselected state $|\phi\rangle$, where

$$|\phi\rangle := \frac{1}{\sqrt{q \cdot R}} \sum_{j \in R} \sum_{i \in \mathcal{C}_j} |j\rangle |i\rangle |v_j(i)\rangle \ .$$

For each $k \in [d]$, we now implement the in-place transformation Π_k according to $\pi_k : [R] \times [N] \to [R]$, where

$$\Pi_k : |j\rangle |i\rangle |v_j(i)\rangle \to |\pi_k(j,i)\rangle |i\rangle |v_j(i)\rangle \ .$$

Recall that the map $(j,i) \mapsto (\pi_k(j,i), i)$ is a permutation. Since we have access both to this permutation and its inverse, we can implement Π_k.

Note that in a satisfiable instance, $\Pi_k |\phi\rangle = |\phi\rangle$. By contrast, if $v_j(i) \neq v_{j'}(i)$, $|j'\rangle |i\rangle |v_j(i)\rangle$ is orthogonal to $|\phi\rangle$. Hence, $\langle\phi| \Pi_k |\phi\rangle$ is the fraction of satisfied "consistency constraints" observed by π_k. We use the *Hadamard test* to measure this value, in a similar way to the *Spectral test* in [7]. Note that unlike the swap test, the Hadamard test only uses one copy of a quantum state.

▶ **Definition 12** (Hadamard test). *Let $|\psi\rangle$ be a quantum state and U a unitary operator.*

1. *Prepend a control qubit to $|\psi\rangle$, to create $|0\rangle |\psi\rangle$.*
2. *Apply a Hadamard on the control qubit, to create $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) |\psi\rangle$.*
3. *Apply U, controlled by the control qubit, to create $\frac{1}{\sqrt{2}} (|0\rangle |\psi\rangle + |1\rangle U |\psi\rangle)$.*
4. *Apply a Hadamard on the control qubit, to create $\frac{1}{2} |0\rangle (|\psi\rangle + U |\psi\rangle) + \frac{1}{2} |1\rangle (|\psi\rangle - U |\psi\rangle)$.*
5. *Measure the control qubit, and accept if the output is 0.*

The success probability is then

$$\frac{1}{4} \| |\psi\rangle + U |\psi\rangle \|^2 = \frac{1}{2} + \frac{1}{4} \langle\psi| U + U^\dagger |\psi\rangle = \frac{1}{2} + \frac{\mathrm{Re} \langle\psi| U |\psi\rangle}{2} \ .$$

We now can describe the constraint tests together:

- **(i)** With probability $\frac{1}{qd\kappa+1}$, check *satisfiability*. This succeeds with probability $1 - \frac{u_s}{R}$.
- **(ii)** With probability $\frac{qd\kappa}{qd\kappa+1}$, generate $|\phi'\rangle$, and measure the second register in the Hadamard basis. If the output state is not $|+\rangle$, reject. Otherwise, choose a random $k \in [d]$, and perform a Hadamard test with Π_k. This succeeds with probability $\frac{1}{\kappa}(\frac{1}{2} + \frac{1}{2}\mathbb{E}_k[\mathrm{Re} \langle\phi| \Pi_k |\phi\rangle]) = \frac{1}{\kappa}(1 - \frac{u_e}{qdR})$.[12]

The overall success probability of the constraint tests is

$$\frac{1}{qd\kappa+1}(1 - \frac{u_s}{R}) + \frac{qd\kappa}{qd\kappa+1} \left(\frac{1}{\kappa} \cdot (1 - \frac{u_e}{qdR}) \right) = \frac{qd+1}{qd\kappa+1} - \frac{u_e + u_s}{R \cdot (qd\kappa+1)} \ .$$

We now show a constant gap between completeness and soundness. In completeness, $u_e = u_s = 0$, so $|\psi\rangle$ passes the constraint tests with probability $C^{\mathrm{YES}} := \frac{qd+1}{qd\kappa+1}$. In soundness, recall that $\mathbf{val}(\mathcal{C}) \leq \delta$, so by Claim 11, any *rigid* quantum proof passes the constraint tests with probability at most $C^{\mathrm{YES}} - \frac{1-\delta}{qd\kappa+1}$. We now apply Lemma 10: any quantum proof that passes *Density test* and *Validity test* with probabilities too similar to that in completeness must pass the constraint tests with probability less than C^{YES}.

▶ **Corollary 13.** *In soundness, if $D(|\psi\rangle) = \frac{1}{\kappa} - d_1$ and $V(|\psi\rangle) = 1 - \frac{1}{\kappa} - d_2$, then*

$$C(|\psi\rangle) \leq C^{\mathrm{YES}} - \frac{1-\delta}{qd\kappa+1} + \left(\kappa d_1 + (\kappa+1)\sqrt{\kappa \cdot d_2} \right)^{1/2} \ .$$

[12] Note that in our protocol, $\langle\phi| \Pi_k |\phi\rangle$ is always real because $|\phi\rangle$ and Π_k have real values.

