The Power of Unentangled Quantum Proofs with Non-negative Amplitudes

Fernando Granha Jeronimo
Institute for Advanced Study
Princeton, NJ, USA
granha@ias.edu

Pei Wu
Institute for Advanced Study
Princeton, NJ, USA
pwu@ias.edu

ABSTRACT

Quantum entanglement is a fundamental property of quantum mechanics and it serves as a basic resource in quantum computation and information. Despite its importance, the power and limitations of quantum entanglement are far from being fully understood. Here, we study entanglement via the lens of computational complexity. This is done by studying quantum generalizations of the class NP with multiple *unentangled* quantum proofs, the so-called QMA(2) and its variants. The complexity of QMA(2) is known to be closely connected to a variety of problems such as deciding if a state is entangled and several classical optimization problems. However, determining the complexity of QMA(2) is a longstanding open problem, and only the trivial complexity bounds QMA $\subseteq$ QMA(2) $\subseteq$ NEXP are known.

In this work, we study the power of *unentangled* quantum proofs with *non-negative* amplitudes, a class which we denote $\mathrm{QMA}^+(2)$. In this setting, we are able to design proof verification protocols for (increasingly) hard problems both using *logarithmic* size quantum proofs and having a *constant* probability gap in distinguishing yes from no instances. In particular, we design *global* protocols for small set expansion (SSE), unique games (UG), and PCP verification. As a consequence, we obtain NP $\subseteq \mathrm{QMA}^+_{\log}(2)$ with a constant gap. By virtue of the new *constant* gap, we are able to "scale up" this result to $\mathrm{QMA}^+(2)$, obtaining the full characterization $\mathrm{QMA}^+(2) = \mathrm{NEXP}$ by establishing stronger explicitness properties of the PCP for NEXP. We believe that our protocols are interesting examples of proof verification and property testing in their own right. Moreover, each of our protocols has a single isolated property testing task relying on non-negative amplitudes which if generalized would allow transferring our results to QMA(2).

One key novelty of these protocols is the manipulation of quantum proofs in a *global* and *coherent* way yielding constant gaps. Previous protocols (only available for general amplitudes) are either *local* having vanishingly small gaps or treating the quantum proofs as classical probability distributions requiring polynomially many proofs. In both cases, these known protocols do not imply non-trivial bounds on QMA(2).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
STOC '23, June 20–23, 2023, Orlando, FL, USA
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9913-5/23/06…$15.00
https://doi.org/10.1145/3564246.3585248

CCS CONCEPTS

• **Theory of computation → Quantum complexity theory**.

KEYWORDS

quantum Merlin-Arthur, QMA(2), PCP, NEXP, unique games conjecture, small-set expansion

ACM Reference Format:

Fernando Granha Jeronimo and Pei Wu. 2023. The Power of Unentangled Quantum Proofs with Non-negative Amplitudes. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23), June 20–23, 2023, Orlando, FL, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3564246.3585248

1 INTRODUCTION

Quantum entanglement is a fundamental property of quantum mechanics and it plays a major role in several fields such as quantum computation, information, cryptography, condensed matter physics, etc [28, 42, 43, 51]. Roughly speaking, quantum entanglement is a distinctive form of quantum correlation that is stronger than classical correlations. Entanglement can lead to surprising (and sometimes counter-intuitive) phenomena as presented in the celebrated EPR paradox [23] and the violation of Bell's (style) inequalities [8, 18]. In a sense, entanglement is necessary to access the full power of quantum computation since it is known that quantum computations requiring "little" entanglement can be simulated classically with small overhead [48]. Entanglement is also crucial in a variety of protocols such as quantum key distribution [9], teleportation [10], interactive proof systems [30], and so on. However, despite this central role, the power and limitations of quantum entanglement are far from being understood. Here, we study quantum entanglement via the lens of computational complexity. More precisely, we investigate the role of entanglement in the context of quantum proof verification.

The notions of provers, proofs, and proof verification play a central role in our understanding of classical complexity theory [3]. The quantum setting allows for various and vast generalizations of these classical notions [49]. For instance, by allowing the proof to be a quantum state of polynomial size and the verifier to be an efficient quantum machine, one obtains the class QMA which is a natural generalization of the class NP [50]. The QMA proof verification

model can be further generalized to two quantum proofs from two *unentangled* provers. This generalization gives rise to a class known as QMA(2) [38] (see Definition 2.1). This latter complexity class is known to be closely connected to a variety of computational problems such as the fundamental problem of deciding whether a quantum state (given its classical description) is entangled or not. It is also connected to a variety of classical optimization problems such as polynomial and tensor optimization over the sphere as well as some norm computation problems [25].

Determining the complexity of QMA(2) is a major open problem in quantum complexity. Contrary to many other quantum proof systems (e.g., QIP [29] and MIP* [30]), we still do not know any non-trivial complexity bounds for QMA(2). On one hand, we trivially have QMA $\subseteq$ QMA(2) since a QMA(2) verifier can simply ignore one of the proofs. On the other hand, a NEXP verifier can guess exponentially large classical descriptions of two quantum proofs of polynomially many qubits and simulate the verification protocol classically in exponential time. Hence, we also have QMA(2) $\subseteq$ NEXP. Despite considerable effort with a variety of powerful techniques brought to bear on this question, such as semidefinite programming hierarchies [6, 22, 26], quantum de Finetti theorems [13, 14, 39], and carefully designed nets [15, 47], only the trivial bounds QMA $\subseteq$ QMA(2) $\subseteq$ NEXP are known.

Even though there are no non-trivial complexity bounds for QMA(2), there are results showing surprisingly powerful consequences of *unentangled* proofs. An early result by Blier and Tapp [12] shows that two *unentangled* proofs of a logarithmic number of qubits suffice to verify the NP-complete problem of graph 3-coloring. The version of QMA(2) with logarithmic-size proofs is known as $\mathrm{QMA}_{\log}(2)$. It is know that $\mathrm{QMA}_{\log}(1) \subseteq$ BQP from the work of Marriott and Watrous [41], and this together with their protocol provides some evidence that having two unentangled proofs of logarithmic size is more powerful than having a single one. This suggests that the lack of quantum entanglement across the proofs can play an important role in proof verification. Furthermore, note that this situation is in sharp contrast with the classical setting where having two classical proofs of logarithmic size is no more powerful than having a single one since two proofs can be combined into a larger one.

The above protocol has a critical drawback, namely, the verifier only distinguishes yes from no instances with a polynomially small probability. This distinguishing probability is known as the *gap* of the protocol. These weak gaps are undesirable for two reasons. First, we cannot obtain tighter bounds on QMA(2) from these protocols since scaling up these results to QMA(2) leads to exponentially small gaps. Such tiny gaps fall short to imply NEXP = QMA(2) as the definition of QMA(2) can tolerate up to only polynomially small gaps. Second, the strength of the various hardness results that can be deduced from these protocols depends on how large the gap is. For instance, we do not know if several of these problems are also hard to approximate within say a more robust universal constant. A series of subsequent works extended Blier and Tapp's result in the context of $\mathrm{QMA}_{\log}(2)$ protocols for NP-complete problems [7, 17, 24]. However, all these protocols suffer from a polynomially small gap.

Another piece of evidence pointing to the additional power of unentangled proofs appears in the work of Aaronson et al. [1]. They show that $\widetilde{O}(\sqrt{n})$ quantum proofs of logarithmic size suffice to decide an NP-complete variant of the SAT problem of size n with a constant gap. Due to the work of Harrow and Montanaro [25], it is possible to convert this protocol into a two-proof protocol where each one has size $\widetilde{O}(\sqrt{n})$ and the gap remains constant. Unfortunately, this converted protocol does not imply tighter bounds for QMA(2) since it only shows NP $\subseteq$ QMA(2).

In this work, we study *unentangled* quantum proofs with *non-negative* amplitudes. We name the associated complexity classes introduced here as $\mathrm{QMA}^+(2)$ and $\mathrm{QMA}^+_{\log}(2)$ (see Section 2.1) in analogy to QMA(2) and $\mathrm{QMA}_{\log}(2)$, respectively. The main question we consider is the following:

What is the power of *unentangled* proofs with *non-negative* amplitudes?

This non-negative amplitude setting is intended to capture several structural properties of the general QMA(2) model while providing some restriction on the adversarial provers in order to gain a better understanding of unentangled proof verification. In this non-negative amplitude setting, we are able to derive much stronger results and fully characterize $\mathrm{QMA}^+(2)$. In particular, we are able to design $\mathrm{QMA}^+_{\log}(2)$ protocols with *constant* gaps for (increasingly) hard(er) problems. Each of these protocols contributes to our understanding of proof verification and leads to different sets of techniques, property testing routines, and analyses.

Our first protocol is for the small set expansion (SSE) problem [4, 45]. Roughly speaking, the SSE problem asks whether all small sets of an input graph are very expanding[1] or if there is a small non-expanding set. The SSE problem arises in the context of the unique games (UG) conjecture. This conjecture plays an important role in the classical theory of hardness of approximation [31–33, 36, 37, 44]. One key reason is that the unique games problem is a (seemingly) more structured computational problem as opposed to more general and provably NP-hard constraint satisfaction problems (CSPs) making it easier to reduce UG to other problems. In this context, the SSE problem is considered an even more structured problem than UG since some of its variants can be reduced to UG. This extra structure of SSE compared to UG can make it even easier to reduce SSE to other problems. So far the hardness of SSE remains an open problem —it has evaded the best known algorithmic techniques [46].

Theorem 1.1 (Informal). *Small set expansion is in* $\mathrm{QMA}^+_{\log}(2)$ *with a constant gap.*

Our second protocol is for the unique games problem. The UG problem is a special kind of CSP wherein the constraints are permutations and it is enough to distinguish almost fully satisfiable instances from those that are almost fully unsatisfiable. The fact that the constraints of a UG instance are bijections which in turn can be implemented as valid (i.e., unitary operators) is explored in our protocol. Although the hardness of UG remains an open problem, a weaker version of the UG problem was recently proven

[1] In terms of edge expansion.

to be NP-hard [5, 20, 35]. From our UG protocol and this weaker version of the problem, we obtain $\mathrm{NP} \subseteq \mathrm{QMA}^+_{\log}(2)$ with a *constant* gap (see Corollary 1.3 below).

Theorem 1.2 (Informal). *Unique Games is in* $\mathrm{QMA}^+_{\log}(2)$ *with a constant gap protocol.*

A key novelty of our protocols is their *global* and *coherent* manipulation of quantum proofs leading to *constant* gaps. The previous protocols for $\mathrm{QMA}_{\log}(2)$ with a logarithmic proof size are *local* in the sense that they need to read *local* information[2] from the quantum proofs thereby suffering from vanishingly small gaps. Furthermore, the previous protocol with a constant gap treats the quantum proofs as classical probability distributions (e.g., relying on the birthday paradox) and this classical treatment ends up requiring polynomially many proofs to achieve the constant gap.

Another interesting feature of our protocols is that they already almost work in the general amplitude case in the sense that each protocol isolates a single property testing task relying on non-negative amplitudes. If such a property testing task can be generalized to general amplitudes, then the corresponding protocol works in $\mathrm{QMA}_{\log}(2)$ as well.

As discussed earlier, by Theorem 1.2 together with the work on the 2-to-2 conjecture, we obtain that NP is contained in $\mathrm{QMA}^+_{\log}(2)$ with a *constant* gap.

Corollary 1.3 (Informal). $\mathrm{NP} \subseteq \mathrm{QMA}^+_{\log}(2)$ *with a constant gap.*

By virtue of the *constant* gaps of our protocols for $\mathrm{QMA}^+_{\log}(2)$, we can "scale up" our results to give an exact characterization of $\mathrm{QMA}^+(2)$ building on top of ideas of very efficient classical PCP verifiers.

Theorem 1.4. $\mathrm{QMA}^+(2) = \mathrm{NEXP}$.

The characterization above is proven by designing a *global* $\mathrm{QMA}^+(2)$ protocol for NEXP. To design this *global* protocol, we not only rely on the properties of the known efficient classical PCP verification for NEXP, but we need additional explicitness and regularity properties. Regarding the explicitness, we call *doubly explicit* the kind of PCP required in our *global* protocol (in analogy to the terminology of graphs). Roughly speaking, doubly explicitness means that we can very efficiently not only determine the variables appearing in any given constraint, but also reverse this mapping by very efficiently determining the constraints in which a variable appears. Here, we prove that these properties can be indeed obtained by carefully combining known PCP constructions.

An intriguing next step is to explore the improved understanding of the unentangled proof verification from our protocols in the general amplitude case. Investigating problems like SSE and UG might provide more structure towards this goal. Characterizing the complexity of QMA(2) would be extremely interesting whatever this characterization turns out to be.

[2]Roughly speaking, they treat a quantum proof as quantum random access codes that encodes n bits using $\log_2(n)$ qubits. By Nayak's bound the probability of recovering a queried position is polynomially small in n.

Organization. This document is organized as follows. In Section 3, we give an overview of our global protocols. In Section 2, we formally define $\mathrm{QMA}^+(2)$ and its variants as well as fix some notation and recall basic facts. In Section 4, we develop some quantum property testing primitives that will be common to our protocols. In Section 5, we present our global protocol for SSE. In Section 6, we present our global protocol for UG and we use it to prove $\mathrm{NP} \subseteq \mathrm{QMA}^+_{\log}(2)$ with a constant gap. In Section 7, we prove the characterization $\mathrm{QMA}^+(2) = \mathrm{NEXP}$.

2 PRELIMINARIES

Let $\mathbb{N}, \mathbb{R}, \mathbb{C}$ stand for the natural, real, and complex numbers. $\mathbb{N}^+$ denotes the positive natural number. For any real number x,

$$\mathrm{sgn}(x) = \begin{cases} 1 & x > 0; \\ 0 & x = 0; \\ -1 & x < 0. \end{cases}$$

In this paper, log stands for the logarithm to base 2. We adopt both the Dirac notation and the usual notation of vectors (whichever seems more appropriate) as we consider both quantum and classical objects. For $p \in [1, \infty)$, we denote the ℓ_p-norm of $u \in \mathbb{C}^n$ as $\|u\|_p$, i.e., $\|u\|_p = \left(\sum_{i=1}^n |u_i|^p\right)^{1/p}$. We omit the subscript for the ℓ_2-norm, i.e., $\|u\| := \|2\|_2$. We denote the ℓ_∞-norm of $u \in \mathbb{C}^n$ as $\|u\|_\infty$, i.e., $\|u\|_\infty = \max_{i\in[n]} |u_i|$. Let $\mathbb{S}_n := \{u \in \mathbb{C}^{n+1} : \|u\| = 1\}$ be the n-dimensional sphere and $\mathbb{S}^+_n := \{u \in (\mathbb{R}_{\geq 0})^{n+1} : \|u\| = 1\}$ be the intersection of the n-dimensional sphere and the non-negative orthant. The subscript will almost always be omitted in this manuscript since it can be confusing and the dimension is normally clear from the context. Adopt the short-hand notation $[n] = \{1, 2, \ldots, n\}$. For any universe U and any subset $S \subseteq U$, let $\overline{S} := U \setminus S$. Denote the characteristic vector of S by $\mathbf{1}_S$, i.e., $\mathbf{1}_S \in \mathbb{R}^U$ and

$$\mathbf{1}_S(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For a logical condition C, we use the Iverson bracket

$$\mathbb{1}[C] = \begin{cases} 1 & \text{if } C \text{ holds}, \\ 0 & \text{otherwise.} \end{cases}$$

2.1 Quantum Merlin-Arthur with Multiple Provers

The class $\mathrm{QMA}(k)$ can be formally defined in more generality as follows.

Definition 2.1 ($\mathrm{QMA}_\ell(k, c, s)$). *Let $k\colon \mathbb{N} \to \mathbb{N}$ and $c, s, \ell\colon \mathbb{N} \to \mathbb{R}^+$ be polynomial time computable functions. A promise problem $\mathcal{L}_{\mathrm{yes}}, \mathcal{L}_{\mathrm{no}} \subseteq \{0,1\}^*$ is in $\mathrm{QMA}_\ell(k, c, s)$ if there exists a* BQP *verifier V such that for every $n \in \mathbb{N}$ and every $x \in \{0,1\}^n$,*

Completeness: *If $x \in \mathcal{L}_{\mathrm{yes}}$, then there exist unentangled states $|\psi_1\rangle, \ldots, |\psi_{k(n)}\rangle$, each on at most $\ell(n)$ qubits, s.t.* $\Pr[V(x, |\psi_1\rangle \otimes \cdots \otimes |\psi_{k(n)}\rangle) \text{ accepts}] \geq c(n)$.

Soundness: *If $x \in \mathcal{L}_{\mathrm{no}}$, then for every unentangled states $|\psi_1\rangle, \ldots, |\psi_{k(n)}\rangle$, each on at most $\ell(n)$ qubits, we have* $\Pr[V(x, |\psi_1\rangle \otimes \cdots \otimes |\psi_{k(n)}\rangle) \text{ accepts}] \leq s(n)$.

Harrow and Montanaro proved that: For any state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \cdots \otimes \mathbb{C}^{d_k}$, if

$$\max_{\phi_i \in \mathbb{C}^{d_i}} \langle \psi \mid \phi_1 \phi_2 \dots \phi_k \rangle = 1 - \varepsilon,$$

then the *product test* rejects $|\psi\rangle^{\otimes 2}$ with probability $\Omega(\varepsilon)$. Based on this product test, Harrow and Montanaro further showed in the QMA protocols, the number of provers can always be reduced to 2.

Theorem 2.2 (Harrow and Montanaro [25]). *For any $\ell, k, 0 \le s < c \le 1$,*

$$\mathrm{QMA}_\ell(k, c, s) \subseteq \mathrm{QMA}_{k\ell}(2, s', c'),$$

where $c' = (1+c)/2$ and $s' = 1 - (1-s)^2/100$.

The class $\mathrm{QMA}^+_\ell(k, c, s)$ is defined exactly the same way, except that the proofs $|\psi_1\rangle, \dots, |\psi_k\rangle$ should have real, non-negative amplitudes. In our work, we are only interested in

$$\mathrm{QMA}^+_{\log}(2) := \bigcup_{c-s=\Omega(1)} \mathrm{QMA}^+_{O(\log n)}(2, c, s),$$
$$\mathrm{QMA}^+(2) := \bigcup_{i \in \mathbb{N},\ c-s=\Omega(1)} \mathrm{QMA}^+_{O(n^i)}(2, c, s).$$

Instead of having only 2 provers, it is much more convenient to consider k provers for some large constant k. This is without loss of generality, as Theorem 2.2 generalizes to QMA^+ as well. As a result, as long as the $\mathrm{QMA}^+(k, c, s)$ protocol is such that $c > 1-(1-s)^2/50$, it can be converted back to a $\mathrm{QMA}^+(2)$ protocol with a constant gap. The condition that $c > 1-(1-s)^2/50$ is also not much of an issue, since by a repetition involving more provers, we can amplify any constant (c, s) gap to a $(1-\varepsilon, \delta)$ gap for ε, δ close to 0. In the remainder of the paper, we will use constantly many proofs without further referring to this result.