3.3 Analysis

In the protocol, Arthur applies *Density test*, *Validity test*, or *constraint tests* with probability p_1, p_2, p_3, respectively, where $p_3 = 1 - p_1 - p_2$.

We start by analyzing the success probability of the protocol in completeness. Here, $\mathbf{val}(\mathcal{C}) = 1$, and the quantum proof $|\psi\rangle = \frac{1}{R}\sum_j |j\rangle |\vec{v}_j\rangle$ is such that $\vec{v}_j$ is a satisfying assignment to the variables that participate in $\mathcal{C}_j$. The success probability for each test is exactly $\frac{1}{\kappa}$, $1 - \frac{1}{\kappa}$, and C^{YES}, respectively. So the success probability of the protocol in completeness is $P_{\mathrm{YES}} = \frac{p_1}{\kappa} + p_2(1 - \frac{1}{\kappa}) + p_3 C^{\mathrm{YES}}$.

We now choose the probabilities p_1, p_2, p_3. Choose $\lambda := \frac{1-\delta}{qd\kappa+1}$.

1. We first set a *distance threshold* $\varepsilon := \frac{\lambda}{\Gamma}$ for a large enough constant $\Gamma(\kappa, q, d, \delta)$ satisfying

$$\left(\kappa\varepsilon + (\kappa+1)\sqrt{\kappa\cdot\varepsilon}\right)^{1/2} \leq \frac{\lambda}{2}\,.$$

2. Let $Z := \frac{1}{2} + 1 + \frac{\varepsilon}{4(1-C^{\mathrm{YES}})}$. Then let

$$p_1 = \frac{1}{2}\cdot\frac{1}{Z} \qquad p_2 = \frac{1}{Z} \qquad p_3 = \frac{\varepsilon}{4(1-C^{\mathrm{YES}})}\cdot\frac{1}{Z}\,.$$

Now we study soundness, i.e. when $\mathbf{val}(\mathcal{C}) \leq \delta$. We again denote the quantum proof as $|\psi\rangle$. We divide up the analysis into a few parts:

1. A quantum proof that is "too sparse" (i.e. $D(|\psi\rangle) = \frac{1}{\kappa} - d$ for any $d \geq \varepsilon$) is detected by *Density test*.

$$\begin{aligned}
P_{\mathrm{NO}} &= p_1(\frac{1}{\kappa} - d) + p_2 V(|\psi\rangle) + p_3 C(|\psi\rangle) \\
&\leq P_{\mathrm{YES}} - p_1 d + p_3(1 - C^{\mathrm{YES}}) \\
&\leq P_{\mathrm{YES}} - p_1\varepsilon + p_3(1 - C^{\mathrm{YES}}) \\
&= P_{\mathrm{YES}} - \frac{\varepsilon}{2Z} + \frac{\varepsilon}{4Z} = P_{\mathrm{YES}} - \frac{\varepsilon}{4Z}\,.
\end{aligned}$$

2. A quantum proof that is "too dense" (i.e. $D(|\psi\rangle) = \frac{1}{\kappa} + d$ for any $d \geq \varepsilon$) is detected by *Validity test*.

$$\begin{aligned}
P_{\mathrm{NO}} &= p_1(\frac{1}{\kappa} + d) + p_2 V(|\psi\rangle) + p_3 C(|\psi\rangle) \\
&\leq p_1(\frac{1}{\kappa} + d) + p_2(1 - \frac{1}{\kappa} - d) + p_3 \\
&= P_{\mathrm{YES}} - (p_2 - p_1)d + p_3(1 - C^{\mathrm{YES}}) \\
&\leq P_{\mathrm{YES}} - (p_2 - p_1)\varepsilon + p_3(1 - C^{\mathrm{YES}}) \\
&= P_{\mathrm{YES}} - \frac{\varepsilon}{2Z} + \frac{\varepsilon}{4Z} = P_{\mathrm{YES}} - \frac{\varepsilon}{4Z}\,,
\end{aligned}$$

   where the first inequality follows from Lemma 7.

3. A quantum proof that is "the right density" (i.e. $D(|\psi\rangle) = \frac{1}{\kappa} + d_1$ for $|d_1| \leq \varepsilon$) but far from "valid" ($V(|\psi\rangle) = 1 - \frac{1}{\kappa} - d_2$ for $d_2 \geq \varepsilon$) is detected by *Validity test* when $p_2 > p_1$.

$$\begin{aligned}
P_{\mathrm{NO}} &\leq p_1(\frac{1}{\kappa} + |d_1|) + p_2(1 - \frac{1}{\kappa} - d_2) + p_3 \\
&\leq P_{\mathrm{YES}} + p_1|d_1| - p_2 d_2 + p_3(1 - C^{\mathrm{YES}}) \\
&\leq P_{\mathrm{YES}} - (p_2 - p_1)\varepsilon + p_3(1 - C^{\mathrm{YES}}) \\
&= P_{\mathrm{YES}} - \frac{\varepsilon}{4Z}\,.
\end{aligned}$$

4. Lastly, a quantum proof that is nearly *rigid* (i.e. $D(|\psi\rangle) = \frac{1}{\kappa} + d_1$ and $V(|\psi\rangle) = 1 - \frac{1}{\kappa} - d_2$ for any $|d_1|, d_2 \leq \varepsilon$) is detected by the constraint tests.

$$\begin{aligned} P_{\text{NO}} &= p_1(\frac{1}{\kappa} + d_1) + p_2(1 - \frac{1}{\kappa} - d_2) + p_3 C(|\psi\rangle) \\ &\leq P_{\text{YES}} + p_1 d_1 - p_2 d_2 + p_3 \left(-\frac{1-\delta}{qd\kappa + 1} + \left(\kappa d_1 + (\kappa+1)\sqrt{\kappa \cdot d_2} \right)^{1/2} \right) \\ &\leq P_{\text{YES}} + p_1 d_1 - p_2 d_2 - p_3 \frac{\lambda}{2} \,. \end{aligned}$$

The first inequality follows from Corollary 13, and the second inequality holds by our choice of ε. Note that $d_2 \geq 0$ by Lemma 6. By Lemma 7, if $d_1 \geq 0$, then $d_1 \leq d_2$; otherwise $d_1 \leq d_2$ trivially. So then

$$P_{\text{NO}} \leq P_{\text{YES}} - (p_2 - p_1) d_2 - p_3 \frac{\lambda}{2} \leq P_{\text{YES}} - \frac{\varepsilon \lambda}{8(1 - C^{\text{YES}})Z} \,.$$

Combining these cases proves the following result:

▶ **Theorem 14.** *Given an instance of* $(1,\delta)$*-*GapCSP*, the* $\mathsf{QMA}^+_{\log}$ *protocol succeeds with probability* P^{YES} *in completeness and at most* $P^{\text{YES}} - \Delta$ *in soundness for some constants* $1 > P^{\text{YES}} > \Delta > 0$*.*

▶ **Corollary 15.** *There exist constants* $1 > P^{\text{YES}} > \Delta > 0$ *such that* $\mathsf{NP} \subseteq \mathsf{QMA}^+_{\log}$ *with completeness* P^{YES} *and soundness* $P^{\text{YES}} - \Delta$*.*

4 QMA$^+$ protocol for NEXP

Our goal in this section is to modify the previous protocol to solve an NEXP-complete problem. Again by the PCP theorem, the *succinct* $(1,\delta)$-GapCSP problem with exponentially many variables and clauses is NEXP-complete. The *succinctness* allows us to efficiently describe the problem input. What remains is to ensure that the verifier's protocol is efficient. Previously, the unitary transformations were efficient because the verifier handled poly(n)-size graphs. Furthermore, the expanders used to check the equality constraints for each variable may have different sizes. Now that there can be exponentially many possibilities for the size of each cluster, naively applying the previous technique is not efficient. These challenges were addressed in [24] by considering a PCP construction for NEXP with strong properties.

▶ **Theorem 16** ([24])**.** *There is a* NEXP*-hard* $(1,\delta)$*-*GapCSP *instance for some* $(N = 2^{\text{poly}(n)}, R = 2^{\text{poly}(n)}, q = O(1), \Sigma = \{0,1\})$*-CSP system* $\mathcal{C}$ *that is both* τ-strongly uniform *for some constant* τ *and* polylog(NR)-doubly explicit*.*

Informally, every constraint in a *succinct* CSP system must be computable in polynomial time. The *doubly explicit* property further requires the existence of efficient maps from variables to constraints *and* from constraints to variables. Intuitively, these maps allow us to efficiently implement the Hadamard test of the consistency checks.

We include the formal definition of these properties. Define $Adj_{\mathcal{C}}(j)$ to be the list of variables participating in $\mathcal{C}_j$, and $Adj_V(i)$ be the list of constraints that depend on variable i.

▶ **Definition 17** (Doubly explicit CSP)**.** *A* (N, R, q, Σ)*-CSP system* $\mathcal{C}$ *is* $Z(N,R)$-doubly explicit *if for all* $i \in [N]$ *and* $j \in [R]$*, the following are computable in time* $Z(N,R)$*:*

- **(i)** *Cardinality of* $Adj_V(i)$ *and* $Adj_{\mathcal{C}}(j)$ *for all* $i \in [N]$ *and* $j \in [R]$*.*
- **(ii)** $Adj^{ind}_{\mathcal{C}} : [R] \times [N] \to [q]$*; if* i *participates in* $\mathcal{C}_j$*, then* $Adj^{ind}_{\mathcal{C}}(j,i) = \imath$ *is the index of* i *in* $Adj_{\mathcal{C}}(j)$*.*