2.2 Trace Distances

A standard notion of distance for quantum states is that of the *trace distance*. The trace distance between $|\psi\rangle$ and $|\phi\rangle$, denoted $\mathrm{D}(|\psi\rangle, |\phi\rangle)$, is

$$\frac{1}{2}\mathrm{Tr}\sqrt{(|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|)^2}.$$

The following fact provides an alternative definition for trace distance.

Fact 2.3. *The trace distance between $|\phi\rangle$ and $|\psi\rangle$ is given by $\mathrm{D}(|\phi\rangle, |\psi\rangle) = \sqrt{1 - |\langle\phi \mid \psi\rangle|^2}$.*

The trace distance remains small under the tensor product.

Fact 2.4. *Let $|\psi_0\rangle, |\phi_0\rangle \in \mathbb{S}_n$ and $|\psi_1\rangle, |\phi_1\rangle \in \mathbb{S}_m$ for arbitrary $n, m \in \mathbb{N}$. Then*

$$\mathrm{D}(|\psi_0\rangle \otimes |\psi_1\rangle, |\phi_0\rangle \otimes |\phi_1\rangle)^2 \le \mathrm{D}(|\psi_0\rangle, |\phi_0\rangle)^2 + \mathrm{D}(|\psi_1\rangle, |\phi_1\rangle)^2.$$

Two states with small trace distance are indistinguishable to quantum protocols.

Fact 2.5. *If a quantum protocol accepts a state $|\phi\rangle$ with probability at most p, then it accepts $|\psi\rangle$ with probability at most $p + \mathrm{D}(|\phi\rangle, |\psi\rangle)$.*

We will use the well-known swap test to compare unentangled quantum states.

Fact 2.6 (Swap Test). *Let $|\phi\rangle$ and $|\psi\rangle$ be two quantum states on the same Hilbert space. Then the acceptance probability of SwapTest$(|\phi\rangle, |\psi\rangle)$ is $\frac{1}{2} + \frac{|\langle\phi \mid \psi\rangle|^2}{2}$.*

We can equivalently state the acceptance probability of the swap test in terms of the trace distance as follows.

Remark 2.7. *Any two quantum states $|\phi\rangle$ and $|\psi\rangle$ pass the swap test with probability $1 - \frac{1}{2}\mathrm{D}(|\phi\rangle, |\psi\rangle)^2$.*

We record the following elementary facts. They are special cases of trace distance made explicit in the inner product language.

Claim 2.8. *Let $u, v, z \in \mathbb{S}^+_d$ for any natural number d. Let $\varepsilon > 0$ be some small real constant.*

(i) *(Closeness preservation) If $\langle u, v\rangle^2 \ge 1 - \varepsilon$. Then*
$$\left|\langle u, z\rangle^2 - \langle v, z\rangle^2\right| \le 3\sqrt{\varepsilon}.$$

(ii) *(Triangle inequality) If $\langle u, z\rangle^2 \ge 1 - \varepsilon$, and $\langle v, z\rangle^2 \ge 1 - \varepsilon$. Then*
$$\langle u, v\rangle^2 \ge 1 - 2\varepsilon.$$

2.3 Expander Graphs

Let $G = (V, E)$ be a d-regular graph. For non-empty sets $S, T \subseteq V$, we denote by $E(S, T)$ the following set of edges $E(S, T) = \{(x, y) \in E \mid x \in S, y \in T\}$.[3] The edge expansion of a non-empty $S \subseteq V$, denoted $\Phi_G(S)$, is defined as

$$\Phi_G(S) := \frac{|E(S, V \setminus S)|}{d\,|S|},$$

and it is a number in the interval $[0, 1]$. For $S \subseteq V$, we refer to relative size $|S|/|V|$ as the *measure* of S. A closely related notion called Cheeger constant for G, is defined as

$$\min_{S \subseteq G : |S| \le |G|/2} \frac{|E(S, V \setminus S)|}{|S|}.$$

3 OVERVIEW OF GLOBAL PROTOCOLS

We now give an overview of our *global* protocols for SSE in Section 3.1, for UG in Section 3.2 and for NEXP in Section 3.3. As alluded earlier, a key insight of these protocols is the manipulation of quantum proofs in a *global* and *coherent* way in order to achieve a *constant* gap. For the problems considered here, there is always an underlying graph to the problem whose edge set can be (or almost) decomposed into perfect matchings. Taking advantage of this collection of perfect matchings will be one of the aspects in allowing for a *global* manipulation of the quantum proofs in these protocols. It will be more convenient to design protocols with constantly many unentangled proofs rather than just two. Recall that due to the result of Harrow and Montanaro [25], these protocols can be converted into two-proof protocols with a constant multiplicative increase in the proof size.

[3] The graphs are usually undirected. In this case, $E(S, S)$ actually counts the same edge twice by the definition.

3.1 Small Set Expansion Protocol

We provide an overview of the SSE protocol in $\mathrm{QMA}^{+}_{\log}(2)$ with a *constant* gap from Section 5. Suppose that we are given an input n-vertex graph G on the vertex set V. Our goal is to decide whether G is a yes or no instance of (η, δ)-SSE. Recall that, in the yes case, there exists a set S of measure δ, such that S essentially does not expand, i.e., $\Phi_G(S) \leq \eta \approx 0$. Nonetheless, in the no case, every set S of measure at most δ has near-perfect expansion, i.e., $\Phi_G(S) \geq 1 - \eta \approx 1$.

In the design of the protocol, we are allowed two *unentangled* proofs on $O_{\eta,\delta}(\log(n))$ qubits. It is natural to ask for one of these proofs to be a state $|\psi\rangle$ "encoding" a uniform superposition of elements of a purported non-expanding set S of the form

$$|\psi\rangle \;=\; \frac{1}{\sqrt{S}} \sum_{i \in S} |i\rangle \,.$$

We now check the non-expansion of the support set of $|\psi\rangle$ as follows. Suppose we could apply the adjacency matrix A of G directly to the vector $|\psi\rangle$. While A is not necessarily a valid quantum operation, it will not be difficult to resolve this issue later. If we are in the yes case and the support of $|\psi\rangle$ indeed encodes a non-expanding set, we would have $\mathrm{supp}(A|\psi\rangle) \cap \mathrm{supp}(|\psi\rangle) \approx \mathrm{supp}(|\psi\rangle)$. However, if we are in the no case, provided the size of $\mathrm{supp}(|\psi\rangle)$ is small (at most a δ fraction of the vertices), the small set expansion property of G would imply $\mathrm{supp}(A|\psi\rangle) \cap \mathrm{supp}(|\psi\rangle) \approx \emptyset$.

How can we check the support conditions above? For this, suppose that we have not only one copy of $|\psi\rangle$ but rather two equal *unentangled* copies $|\psi_1\rangle = |\psi_2\rangle$. We apply A to $|\psi_1\rangle$ and then measure the correlation between $A|\psi_1\rangle$ and $|\psi_2\rangle$. In the yes case, the two vectors are almost co-linear, whereas in the no case they are almost orthogonal. It is well-known that co-linearity versus orthogonality of two *unentangled* quantum states can be tested via the swap test.

We now address the issue that the adjacency matrix A may not be a unitary matrix, and hence it is not necessarily a valid quantum operation. Nonetheless, the adjacency matrix of a d-regular graph can always be written as a sum of d permutation matrices $P_1, \ldots, P_d$, which are special unitary matrices. In terms of the support guarantees described above, it is possible to show that applying one of these permutation matrices uniformly at random in the protocol leads to a similar behavior as applying A.

In the yes case, it can be shown that all goes well with the above strategy. However, in the no case, things become more delicate starting with the fact that $|\psi\rangle$ is an arbitrary adversarial state of the form

$$|\psi\rangle = \sum_{i \in S} \alpha_i |i\rangle \,,$$

where we have no control over the amplitudes α_i's magnitudes and phases.

One important issue is that the support of $|\psi\rangle$ may not be small (i.e., at most a δ fraction), and the graph G may have large non-expanding sets even in the no case. We design a sparsity test to enforce that its support is indeed small. The soundness of this sparsity test takes advantage of the non-negative amplitudes assumption to achieve dimension-independent parameters and this is the only test of the protocol that rely on the non-negative assumption. This points to a very natural question in quantum property testing: how efficiently can we test sparsity[4] with the help of a prover in the general amplitude case?

In our protocol, the support conditions from above are actually checked by considering the average magnitude of the overlap between $P_r|\psi\rangle$ and $|\psi\rangle$. This overlap governs (part of) the acceptance probability of the protocol which can be bounded as

$$\mathop{\mathbb{E}}_{r \in [d]} \left[\left| \langle P_r \psi \,|\, \psi \rangle \right| \right] \leq \frac{1}{d} \sum_{i,j} A_{i,j} \left| \alpha_i \right| \left| \alpha_j \right| = \frac{1}{d} \left\langle A \left| \psi \right| \,\middle|\, \left| \psi \right| \right\rangle \,,$$

where $||\psi|\rangle = \sum_{i \in S} |\alpha_i| \, |i\rangle$. With this bound, phases are no longer relevant.