(iii) $Adj_{\mathcal{C}}^{id} : [R] \times [q] \to [N]$*;* $Adj_{\mathcal{C}}^{id}(j, \imath) = i$ *is the* $\imath$*-th variable of* $Adj_{\mathcal{C}}(j)$*.*
(iv) $Adj_V^{ind} : [N] \times [R] \to [R]$*; if* i *participates in* $\mathcal{C}_j$*, then* $Adj_V^{ind}(i, j) = \jmath$ *is the index of* j *in* $Adj_V(i)$*.*
(v) $Adj_V^{id} : [N] \times [R] \to [R]$*;* $Adj_V^{id}(i, \jmath) = j$ *is the* $\jmath$*-th variable* $Adj_V(i)$*.*

This property alone is not enough for efficient regularization: the verifier must know how to implement an expander of size $|Adj_V(i)|$ for *all* variables i. The *strongly uniform* property resolves this complication.

▶ **Definition 18** (Strongly uniform CSP)**.** *Let* $\tau \in \mathbb{N}$*. A* (N, R, q, Σ)*-CSP system* $\mathcal{C}$ *is* τ*-strongly uniform if the variable set* $[N]$ *can be partitioned into at most* τ *different subsets* $\bigcup_y V_y$ *such that* $|Adj_V(i)| = |Adj_V(j)| = n_k$ *if* i *and* j *belong to the same part* V_k*. Furthermore, the part* $k \in [\tau]$ *can be determined in time* $\mathrm{polylog}(NR)$*.*

A τ-strongly uniform CSP system allows the verifier to use τ different (possibly exponential size) d-regular expanders. These can be constructed in polynomial time:

▶ **Theorem 19** (Doubly explicit expander graphs [28, 4])**.** *There is a constant* d *such that the following explicit constructions of expander graphs exist:*

1. *For every* n*, there is a* d*-regular graph on* n *vertices.*
2. *For every prime* $p > 17$*, there is a* d*-regular graph on* $n = p(p^2 - 1)$ *vertices, and the graph can be decomposed into* d *permutations* $\pi_1, \ldots, \pi_d$ *that can each be evaluated in time* $\mathrm{polylog}(n)$*.*[13]

Furthermore, the neighbors of each variable can be listed in $\mathrm{polylog}(n)$*, and the graphs have Cheeger constant at least 2.*

With this theorem, the verifier can choose a large constant n_0, and use Construction 1 if $n_i \leq n_0$. Otherwise, the verifier can cover *almost* all n_i vertices with an explicit expander using Construction 2.

▶ **Theorem 20** (Primes in short intervals [12])**.** *There is an absolute constant* k_0 *such that for any integer* $k > k_0$*, there is a prime in the interval* $[k - 4k^{2/3}, k]$*.*

We modify the protocol to expect the quantum proof $|p_1, p_2, \ldots, p_\tau\rangle \otimes \frac{1}{R} \sum_{j \in R} |j\rangle\,|\vec{v}_j\rangle$ such that each $p_i \in [\lfloor n_i^{1/3} \rfloor - 4\lfloor n_i^{1/3} \rfloor^{2/3}, \lfloor n_i^{1/3} \rfloor]$. The verifier measures the primes, and can always check that every p_i is a prime number in the required range. The rest of the analysis is similar to the NP protocol, but regularized using these efficient expanders. We first explain how to efficiently implement the *constraint tests*, and then analyze the QMA^+ protocol.

4.1 Efficient constraint tests

We show how to efficiently implement *consistency* checks that imply a version of Claim 11. Fix any vertex i. Let n be the number of constraints that depend on i and p be the corresponding prime. Let n_0 be a large enough constant. If $n \leq n_0$, then use Construction 1 d-regular expander to wire new copies of the vertex together, just as for NP. Otherwise, use Construction 2 to generate a d-regular expander graph of size $p(p^2-1) \in [n - O(n^{8/9}), n]$ that wires nearly all copies of the vertex together. Then add d self-loops for the remaining vertices. The number of vertices with self loops is at most some ηn (for some small constant η) since $p(p^2 - 1) \geq n - O(n^{8/9})$; we can make η arbitrarily small by choosing a large enough n_0.

[13] In fact, since these graphs are Cayley graphs, both the permutations and their inverses can be evaluated in time $\mathrm{polylog}(n)$. We use both π_k and π_k^{-1} in the constraint tests to implement the unitary $\mathcal{M}_k$.

▷ Claim 21 ([24]). Consider a (N, R, q, Σ)-CSP system $\mathcal{C}$, and apply *efficient regularization*. If $\mathbf{val}(\mathcal{C}) = 1$, then all "consistency constraints" can be simultaneously satisfied. If $\mathbf{val}(\mathcal{C}) \leq \delta$, then the total number of unsatisfied constraints is at least $(1 - \delta - q\eta)R$.

The analysis of Claim 21 is similar to Claim 11. The additional factor of $q\eta$ comes from the self-loop constraints; these can be satisfied without violating any "consistency constraint".

After measuring the primes, let the verifier act on the space $\mathbb{C}^R \otimes \mathbb{C}^\kappa \otimes \mathbb{C}^N \otimes \mathbb{C}^{|\Sigma|}$. We explicitly define the unitary operators that are used in the NP protocol. These definitions exactly match [24]. First, operator $\mathcal{A}$ expands the values from the list of values of variables involved in each constraint:

$$\mathcal{A} : |j\rangle |v\rangle |0\rangle |0\rangle \to \frac{1}{q} \sum_{r=1}^{q} |j\rangle |\vec{v}\rangle |i_r\rangle |v_r\rangle \ ,$$

Here, v is the list of values of variables involved in constraint j, i_r is the r-th variable involved in j, and v_r is the value of i_r according to v. Next, define the permutation operators $\mathcal{M}_k$ for each $k \in [d]$ that implement the d permutations of each efficiently constructed expander:

$$\mathcal{M}_k : |j\rangle |v\rangle |i\rangle |v'\rangle \to |\Pi_k(j, i)\rangle |v\rangle |i\rangle |v'\rangle$$

The last operation computes the constraints in superposition:

$$\mathcal{B} |j\rangle |v\rangle |0\rangle \to |j\rangle |v\rangle |\mathcal{C}_j(v)\rangle$$

▶ **Theorem 22** ([24]). *$\mathcal{A}, \mathcal{B}, \mathcal{M}_k$ can be implemented by* BQP *circuits.*

4.2 Analysis

The analysis is nearly the same as in the NP protocol, with the minor difference that the verifier also receives $p_1, \ldots, p_\tau$ in the quantum proof. The rigidity tests are unchanged. For the constraint tests, the verifier can use the explicit operators $\mathcal{A}, \mathcal{B}, \mathcal{M}_k$:

- **(i)** For *satisfiability*, the prover computes $\mathcal{B} |\psi\rangle |0\rangle$ and measures the second qubit in the standard basis.
- **(ii)** For *consistency*, the prover computes $\mathcal{A} |\psi\rangle$, selects random $d \in [k]$, and uses $\mathcal{M}_k$ in Hadamard test.

We already know that for a quantum proof of the valid form, we can write the success probability as:

$$C(|\psi\rangle) = \frac{qd + 1}{qd\kappa + 1} - \frac{u_e + u_s}{R \cdot (qd\kappa + 1)}$$

To be able to analyze soundness, all that is left is to reprove Corollary 13 to handle the subtle difference between Claim 21 and Claim 11. Let $D(|\psi\rangle) = \frac{1}{\kappa} \pm d_1$ and $V(|\psi\rangle) = 1 - \frac{1}{\kappa} - d_2$, then we can write:

$$\begin{aligned} C(|\psi\rangle) &= C^{\text{YES}} - \frac{u_e + u_s}{R \cdot (qd\kappa + 1)} \\ &\leq C^{\text{YES}} - \frac{R \cdot (1 - \delta - q\eta)}{R \cdot (qd\kappa + 1)} + \left(\kappa d_1 + (\kappa + 1)\sqrt{\kappa \cdot d_2}\right)^{1/2} . \end{aligned}$$

We can choose a small enough η (via large enough n_0) so that $\eta q < \frac{1-\delta}{2}$.

▶ **Corollary 23.** *If $D(|\psi\rangle) = \frac{1}{\kappa} \pm d_1$ and $V(\psi) = 1 - \frac{1}{\kappa} - d_2$, then:*

$$C(|\psi\rangle) \leq C^{\text{YES}} - \frac{1 - \delta}{2\,(qd\kappa + 1)} + \left(\kappa d_1 + (\kappa + 1)\sqrt{\kappa \cdot d_2}\right)^{1/2}$$

For soundness, the same analysis of Section 3.3 goes through by reducing λ by a factor of 2; this change comes from Corollary 23, which handles the extra $q\eta$ self-loop constraints. All together:

▶ **Theorem 24.** *Consider* $(1,\delta)$*-*GapCSP *with a* $(N = 2^{\mathrm{poly}(n)}, R = 2^{\mathrm{poly}(n)}, q = O(1), \Sigma = \{0,1\})$*-CSP system that is* $\mathrm{polylog}(NR)$*-doubly explicit and* $O(1)$*-strongly uniform. The* QMA^+ *protocol solves this problem with completeness* c *and soundness* s *for some constants* $1 > c > s > 0$*.*

▶ **Corollary 25.** *There exist constants* $1 > c > s > 0$ *such that* $\mathsf{NEXP} \subseteq \mathsf{QMA}^+_{c,s}$*.*

5 Subtle features of QMA^+

5.1 Promise symmetry matters

One can imagine restricting to proofs with non-negative amplitudes *only* in completeness. But this class is equal to QMA:

▶ **Fact 26.** *Consider the class* $\mathsf{QMA}^{+'}$*, where the proof must have non-negative amplitudes* only in completeness*. Since subset states have non-negative amplitudes,* $\mathsf{SQMA} \subseteq \mathsf{QMA}^{+'} \subseteq \mathsf{QMA}$*. By [21],* $\mathsf{SQMA} = \mathsf{QMA}^{+'} = \mathsf{QMA}$*.*

Instead, QMA^+ also restricts the proof in soundness, which reduces the ways Merlin can "deceive" Arthur. This increases the power of the complexity class:

▶ **Corollary 27.** *Notice that* $\mathsf{QMA}^{+'}_{c,s} \subseteq \mathsf{QMA}^+_{c,s}$ *for any choice of* $0 \leq c, s \leq 1$*, since any* $\mathsf{QMA}^{+'}_{c,s}$ *protocol is also a* $\mathsf{QMA}^+_{c,s}$ *protocol. Since* $\mathsf{QMA} = \mathsf{QMA}^{+'}$*,* $\mathsf{QMA} \subseteq \mathsf{QMA}^+_{c,s}$ *whenever* $c < 1$ *and* $c - s \geq \frac{1}{p(n)}$ *for any polynomial* $p(n)$*.*

In general, suppose $\mathcal{R}$ is a set of quantum states that approximate all quantum states (i.e. an ϵ-covering) by at least an inverse polynomial in number of qubits. Then QMA is equal to QMA restricted to $\mathcal{R}$ in completeness, and at most QMA restricted to $\mathcal{R}$ in both completeness and soundness.

Furthermore, classes that modify QMA only in completeness enjoy promise gap amplification through parallel repetition. This does not hold for promise-symmetric modifications. We provide a simple example of how parallel repetition fails to amplify the promise gap of QMA^+:

▶ **Fact 28.** *Consider a Hermitian and positive semidefinite matrix* M*. Let* $\|M\|_+ := \max_{|v\rangle \geq 0} \frac{\||M|v\rangle\|_2}{\||v\rangle\|_2}$ *be the maximum value among real non-negative vectors. Then it is possible for* $\|M \otimes M\|_+ > \|M\|_+^2$*.*

Proof. Consider two qubits and the projector $M = |x_-\rangle\langle x_-|$, where M projects into the Pauli-X basis (i.e. $|x_-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$). Then $\|M\|_+ = \frac{1}{\sqrt{2}}$, maximized at $|0\rangle$ or $|1\rangle$. But using the state $|\chi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, $\|M \otimes M\|_+ \geq \|M|\chi\rangle\| = \frac{1}{\sqrt{2}} > \|M\|_+^2$. ◀

5.2 QMA^+ at some constant gap equals QMA

Perhaps surprisingly, $\mathsf{QMA}^+_{c,s}$ *equals* QMA for some constants $1 > c > s > 0$. This is because every quantum state can be approximated (up to a constant) by a quantum state *without relative phase*:

▶ **Proposition 29.** $\mathsf{QMA}^+_{c,s} \subseteq \mathsf{QMA}_{c,4s}$*.*

Proof. Consider a problem in $\mathsf{QMA}^{+}_{c,s}$, and let Π_1 be its accepting operator. We will use the same circuit in QMA, and analyze the new completeness and soundness.

- *(completeness)* The same completeness proof is a valid proof for QMA, accepting with completeness c.
- *(soundness)* Recall that $\langle\chi|\,\Pi_1\,|\chi\rangle \leq s$ for any $|\chi\rangle$ with non-negative amplitudes. Consider any state $|\psi\rangle$ with *real* (but possibly negative) amplitudes. Separate and normalize its positive entries and negative entries, i.e. $|\psi\rangle = \sqrt{p}\,|\psi_+\rangle - \sqrt{1-p}\,|\psi_-\rangle$. Notice that $\langle\psi_+|\psi_-\rangle = 0$, so $|\psi\rangle$ is of unit norm. Then

$$\begin{aligned}\langle\psi|\,\Pi_1\,|\psi\rangle &= p\,\langle\psi_+|\,\Pi_1\,|\psi_+\rangle + (1-p)\,\langle\psi_-|\,\Pi_1\,|\psi_-\rangle - \sqrt{p(1-p)}(\langle\psi_-|\,\Pi_1\,|\psi_+\rangle + \langle\psi_+|\,\Pi_1\,|\psi_-\rangle)\\ &\leq p\,\langle\psi_+|\,\Pi_1\,|\psi_+\rangle + (1-p)\,\langle\psi_-|\,\Pi_1\,|\psi_-\rangle + 2\sqrt{p(1-p)}\sqrt{\langle\psi_-|\,\Pi_1\,|\psi_-\rangle\,\langle\psi_+|\,\Pi_1\,|\psi_+\rangle}\\ &\leq s + 2s\sqrt{p(1-p)} \leq 2s\,,\end{aligned}$$

  where the first inequality holds by Cauchy-Schwarz because Π_1 is positive semidefinite. Similarly, consider any state with arbitrary amplitudes. Separate and normalize its real and imaginary entries, i.e. $|\phi\rangle = \sqrt{p'}\,|\phi_{\mathbb{R}}\rangle + i\sqrt{1-p'}\,|\phi_{i\mathbb{R}}\rangle$. Notice that $|\phi\rangle$ is still unit norm. By the same calculation, one finds that $\langle\phi|\,\Pi_1\,|\phi\rangle \leq 4s$. ◀