A second important and more delicate issue is that the magnitude of the amplitudes α_i's of $|\psi\rangle$ may be very far from flat. By definition, the SSE property of the graph G only states that for every "flat" indicator vector $\mathbf{1}_S$, where S is any vertex set of measure at most δ, we have

$$\frac{1}{d} \left\langle A \frac{\mathbf{1}_S}{\sqrt{|S|}} \,\middle|\, \frac{\mathbf{1}_S}{\sqrt{|S|}} \right\rangle \;\approx_{\eta,d}\; 0 \,.$$

Nonetheless, in order to not be fooled by the provers, we need a stronger *analytic* condition

$$\max_{u \colon \|u\|_2 = 1, |\mathrm{supp}(u)| \leq \delta |V|} \frac{1}{d} \left\langle Au \,|\, u \right\rangle \approx 0 \,,$$

where u ranges over arbitrary unit vectors. For every disjoint set $S, T \subseteq V$ of combined measure at most δ, the SSE property of G allows us to deduce

$$\frac{1}{d} \left\langle A \frac{\mathbf{1}_S}{\sqrt{|S|}} \,\middle|\, \frac{\mathbf{1}_T}{\sqrt{|T|}} \right\rangle \;\approx_{\eta,d}\; 0 \,. \tag{3.1}$$

Ideally, we would like to leverage the bounds we have for flat indicator vectors of small sets from (3.1) to conclude that arbitrary unit vectors of small support have a bounded quadratic form. The seminal work on 2-lifts [11] of Bilu and Linial dealt with a similar question, but without the sparse support conditions. Surprisingly, they gave sufficient conditions for this phenomenon. Here, we prove that the same phenomenon also happens for the sparse version of the problem. In particular, this shows that SSE graphs satisfy the more "robust" *analytic* SSE property. Using this robust property, we conclude the soundness of the protocol.

3.2 Unique Games Protocol

We provide an overview of the UG protocol in $\mathrm{QMA}^{+}_{\log}(2)$ with a *constant* gap from Section 6. Suppose that we are given an input UG instance with alphabet Σ, namely, an n-vertex d-regular graph $G = (V, E)$, where each directed[5] edge $e \in E$ is associated with a permutation constraint $f_e \colon \Sigma \to \Sigma$. We say that an assignment $\ell \colon V \to \Sigma$ satisfies an edge $e = (i, j)$ if $f_e(\ell(i)) = \ell(j)$. This means that for each assigned value for i there is a unique value for j and

[4] For this task, we can have multiple unentangle copies of the state to be tested as well multiple unentangle proofs to help the tester.

[5] The reverse edge of e is typically associated with the constraint f_e^{-1}.

vice-verse satisfying the permutation constraint of edge e. The goal is to distinguish between (yes) there exists an assignment satisfying at least $1-\eta$ fraction of the constraints, and (no) every assignment satisfies at most a δ fraction of constraints.

In the yes case, the protocol expects from the unentangled provers copies of a quantum state $|\psi\rangle$ encoding an assignment ℓ of value at least $1-\eta$ of the form

$$|\psi\rangle = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} |i\rangle|\ell(i)\rangle\,. \tag{3.2}$$

We will again explore the underlying graph structure of the problem to make the proof verification *global* leading to a constant gap. Similarly to the SSE protocol, we will also use the fact that the adjacency matrix A of a d-regular graph can be written as a sum of d permutation matrices $P_1, \ldots, P_d$ and these matrices are special cases of unitary operators. Using a permutation matrix P_r and the UG constraints, we will define a unitary operator Π_r intended to help us check the constraints along the edges of P_r. Each Π_r is defined as follows

$$\Pi_r|i\rangle|a\rangle \mapsto (P_r|i\rangle)\,|f_{(i,P_r i)}(a)\rangle\,,$$

where i ranges in V and a ranges in Σ. The crucial observation is that if the constraints along the edges of P_r are almost fully satisfied by ℓ, we should have $|\psi\rangle \approx \Pi_r|\psi\rangle$ whereas if they almost fully unsatisfied by ℓ, we should have $|\psi\rangle$ almost orthogonal to $\Pi_r|\psi\rangle$. By sampling a uniformly random Π_r and checking this approximate co-linearity versus orthogonality property, we obtain a *global* test to check if an assignment is good.

In the no case, there is no reason the adversarial provers will provide proofs of the form (3.2) encoding a valid assignment. In general, we will have an arbitrary state of the form

$$|\psi\rangle = \sum_{i=1}^{n} \alpha_i |i\rangle \left(\sum_{a\in\Sigma} \beta_{i,a}|a\rangle \right).$$

There are two main issues. First, the adversary can omit the assignment to several vertices by making $\alpha_i \approx 0$. Second, even if all the vertices are present in the superposition with amplitudes $\alpha_i = 1/\sqrt{n}$, the prover can assign a superposition of multiple values to each position as in

$$|\psi\rangle = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} |i\rangle \left(\sum_{a\in\Sigma} \beta_{i,a}|a\rangle \right).$$

Fortunately, both of these issues can be handled in a global way. In addressing the second issue, we currently rely on the non-negative amplitudes assumption. To give a flavor of why non-negative amplitudes can be helpful, consider the following simplified scenario that $\Sigma = \{0,1\}$ and

$$|\psi\rangle = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} |i\rangle \left(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \right).$$

Suppose that we measure the second register (containing the values in Σ) of two copies of $|\psi\rangle$ obtaining $|0\rangle$ and $|1\rangle$, and let $|\psi_0\rangle$ and $|\psi_1\rangle$ be the resulting states on the first register containing the indices of the vertices, respectively. In the ideal scenario, if each vertex has a single well defined value in $|\psi\rangle$ (which is not the case in this example), we should have $|\psi_0\rangle \perp |\psi_1\rangle$. If not (as in this toy example), the supports of $|\psi_0\rangle$ and $|\psi_1\rangle$ are not disjoint. With non-negative amplitudes, if there is substantial "mass" in the intersection of their supports, then this condition can be tested using a swap test since $\langle\psi_0\,|\,\psi_1\rangle$ will be large (in this toy example it is 1 as $|\psi_0\rangle = |\psi_1\rangle = \sum_{i=1}^{n} 1/\sqrt{n}|i\rangle$).

With this UG protocol and the recent proof[6] of the NP-hardness of deciding UG with parameters $\eta = 1/2$ and $\delta > 0$ an arbitrarily small chosen constant, we can deduce that $\mathrm{NP} \subseteq \mathrm{QMA}^{+}_{\log}(2)$.

3.3 PCP Verification Protocol for NEXP

We provide an overview of the NEXP protocol in $\mathrm{QMA}^{+}(2)$ with *constant* gap from Section 7. Recall that scaling up to $\mathrm{QMA}(2)$ the previous protocols for $\mathrm{QMA}_{\log}(2)$ from literature leads to exponentially small gaps which are intolerable to $\mathrm{QMA}(2)$. This motivates our study of *constant* gap protocols for hard problems in $\mathrm{QMA}^{+}_{\log}(2)$. Our new constant gap protocols can be indeed scaled up to $\mathrm{QMA}^{+}(2)$ and the gap remains constant! Another issue unresolved in the previous work is that if we scale up the protocol naively, the running time of the verifier becomes exponential and this is also intolerable to $\mathrm{QMA}(2)$ (or $\mathrm{QMA}^{+}(2)$) which requires a polynomial-time BQP verifier. Simultaneously achieving a constant gap with a polynomial-time verifier is quite interesting since this requires considering very efficient forms of quantum proof verification.

Classically, it is known that NEXP admits polynomial-time proof verification protocols with a constant gap, i.e., very efficient PCPs. Note that the proof size is exponentially large in the input size and the verification runs in *polylogarithmic* time in the size of the proof. These protocols manipulate exponentially large objects given in very succinct and explicit forms. We will build on some of these PCPs results to design our $\mathrm{QMA}^{+}(2)$ protocol for NEXP, but our *global* verification of quantum proofs will require even stronger explicitness and regularity properties of these objects. In this work, we prove these additional properties by carefully investigating the composition of known PCP constructions.

A PCP protocol naturally gives rise to a label cover CSP (via a simple and standard argument). We give a *global* $\mathrm{QMA}^{+}(2)$ protocol for label cover arising from the PCP for NEXP with the additional explicitness and regularity properties alluded above. Recall that a label cover instance is given by a bipartite graph $G = (L \sqcup R, E)$ with a left and right vertex partitions L and R, left and right alphabets Σ_L and Σ_R and constraints $f_e \colon \Sigma_L \to \Sigma_R$ on the edges $e \in E$. Given assignments to the left and right partitions $\ell_L \colon L \to \Sigma_L$ and $\ell_R \colon R \to \Sigma_R$, a constraint on edge $e = (i,j)$ is satisfied if $f_e(\ell_L(i)) = \ell_R(j)$. In this correspondence of PCP and label cover, the left vertices correspond to the constraints of the PCP verifier and the right vertices correspond to the symbols of the proof which are the variables in the PCP constraints.