▶ **Corollary 30.** *For any* $0 < \varepsilon < 0.2$, $\mathsf{QMA}^{+}_{0.8+\varepsilon,0.2} = \mathsf{QMA}$.

Proof. $\mathsf{QMA}^{+}_{0.8+\varepsilon,0.2} \subseteq \mathsf{QMA}$ follows from Proposition 29. The other direction follows from Corollary 27. ◀

This is a strange phenomenon: depending on the choice of constants $c > s$, $\mathsf{QMA}^{+}_{c,s}$ could be as small as QMA and as large as NEXP![14] See Figure 2 for a pictorial description. An implication of our work is that assuming $\mathsf{EXP} \neq \mathsf{NEXP}$, QMA^{+} simply *cannot* be amplified.

6 Open questions

1. **What is the relationship of QMA and QMA(2)?** Our result does not immediately say anything about QMA and $\mathsf{QMA}(2)$. It only suggests that for QMA, the restriction of *relative phase* is maximally strong. For example, it is possible that $\mathsf{QMA}(2) = \mathsf{NEXP}$; i.e. the restriction of *entanglement* across a fixed barrier may be just as powerful. In fact, showing $\mathsf{QMA}(2) = \mathsf{QMA}^{+}(2)$ is still an open route to proving $\mathsf{QMA}(2) = \mathsf{NEXP}$, but in light of this work, amplification for $\mathsf{QMA}^{+}(2)$ must crucially rely on the unentanglement promise.
2. **Are other complexity classes sensitive to different constant-sized promise gaps?** We show that for $\mathsf{QMA}^{+}_{c,s}$, the parameter $\frac{c}{s}$ can be "tuned" to change the power of the class from QMA to NEXP (see also Figure 2). Do other complexity classes drastically change power with different promise gaps? One similar class is SBQP [27], which equals $\mathsf{BQP}_{c,s}$ when $\frac{c}{s} = 2$ (but unlike our work, c and s are exponentially small). However, when $\frac{c}{s}$ is allowed to be any number above 1, $\mathsf{BQP}_{c,s}$ is equal to PP [16]. Note that relative to oracles, SBQP is not closed under intersection, which was used to separate it from QMA [3].
3. **State complexity vs. decision complexity?** Although we prove that there are constants c_1, s_1, c_2, s_2 such that $\mathsf{QMA}^{+}_{c_1,s_1} = \mathsf{QMA}^{+}(2)_{c_2,s_2}$, we do not prove the existence of any *product test* (as in [22]). In fact, it is possible that no product test exists! This

[14] Note that the same phenomenon holds for $\mathsf{QMA}^{+}(2)$ and $\mathsf{QMA}(2)$ with nearly the same proof. This is why [24] was perceived as "just a constant gap away" from proving $\mathsf{QMA}(2) = \mathsf{NEXP}$.

could lead to a separation between the complexity of *decision problems* and *state synthesis problems*; i.e. $\mathsf{QMA}^+ = \mathsf{QMA}^+(2)$ but $\mathsf{stateQMA}^+ \neq \mathsf{stateQMA}^+(2)$. In fact, it is even possible that $\mathsf{QMA} = \mathsf{QMA}(2)$ but $\mathsf{stateQMA} \neq \mathsf{stateQMA}(2)$. This inquiry can help us understand whether (or how) *the power of unentanglement* is useful when solving *decision problems*.

References

1 Scott Aaronson. Open problems related to quantum query complexity, 2021. arXiv:2109.06917.

2 Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. The power of unentanglement, 2008. arXiv:0804.0802.

3 Scott Aaronson, Robin Kothari, William Kretschmer, and Justin Thaler. Quantum lower bounds for approximate counting via laurent polynomials, 2020. doi:10.4230/LIPICS.CCC.2020.7.

4 Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, pages 1–17, 2021. arXiv:2003.11673.

5 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. doi:10.1145/278298.278306.

6 Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs; A new characterization of NP. In *33rd Annual Symposium on Foundations of Computer Science, Pittsburgh, Pennsylvania, USA, 24-27 October 1992*, pages 2–13. IEEE Computer Society, 1992. doi:10.1109/SFCS.1992.267824.

7 Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. On the power of nonstandard quantum oracles. In *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*, 2023. doi:10.4230/LIPIcs.TQC.2023.11.