We now give an abstract simplified description of our protocol to convey some intuition and general ideas. The precise protocol is actually more involved and somewhat different (see Section 7 for its full description). In the yes case our $\mathrm{QMA}^{+}(2)$ protocol expects

[6] Coming from the proof of the 2-to-2 conjecture.

to receive copies of the state $|\psi_L\rangle$ and from it obtain copies of a state similar to $|\psi_R\rangle$ both described below

$$|\psi_L\rangle = \sum_{i\in L} \frac{1}{\sqrt{|L|}} |i\rangle|\ell_L(i)\rangle \qquad \text{and} \qquad |\psi_R\rangle = \sum_{j\in R} \frac{1}{\sqrt{|R|}} |j\rangle|\ell_R(j)\rangle. \tag{3.3}$$

Note that the left assignment ℓ_L specifies the values of all variables appearing in each PCP constraint, and ℓ_R specifies the values of variables appearing in the PCP proof. In this case, checking the constraints (essentially) amounts to testing consistency of these various assignments to the variables. To accomplish this goal, we design two operations[7] Γ_L and Γ_R such that,[8] if the label cover instance is fully satisfiable (with ℓ_L and ℓ_R), then $\Gamma_L(|\psi_L\rangle) \approx \Gamma_R(|\psi_R\rangle)$, otherwise $\Gamma_L(|\psi_L\rangle)$ will be approximately orthogonal to $\Gamma_R(|\psi_R\rangle)$. In a vague sense, Γ_L tries to extract the value of some variables in the constraints and Γ_R tries to replicate the values of each variable in a quantum superposition so that $\Gamma_L(|\psi_L\rangle)$ and $\Gamma_R(|\psi_R\rangle)$ become equal if ℓ_L, ℓ_R are fully satisfying assignments and they become close to orthogonal if the CSP instance is far from satisfiable (regardless of ℓ_L, ℓ_R). At a high level, there is some parallel[9] with the SSE and UG protocols. There, we had $|\psi_L\rangle = |\psi_R\rangle$, Γ_L being the identity and Γ_R being either P_r (in SSE) or Π_r (in UG).

A crucial point is that to make the operations Γ_L and Γ_R efficient, we need to be able to determine (1) the neighbors of any given vertex in L in polynomial time, and (2) the neighbors of any given vertex in R in polynomial time. We call an instance satisfying (1) and (2) *doubly explicit*. While (1) follows easily from the definition of PCP, to get property (2) we need to carefully compose known PCP protocols and prove that this property holds.

Similarly to the UG protocol, we also need to check that the quantum proofs are close to a valid encoding of an assignment to the variables. The provers should not (substantially) omit the values of variables nor provide a superposition of multiple values for the same variable. Similarly, checking this second condition is the part of the protocol that currently relies on non-negative amplitudes.

4 PROPERTY TESTING PRIMITIVES

In this section, we prove some property testing primitives that we will use as the building blocks in designing protocols for general problems.

The first test is the *symmetry* test. In many situations, we ask the prover to provide a supply of constantly many copies of a state. To make sure that all copies are approximately the same state, the symmetry test will be invoked. The symmetry test in general can be applied in any quantum protocol. A similar symmetry test has been considered previously in [1]. Here we provide a stronger version.

The second test is the *sparsity* test. Consider the scenario where we ask the prover to provide a state that is supposed to be some *subset state*. In particular, let $\mathcal{S}_\gamma \subseteq \mathbb{C}^n$ be the set of subset state spanning a γ fraction of computational basis, i.e.,

$$\mathcal{S}_\gamma := \left\{ \frac{1}{\sqrt{\gamma n}} \sum_{i\in S} |i\rangle : S \subseteq [n], |S| = \gamma n \right\}.$$

We call γ the *sparsity* of the subset state in $\mathcal{S}_\gamma$. The sparsity test is used to determine whether a given state is close to $\mathcal{S}_\gamma$. Our sparsity test relies on the fact that the amplitudes of the quantum proofs are non-negative.

The third test is the *validity* test. A natural quantum proof for many problems like the 3-SAT or 3COLOR problem is to put the variables/vertices together with their values/colors in superpositions. For example, for 3-SAT on n variables, such that variable i has value x_i, a valid proof should look like

$$|\phi\rangle = \frac{1}{\sqrt{n}} \sum_{i\in[n]} |i\rangle|x_i\rangle.$$

This can be generalized for an arbitrary set of variables X and an arbitrary value domain Σ of the variables. Then the valid set would be

$$\mathcal{V} = \left\{ \frac{1}{\sqrt{|X|}} \sum_{i\in X} |i\rangle|x_i\rangle : \forall i \in X, x_i \in \Sigma \right\}.$$

The validity test tells whether a given state is close to a valid state. Our validity test works only in the situation when the given state is close to a state in $\mathcal{S}_{|\Sigma|^{-1}}$, which is guaranteed by the sparsity test. Thus, this validity test does not generalize.

4.1 ε-Tilted States

Before we discuss the tests, let's make the following definition first.

Definition 4.1 (ε-tilted states)**.** *A family of states $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle$ defined on a same space is an ε-tilted state if there is a subset $R \subseteq [k]$ such that $|R| \geq (1-\varepsilon)k$ and for any $i, j \in R$,*

$$\mathrm{D}(|\psi_i\rangle, |\psi_j\rangle) \leq \sqrt{\varepsilon}.$$

Furthermore, we call $|\psi_i\rangle$ a representative state *for any $i \in R$, and the subset $\{|\psi_i\rangle : i \in R\}$ the* representative set.

Note that a 0-tilted state is simply a set of equal states, and any ε-tilted state is also a δ-tilted state for any $\delta > \varepsilon$. The name ε-tilted state may be confusing. Our message is that instead of treating this object as a set of states, we should simply treat them as a single state conceptually (for example, think of it as a representative state tilted a little bit). As we will see later in Section 4.2, when the symmetry test passes, we are supplied with an ε-tilted state with high probability. Having a large number of (almost) equal states is very convenient, therefore we always take advantage of the symmetry test and work with ε-tilted states. We reserve the capital letters, i.e., $|\Psi\rangle$ or simply Ψ,[10] to denote an ε-tilted state. The *size* of Ψ, denoted $|\Psi|$, is the size of Ψ viewed as a set of states.

The tilted states tensorize. In particular, for two sets of states $\Psi = \{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle\}$ and $\Phi = \{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_k\rangle\}$ of the same size, let $\Psi \otimes \Phi$ denote the set of states $\{|\psi_1, \phi_1\rangle, \ldots, |\psi_k, \phi_k\rangle\}$ (if there is not a default order, the order can be set arbitrarily).

[7] We stress that this is a simplistic view of the protocol. See Section 7 for the precise technical details.

[8] Assuming $|\psi_L\rangle$ and $|\psi_R\rangle$ are of the above form.

[9] As in the SSE and UG protocols, there is also distribution on pairs of operator (Γ_L, Γ_R) here.

[10] In this paper, we never use the density operator, so there should be no confusion.

Proposition 4.2 (Tensorization of tilted states). *If Ψ is an ε-tilted state and Φ is a γ-tilted state, and $|\Psi| = |\Phi| = k$. Then $\Psi \otimes \Phi$ is an $(\varepsilon+\gamma)$-tilted state.*

4.2 Symmetry Test

The symmetry test is described below.

Symmetry Test
Input: $\Psi = \{a_1, a_2, \ldots, a_k\} \subseteq \mathbb{S}$ for some even number k.
(i) Sample a random matching π within $1, 2, \ldots, k$.
(ii) SwapTest on the pairs based on the matching π.
Accept if all SwapTests accept.

Theorem 4.3 (Symmetry test). *Suppose Ψ is not an ε-tilted state. Then the symmetry test passes with probability at most $\exp(-\Theta(\varepsilon^2 k))$. On the contrary, for 0-tilted state Ψ, the symmetry test accepts with probability 1.*

4.3 Sparsity Test

Now we move on to the sparsity test, where the non-negative assumption is used crucially. In the sparsity test, aside from the state that we want to test whether it's close to some subset state, the prover will provide an auxiliary proof to assist the verifier.

In what follows, we provide two versions of the sparsity tests. In the first version, we want to know if a given state $|\psi\rangle$ is close to some subset state without prior knowledge of the sparsity γ. In the second version, there is a target sparsity γ, and we want to know if $|\psi\rangle$ is close to $\mathcal{S}_\gamma$. We describe the first version below.

Sparsity test I (with precision ε)
Input: $\Psi = \{u_1, \ldots, u_{2k}\} \subseteq \mathbb{S}^+, \Phi = \{v_1, \ldots, v_{2k}\} \subseteq \mathbb{S}^+$.
Partition Ψ into Ψ_0 and Ψ_1 of equal size, and partition Φ into Φ_0 and Φ_1 of equal size.
(i) SwapTest on $(\Psi_0, \mathbf{1}_{[n]}/\sqrt{n})$;
(ii) SwapTest on $(\Phi_0, \mathbf{1}_{[n]}/\sqrt{n})$;
(iii) SwapTest on (Ψ_1, Φ_1).
Accept if and only if $\alpha+\beta \in [3/2-\sqrt{\varepsilon}, 3/2+\sqrt{\varepsilon}]$ and $\lambda \le 1/2+\sqrt{\varepsilon}$, where α, β and λ are the fractions of accepted SwapTests in (i), (ii), and (iii), respectively.
Output: α, β, λ.