8 Hugue Blier and Alain Tapp. A quantum characterization of np, 2010. arXiv:0709.0738.

9 Anne Broadbent and Eric Culf. Rigidity for monogamy-of-entanglement games, 2023. doi:10.4230/LIPICS.ITCS.2023.28.

10 Samuel Burer. Copositive programming. In *Handbook on semidefinite, conic and polynomial optimization*, pages 201–218. Springer, 2011. URL: https://link.springer.com/chapter/10.1007/978-1-4614-0769-0_8.

11 Jing Chen and Andrew Drucker. Short multi-prover quantum proofs for sat without entangled measurements, 2010. arXiv:1011.0716.

12 Yuanyou Furui Cheng. Explicit estimate on primes between consecutive cubes, 2013. arXiv:0810.2113.

13 Alessandro Chiesa and Michael A Forbes. Improved soundness for qma with multiple provers, 2011. arXiv:1108.2098.

14 John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23(15):880, 1969. doi:10.1103/PhysRevLett.23.880.

15 David Cui, Arthur Mehta, Hamoon Mousavi, and Seyed Sajjad Nezhadi. A generalization of CHSH and the algebraic structure of optimal strategies. *Quantum*, 4:346, October 2020. doi:10.22331/q-2020-10-21-346.

16 Abhinav Deshpande, Alexey V Gorshkov, and Bill Fefferman. Importance of the spectral gap in estimating ground-state energies. *PRX Quantum*, 3(4):040327, 2022. arXiv:2007.11582.

17 Irit Dinur. The pcp theorem by gap amplification. *Journal of the ACM (JACM)*, 54(3):12–es, 2007. URL: https://dl.acm.org/doi/abs/10.1145/1236457.1236459.

18 François Le Gall, Shota Nakagawa, and Harumichi Nishimura. On QMA protocols with two short quantum proofs. *Quantum Inf. Comput.*, 12(7-8):589–600, 2012. doi:10.26421/QIC12.7-8-4.

19 Sevag Gharibian. Strong np-hardness of the quantum separability problem, 2009. arXiv:0810.4507.

20 Sevag Gharibian, Miklos Santha, Jamie Sikora, Aarthi Sundaram, and Justin Yirka. Quantum generalizations of the polynomial hierarchy with applications to qma(2), 2018. doi:10.4230/LIPICS.MFCS.2018.58.

21 Alex B. Grilo, Iordanis Kerenidis, and Jamie Sikora. Qma with subset state witnesses, 2014. doi:10.48550/ARXIV.1410.2882.

22 Aram W. Harrow and Ashley Montanaro. Testing product states, quantum merlin-arthur games and tensor optimization. *Journal of the ACM (JACM)*, 60(1):1–43, 2013. arXiv:1001.0017.

23 Prahladh Harsha. *Robust PCPs of proximity and shorter PCPs*. PhD thesis, Massachusetts Institute of Technology, 2004. URL: https://dspace.mit.edu/bitstream/handle/1721.1/26720/59552830-MIT.pdf.

24 Fernando Granha Jeronimo and Pei Wu. The power of unentangled quantum proofs with non-negative amplitudes. *55th Annual ACM Symposium on Theory of Computing*, 2023. doi:10.1145/3564246.3585248.

25 Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Mip*= re. *Communications of the ACM*, 64(11):131–138, 2021. arXiv:2001.04383.

26 Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. Quantum merlin-arthur proof systems: Are multiple merlins more helpful to arthur?, 2003. arXiv:quant-ph/0306051.

27 Greg Kuperberg. How hard is it to approximate the jones polynomial? *Theory of Computing*, 11(1):183–219, 2015. doi:10.4086/toc.2015.v011a006.

28 Alexander Lubotzky. Finite simple groups of lie type as expanders, 2009. arXiv:0904.3411.

29 Chris Marriott and John Watrous. Quantum arthur-merlin games, 2005. arXiv:cs/0506068.

30 Dominic Mayers and Andrew Yao. Self testing quantum apparatus, 2004. arXiv:0307205.

31 Matthew McKague. On the power quantum computation over real hilbert spaces. *International Journal of Quantum Information*, 11(01):1350001, February 2013. doi:10.1142/s0219749913500019.

32 Anand Natarajan and Tina Zhang. Quantum free games. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1603–1616, 2023. arXiv:2302.04322.

33 Attila Pereszlényi. Multi-prover quantum merlin-arthur proof systems with small gap, 2012. arXiv:1205.2761.

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

35 Adrian She and Henry Yuen. Unitary property testing lower bounds by polynomials, 2022. arXiv:2210.05885.

36 Boris S Tsirelson. Some results and problems on quantum bell-type inequalities. *Hadronic Journal Supplement*, 8(4):329–345, 1993. URL: http://www.math.tau.ac.il/~tsirel/download/hadron.pdf.

37 Jianwei Xu. Quantifying the phase of quantum states, 2023. arXiv:2304.09028.