Theorem 4.4 (Sparsity test). *Given $\Psi = \{u_i \in \mathbb{S}^+_n\}_{i\in[2k]}, \Phi = \{v_i \in \mathbb{S}^+_n\}_{i\in[2k]}$ two ε-tilted states for $\varepsilon < 1/2$. Let α, β, and λ be the outputs.*

(Completeness) For any 0-tilted states Ψ and Φ, such that $\Psi \in \mathcal{S}_\delta$, $\Phi \in \mathcal{S}_{1-\delta}$, and $\Psi \perp \Phi$. Then with probability at least $1-\exp(-\Theta(\varepsilon k))$ the sparsity test accepts, furthermore,

$$|2\alpha - 1 - \delta| \le \sqrt{\varepsilon},$$
$$|2\beta - 1 - (1-\delta)| \le \sqrt{\varepsilon}.$$

(Soundness) The sparsity test accepts with probability at most $\exp(-\varepsilon k)$, if either of the following fails to hold:

(i) *There is $S \subseteq [n]$, such that for any $\gamma > 0$,*

$$|S| \le (2\alpha-1)n + 9\varepsilon^{1/4} n/\gamma,$$

and for any representative $u \in \Psi$,

$$\|u|_S\|^2 \ge 1 - \gamma - 2\sqrt{\varepsilon}.$$

(ii) *There is $S \subseteq [n]$, such that*

$$||S| - (2\alpha-1)n| \le O(\varepsilon^{1/12}(2\alpha-1)^{1/3})n,$$

and for any representative $u \in \Psi$,

$$\mathrm{D}\left(u, \mathbf{1}_S/\sqrt{|S|}\right) = O\left(\frac{\varepsilon^{1/24}}{(2\alpha-1)^{1/3}}\right).$$

Key to the analysis of the sparsity test is the following lemma.

Lemma 4.5. *Let $u, v \in \mathbb{S}^+_n$ for an arbitrary natural number n. Let $\delta \in (0,1)$ be some constant. If for some small constant $\varepsilon > 0$, the following items are true:*

(i) $\langle u, v\rangle^2 \le \varepsilon$,
(ii) $|\langle u, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 - \delta| \le \varepsilon$,
(iii) $|\langle v, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 - (1-\delta)| \le \varepsilon$.

Then, for any $0 < \gamma < 1/2$, and some $|S| \le (\delta + 2\sqrt{\varepsilon}/\gamma)n$,

$$\|u|_S\|^2 \ge 1 - \gamma. \tag{4.1}$$

Furthermore, for some $S \subseteq [n]$ with

$$(\delta - O(\varepsilon))n \le |S| \le (\delta + O(\varepsilon^{1/6}\delta^{1/3}))n$$

we have

$$\langle u, \mathbf{1}_S/\sqrt{|S|}\rangle \ge 1 - O\left(\frac{\varepsilon^{1/6}}{\delta^{2/3}}\right).$$

Suppose that we have a target sparsity γ, a constant number in $(0,1)$. We adapt the previous sparsity test slightly to test whether some given state is close to $\mathcal{S}_\gamma$.

Sparsity test II (with target sparsity γ and precision ε)
Input: $\Psi = \{u_1, \ldots, u_{2k}\}, \Phi = \{v_1, \ldots, v_{2k}\}$
(i) Sparsity test I on (Ψ, Φ) with precision ε.
Accept if the sparsity test I accepts and its output satisfies: $2\alpha - 1 \in [\gamma - \sqrt{\varepsilon}, \gamma + \sqrt{\varepsilon}]$.

Theorem 4.6 (Sparsity test with target sparsity γ). *Let $\varepsilon > 0$ be such that $\varepsilon < \gamma^{4/5}$. Suppose that Ψ and Φ are ε-tilted states. Then the sparsity test accepts with probability at most $\exp(-\varepsilon k)$ if the following fails to hold:*

$$\mathrm{D}(\Psi, \mathcal{S}_\gamma) \le O\left(\frac{\varepsilon^{1/24}}{\gamma^{1/3}}\right). \tag{4.2}$$

If Ψ is the 0-tilted states from $\mathcal{S}_\gamma$, then there is Φ such that the sparsity test accepts with probability $1 - \exp(-\Theta(\varepsilon k))$

4.4 Validity Test

Consider the variable set $X = \{1, 2, \ldots, n\}$, and domain $\Sigma = \{1, 2, \ldots, q\}$. Recall that the valid set is the following

$$\mathcal{V} = \left\{ \frac{1}{\sqrt{n}} \sum_{i\in[n]} |i\rangle|x_i\rangle : \ \forall i \in [n], x_i \in \Sigma \right\}.$$

The goal is to test whether a state is close to $\mathcal{V}$.

Validity test (with precision d)
Input: $\Psi = \{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle\} \subseteq \mathbb{S}^+$.
(i) Apply discrete Fourier transform to the second register of Ψ.
(ii) Measure the second register.
Accept if $\alpha \leq 1/q + d$, where α is the fraction of $|0\rangle$ observed after measuring.

Theorem 4.7 (Validity test). *Suppose that Ψ is an ε-tilted state for some small $\varepsilon > 0$. Further suppose that for any representative state $|\psi\rangle \in \Psi$, $\mathrm{D}(|\psi\rangle, \mathcal{S}_{1/q}) \leq d$ for $2\varepsilon \leq d < 1/q$. Then the probability that in the validity test the fraction of measured $|0\rangle$ is less than $(1+qd)/q$ is at most $\exp(-\Theta(qd^2k))$, if*

$$\mathrm{D}(|\psi\rangle, \mathcal{V}) \geq \sqrt{6qd} + d.$$

If Ψ is a 0-tilted state from $\mathcal{V}$, then the validity test accepts with probability at least $1 - \exp(-\Theta(qd^2k))$.

5 SSE $\in$ $\mathbf{QMA}^+_{\log}(2)$

Definition 5.1 ((η,δ)-SSE graph)**.** *Let $\eta, \delta \in (0,1)$. We say that G is a (η,δ) small set expander, or simply (η,δ)-SSE for short, if for every $\emptyset \neq S \subseteq V$ of size $|S| \leq \delta |V|$ we have $\Phi_G(S) \geq 1 - \eta$.*

Definition 5.2 ((η,δ)-SSE)**.** *Let $\eta, \delta \in (0,1)$. An instance of (η,δ) small set expansion (SSE) problem is a graph G on the vertex set V such that*

(Yes) *There exists $S \subseteq V$ with measure at most δ and $\Phi_G(S) \leq \eta$;*
(No) *Every set $S \subseteq V$ of measure at most δ has expansion $\Phi_G(S) \geq 1 - \eta$.*

We now show that SSE can be verified with constant copies of unentangled proofs of non-negative amplitudes and a logarithmic number of qubits with *constant* completeness-soundness gap. More precisely, we prove the following theorem.

Theorem 5.3. *The (η,δ)-SSE problem is in $\mathrm{QMA}^+_{O_\delta(\log(n))}(2, c, s)$ with completeness $c \geq 1 - \eta$ and soundness $s \leq 5/6 + O(\sqrt{\eta}\log(1/\eta))$.*

We will prove the theorem by showing that the $\mathrm{QMA}_{\log}(2)$ protocol described in Algorithm 5.4 is complete and sound for (η,δ)-SSE.

Algorithm 5.4: (η,δ)-SSE Protocol
Let $\varepsilon = \eta^8\delta^4/C$, and $k = C\log(1/\eta)/\varepsilon^2$ for some large enough constant C.
Let S be the vertex set such that $|S| \leq \delta n$ and $\Phi_G(S) \leq \eta$.
Provers: Send
(i) $2k$ copies of the superpositions of the non-expanding set S, i.e.,

$$|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_{2k}\rangle = \frac{1}{\sqrt{\delta n}} \sum_{i \in S} |i\rangle.$$

(ii) $2k$ copies of the superpositions of the complement of S, i.e.,

$$|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_{2k}\rangle = \frac{1}{\sqrt{(1-\delta)n}} \sum_{i \notin S} |i\rangle.$$

Verifier: Choose uniformly at random one of the following tests.
(i) Symmetry test on $\{|\psi_i\rangle\}$ and symmetry test on $\{|\phi_i\rangle\}$.
(ii) Sparsity test on $(\{|\psi_i\rangle\}, \{|\phi_i\rangle\})$ with precision ε. If the output α is such that $2\alpha - 1 > (1+\eta)\delta$, *reject.*
(iii) Expansion test on $|\psi_i\rangle$ and $|\psi_j\rangle$ for two distinct random $i, j \in \{1, 2, \ldots, 2k\}$.

Since G is a d regular graph, its adjacency matrix A can be written as a sum of d permutation matrices $P_1, \ldots, P_d$. This representation as a sum of unitary matrices will be important to view these matrices as valid quantum operations. To test the lack of expansion of the support of $|\psi_1\rangle$, we apply to this state a permutation P_i, chosen uniformly at random. Then, we test if the resulting state (mostly) overlaps with $|\psi_2\rangle$ (which is supposed to encode the same set in its support). This test is described in Algorithm 5.5.

Algorithm 5.5: Expansion Test
Input: $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{S}^+$
(i) Choose $r \in [d]$ uniformly at random;
(ii) Compute $P_r|\psi_1\rangle$;
(iii) SwapTest$(P_r|\psi_1\rangle, |\psi_2\rangle)$.
Accept if the swap test accepts.

6 GapUG $\in$ $\mathbf{QMA}^+_{\log}(2)$ AND NP $\subseteq$ $\mathbf{QMA}^+_{\log}(2)$

Definition 6.1 (Unique Games)**.** *A unique game instance $\mathfrak{I}$ consists of a d-regular graph $G = (V, E)$. Each edge $e = (a, b) \in E$ is associated with a bijective constraint $f_e : \Sigma \to \Sigma$, where $\Sigma = \{1, 2, \ldots, q\}$ for some constant q.*

For any labeling $\ell : [n] \to \Sigma$, the value associated with the labeling is the fraction of edge constraints satisfied by the labeling, i.e.,

$$\frac{1}{nd}|\{(a,b) \in E : f_{(a,b)}(\ell(a)) = \ell(b)\}|.$$[11]

The value of $\mathfrak{I}$, denoted val$(\mathfrak{I})$*, is the max value over all possible labelings.*

[11] Though we can think of the graph in the definition being undirected, when we describe an edge constraint for $e = (a, b)$ using a bijection, we need labels of one vertex as the domain and labels of the other as the range of f. So when we say f_e, we always have an implicit orientation of the edge. So the set here counts each edge twice, that is val can take the value up to 1.

Definition 6.2 ($(1-\delta,\eta)$-GapUG problem)**.** *Given any unique games instance $\mathfrak{I}$. Determine which of the following two cases is true:* ***(Yes)*** $\mathrm{val}(\mathfrak{I}) \geq 1-\delta$; ***(No)*** $\mathrm{val}(\mathfrak{I}) \leq \eta$.

The purpose of this section is to establish the following theorem.

Theorem 6.3. *For any $\delta, \eta \in (0,1)$ such that $(1-\delta)^2 > \eta$, then*

$$(1-\delta,\eta)\text{-GapUG} \in \mathrm{QMA}^{+}_{\log}(2).$$

It suffices to present a $\mathrm{QMA}^{+}_{\log}(k)$ protocol (see Algorithm 6.4) for some constant k for the $(1-\delta,\eta)$-GapUG problem. For the given graph $G=(V,E)$, say $V=\{1,2,\ldots,n\}$. Since G is a regular graph, we can partition E into d permutations $\pi_1,\pi_2,\ldots,\pi_d : \{n\} \to \{n\}$. The permutation can also be thought of as a perfect matching between two vertex sets L and R with $L=R=V$. We find the matching view more convenient, so we often call π a matching. For any labeling $\ell : [n] \to \Sigma$, we represent it by the following quantum state

$$|\psi\rangle = \frac{1}{\sqrt{n}} \sum_{i\in[n]} |i\rangle|\ell(i)\rangle.$$

Recall that $\mathcal{V} \subseteq \mathcal{S}_{1/q}$ denote the set of all valid labelings, i.e.,

$$\mathcal{V} := \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |i\rangle|v_i\rangle : v_i \in \Sigma \right\}.$$

Let Π_r be the unitary map associated with the matching π_r, such that for any $r \in [d]$, $i \in [n]$, and $v \in \Sigma$:

$$\Pi_r\, |i\rangle|v\rangle \mapsto |\pi_r(i)\rangle|f_{(i,\pi_r(i))}(v)\rangle.$$

In words, when we pick a matching π_r and a labeling $|\psi\rangle$ on L, then $\Pi_r|\psi\rangle$ represents the unique labeling on R that satisfies all the edge constraints for the edges in π_r. In reality, L and R are the same vertex set, they have the same labeling. Let

$$\theta = \frac{1}{2}\left(\frac{1+(1-\delta)^2}{2} + \frac{1+\eta}{2}\right), \quad \lambda = \frac{(1-\delta)^2}{2} - \frac{\eta}{2}.$$

Algorithm 6.4: $(1-\delta,\eta)$-GapUG Protocol

Let $\varepsilon = \lambda^{48}/(Cq^{32})$, and $k = C/\varepsilon^2$ for some large enough constant C.

Provers: send

(i) $2k$ copies of labelings realize $\mathrm{val}(\mathfrak{I})$, i.e.,

$$|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_{2k}\rangle = \frac{1}{\sqrt{n}} \sum_{i\in[n]} |i\rangle|\ell(i)\rangle.$$

(ii) $2k$ copies of the labelings but complemented, i.e.,

$$|\gamma_1\rangle, |\gamma_2\rangle, \ldots, |\gamma_{2k}\rangle = \frac{1}{\sqrt{n}} \sum_{i\in[n]} |i\rangle \frac{1}{\sqrt{q-1}} \sum_{v\neq\ell(i)} |v\rangle.$$

Verifier: Let $\Psi = \{|\psi_1\rangle, \ldots, |\psi_{2k}\rangle\}$, and similarly for Γ. Run a uniformly random test of the following

(i) Two symmetry tests on Ψ and Γ.

(ii) Sparsity test on (Ψ,Γ) with target sparsity $1/q$ and precision ε.

(iii) Validity test on Ψ with precision $\nu = \varepsilon^{1/24} q^{1/3}$.

(iv) Labeling test on Ψ_0, Ψ_1, where Ψ_0 and Ψ_1 are partition of Ψ into two subsets with equal size.

The labeling test is described below.

Labeling Test

Input: $\Psi = \{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle\}, \Phi = \{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_k\rangle\}$.

(i) For i from 1 to k, SwapTest on $(\Pi_r|\psi_i\rangle, |\phi_i\rangle)$ for uniformly random $r \in [d]$ (each iteration with a fresh random choice).

Accept if more than a θ fraction the SwapTests accept.

We record the following lemma about the labeling test.

Lemma 6.5 (Labeling test)**.** *Suppose $\mathrm{val}(\mathfrak{I}) \leq \eta$. Given ε-tilted states Ψ such that any representative state $|\psi\rangle$ satisfies $\mathrm{D}(|\psi\rangle,\mathcal{V})$ and ε sufficiently small (for example, $\mathrm{D}(|\psi\rangle,\mathcal{V}) \leq \lambda/8$ and $\varepsilon \leq \lambda^2/256$). Then the labeling test accepts Ψ with probability at most $\exp(-\Theta(\lambda^2 k))$.*

Regularization—NP $\subseteq$ QMA$^{+}_{\log}(2)$. Due to the works [20, 21, 34, 35], it is known that the $(1/2,\eta)$-GapUG problem is NP-hard. An optimistic reader would happily conclude that NP $\subseteq \mathrm{QMA}^{+}_{\log}(2)$. This is indeed the case, with a small caveat though: In our previous discussion, we assumed the graph instance to be regular. However, when we convert a general graph into a regular one, the value of the game will change. We address this issue here.

Theorem 6.6 (Regularization [19]). *For any general unique games instance $\mathfrak{I}$, there is a new unique games instance $\mathfrak{I}'$ that is polynomial time constructible such that*

$$\mathrm{val}(\mathfrak{I}) \geq \frac{1}{2} \Longrightarrow \mathrm{val}(\mathfrak{I}') \geq 1 - \frac{1}{2(d+1)}, \tag{6.1}$$

$$\mathrm{val}(\mathfrak{I}) \leq \eta \Longrightarrow \mathrm{val}(\mathfrak{I}') \leq 1 - \frac{1-\eta}{d+1}. \tag{6.2}$$

The regularization process follows closely that of Dinur's treatment [19]. Define a new graph $G'=(V',E')$, such that

$$V' = \{(v,e) \in V \times E : v \text{ is incident to } e\}$$
$$E' = E'' \cup \bigcup_{v\in V} E_v,$$

where $E'' = \{((v,e),(u,e)) : (v,u) = e \in E\}$ and E_v is the set of edges in the d-regular expander graph $G_v = (V_v = \{(v,e) \in V'\}, E_v)$, for some constant d, whose Cheeger constant is at least 2.[12] In words, we replace every vertex v with a cluster of vertices of size equal to the number of edges that v is incident to in G. Within each cluster, the vertices are connected based on expander graphs. For every edge, $e=(u,v)$ in the original graph, connect the vertex (u,e) with vertex (v,e) in the new graph. The constraints f' on E'' will be like that of f_e on E. In particular, $f'_{((u,e),(v,e))} = f_{(u,v)}$. Further, the constraints on edges E_v will be the equality constraints, which can be represented as a bijective map. This new UG instance $\mathfrak{I}'$ satisfies that described in Theorem 6.6. Therefore, for the regular graph, $(1-\frac{1}{2(d+1)}, 1-\frac{1-\eta}{d+1})$-GapUG problem is NP-hard.

We verify that for any $\eta < 1/4(d+1)$,

$$\left(1 - \frac{1}{2(d+1)}\right)^2 > 1 - \frac{1-\eta}{d+1}.$$

Therefore, by Theorem 6.3, we have

[12] A random graph G_v would be good, and various explicit constructions are known. We refer interested readers to the wonderful survey on this topic [27].

Theorem 6.7. *With constant completeness and soundness gap,* $\mathrm{NP} \subseteq \mathrm{QMA}^{+}_{\log}(2)$.

7 NEXP = QMA$^{+}(2)$

In this section, we scale up our previous result to NEXP = $\mathrm{QMA}^{+}(2)$. The direction that $\mathrm{QMA}^{+}(2) \subseteq \mathrm{NEXP}$ follows the same trivial argument that $\mathrm{QMA}(2) \subseteq \mathrm{NEXP}$—guess the quantum proofs. Our focus will be on the other direction. The starting point would be a PCP for NEXP. For the moment, we abstract things out and focus on the constraints satisfaction problem (CSP) with the understanding that the CSP system will come from the corresponding PCP.

Definition 7.1. *An (N, R, q, Σ)-CSP system $\mathfrak{C}$ on N variables with values in Σ consists of a set (possibly a multi-set) of R constraints $\{C_1, C_2, \ldots, C_R\}$, and the arity of each constraint is exactly q. The value of $\mathfrak{C}$, denoted $\mathrm{val}(\mathfrak{C})$, is the maximum fraction of the satisfiable constraints over all possible assignment $\sigma : [N] \to \Sigma$. The $(1, \delta)$-GapCSP problem is to distinguish whether a given system $\mathfrak{C}$ is such that (**Yes**) $\mathrm{val}(\mathfrak{C}) = 1$ or (**No**) $\mathrm{val}(\mathfrak{C}) \leq \delta$.*

For any CSP system $\mathfrak{C}$, we think of a bipartite graph $G_{\mathfrak{C}}$ where the left vertices are the constraints and the right vertices are the variables. Whenever a constraint queries a variable there is an edge in the graph between the corresponding vertices. For any $j \in [R]$, let $\mathrm{Adj}_C(j)$ denote the list of variables that C_j queries; and for any $i \in [N]$, let $\mathrm{Adj}_V(i)$ denote the list of constraints that query variable i. An efficient CSP system $\mathfrak{C}$ should satisfy that for any $j \in [R]$, there is an algorithm that compute C_j in time $\mathrm{poly}\log(NR)$. That includes deciding which variables are queried by C_j, and based on the values of the relevant variables compute C_j. For our purpose, we require stronger properties, which we refer to as *double explicitness*. Informally, we require that given any variable i, we can also "list" the constraints that query i efficiently.

Definition 7.2 (Doubly explicit CSP)**.** *For any (family of) (N, R, q, Σ)-CSP system $\mathfrak{C}$, we say that $\mathfrak{C}$ is doubly explicit if the following are computable in time $\mathrm{poly}\log(NR)$:*

(i) *The cardinality of $\mathrm{Adj}_C(j)$ for any $j \in [R]$ and the cardinality of $\mathrm{Adj}_V(i)$ for any $i \in [N]$.*

(ii) $\mathrm{Adj}_C^{\mathrm{global}\to\mathrm{local}} : [R] \times [N] \to [q]$, *such that* $\mathrm{Adj}_C^{\mathrm{global}\to\mathrm{local}}(j, i) = \iota$ *if i is ιth variable that C_j queries.*[13]

(iii) $\mathrm{Adj}_C^{\mathrm{local}\to\mathrm{global}} : [R] \times [q] \to [N]$, *such that* $\mathrm{Adj}_C^{\mathrm{local}\to\mathrm{global}}(j, \iota)$ *is the ιth variable that C_j queries.*

(iv) $\mathrm{Adj}_V^{\mathrm{global}\to\mathrm{local}} : [N] \times [R] \to [R]$, *such that* $\mathrm{Adj}_V^{\mathrm{global}\to\mathrm{local}}(i, j) = \iota$ *if ι is the index of constraint j in $\mathrm{Adj}_V(i)$.*

(v) $\mathrm{Adj}_V^{\mathrm{local}\to\mathrm{global}} : [N] \times [R] \to [R]$ *such that for any $i \in [N]$ and $\iota \in [|\mathrm{Adj}_V(i)|]$, let $j = \mathrm{Adj}_V^{\mathrm{local}\to\mathrm{global}}(i, \iota)$, then ιth constraints in $\mathrm{Adj}_V(i)$ is C_j.*

[13] If C_j does not query i, we don't care about the value of $\mathrm{Adj}_C^{\mathrm{global}\to\mathrm{local}}$. Similarly for $\mathrm{Adj}_V^{\mathrm{global}\to\mathrm{local}}$.

In words, in the bipartite graph $G_{\mathfrak{C}}$. For each vertex, say $i \in [N]$, there is a local view of its neighborhood $\mathrm{Adj}_V(i)$. We should be able to efficiently switch from this local representation to a global representation, by $\mathrm{Adj}_V^{\mathrm{local}\to\mathrm{global}}(i, \cdot)$, and vice versa.

Another property we require is the *uniformity*, defined below.

Definition 7.3 (T-Strongly uniform CSP)**.** *For any (N, R, q, Σ)-CSP system $\mathfrak{C}$ and $T \in \mathbb{Z}$, we say that $\mathfrak{C}$ is T-strongly uniform if the variable set $[N]$ can be partitioned into at most T subsets $V_1 \cup V_2 \cup \cdots \cup V_T$, such that the cardinality of $\mathrm{Adj}_V(i)$ for any variable i only depends on which subset it belongs to. Furthermore, let $\tau : [N] \to [T]$, such that $\tau(i) = j$ if $i \in V_j$. Then $\tau(i)$ can be computed in time $\mathrm{poly}\log(NR)$.*

Given some $(N, R, q, \{0,1\})$-CSP system $\mathfrak{C}$ that is T-strongly uniform for some constant T and is strongly explicit. Then it is NEXP-hard to decide whether $\mathrm{val}(\mathfrak{C}) = 1$ or $\mathrm{val}(\mathfrak{C}) < \delta$ for some absolute constant δ. This CSP $\mathfrak{C}$ comes from the efficient PCP for NEXP. Although not all PCP satisfies doubly explicitness or uniformity, there is some PCP construction that enjoys these properties. We discuss such PCP in more detail and prove the related properties in the full paper.

Theorem 7.4 (PCP for NEXP). *There is a* PCP *system for a* NEXP*-complete problem, in which the verifier tosses* $\mathrm{poly}(n)$ *random bits and makes a constant number of queries to the proof* Π *such that if the input is a "Yes" instance, then the verifier always accept; if the input is a "no" instance, then the verifier accepts with probability at most δ for some constant δ. Furthermore, this* PCP *is doubly explicit and T-strongly uniform for some constant T.*

This PCP gives rise to a $(1, \delta)$-GapCSP instances for some $(N = 2^{\mathrm{poly}(n)}, R = 2^{\mathrm{poly}(n)}, q = O(1), \{0,1\})$-CSP system that are T-strongly uniform for some constant T and doubly explicit. In the remainder of the section, our goal is to prove the following theorem:

Theorem 7.5. *For any constant strongly uniform and doubly explicit (N, R, q, Σ)-CSP system $\mathfrak{C}$, there is a* $\mathrm{QMA}^{+}(2)$ *protocol that solves the $(1, \delta)$-GapCSP problem for $\mathfrak{C}$ with constant completeness and soundness gap.*

Theorem 7.4 together with Theorem 7.5 imply that

Theorem 7.6. $\mathrm{NEXP} \subseteq \mathrm{QMA}^{+}(2)$ *with constant completeness and soundness gap.*

7.1 Explicit Regularization

The first step towards proving Theorem 7.5 is regularization for the CSP $\mathfrak{C}$, very much like that in Theorem 6.6. The main technical issue is that everything happening in the previous case needs to be efficient for the exponentially large expander graphs. Fortunately, explicit constructions of expander graphs are very well-studied.

Theorem 7.7 (Explicit regular expander graphs [2, 40]). *There is some constant d, for which we have the following explicit constructions on expander graphs with Cheeger constant at least 2:*

(i) *For any n, there is a d-regular expander graph on n vertices.*

(ii) *For any prime $p > 17$, there exists a d-regular expander graph on $n = p(p^2-1)$ vertices. Furthermore, the graph G can be decomposed into d matchings $\pi_1, \pi_2, \ldots, \pi_d$, such that given $i \in [n]$ and $j \in [d]$, there is a* $\mathrm{poly}\log(n)$*-time algorithm $\Pi_G : [n] \times [d] \to [n]$, such that*

$$\Pi_G(i,j) = \pi_j(i).$$

For both constructions, given $i \in [n]$, the neighbors of i can be listed in time $\mathrm{poly}\log(n)$.

Since the second construction of expander graphs from the above theorem does not work for any number of vertices, we also need the following theorem about primes in short intervals.

Theorem 7.8 (Primes in short intervals [16]). *There is some absolute constant n_0, such that for any integer $n > n_0$, there is a prime between the interval $[n - 4n^{2/3}, n]$.*

With the above tools at our disposal, we discuss the explicit regularization for this exponentially large CSP $\mathfrak{C}$. Replace the variable i with a cluster of variables labeled (i, ι) for $\iota \in [n_i]$, where $n_i = |\mathrm{Adj}_V(i)|$. If $n_i < n_0$ for some absolute constant n_0 (this can be a larger constant than that in Theorem 7.8), then we can simply use the expander graph provided by Theorem 7.7 (i). For $n_i \geq n_0$, we use the expander graph provided by Theorem 7.7 (ii). In particular, let p_i be some prime such that

$$p_i \in [\lfloor n_i^{1/3} \rfloor - 4\lfloor n_i^{1/3} \rfloor^{2/3}, \lfloor n_i^{1/3} \rfloor].$$

The existence of p_i is guaranteed by Theorem 7.8. Let $n_i' := p_i(p_i^2 - 1) \in [n - O(n^{8/9}), n]$, and let

$$V_i' = \{(i,j) : j \leq n_i'\},$$
$$V_i'' = \{(i,j) : n_i' < j \leq n_i\}.$$

Depending on n_0, $|V_i''| \leq \eta n_i$ for $\eta = \eta(n_0)$. As we set n_0 to be a large enough constant, η is arbitrarily small. Connect the vertices in V_i' by a d-regular expander graph G_i, whose existence is guaranteed by Theorem 7.7 (ii). For all vertices in V_i'', add d self-loops. Associate these edges with equality constraints. Let $\mathfrak{C}'$ denote the new CSP instance. Recall that q is the number of variables queried by each constraint in $\mathfrak{C}$

Claim 7.9. *If* $\mathrm{val}(\mathfrak{C}) = 1$, *then* $\mathrm{val}(\mathfrak{C}') = 1$. *If* $\mathrm{val}(\mathfrak{C}) = \delta < 1$, *then the total number of unsatisfied constraints in $\mathfrak{C}'$ is at least* $(1 - \delta - q\eta)R$.

7.2 The Protocol

In the protocol, the provers are supposed to provide the following state:

$$|\psi\rangle = \sum_{j \in [R]} |j\rangle|v_j\rangle, \tag{7.1}$$

where $v_j \in \mathbb{C}^{|\Sigma^q|}$, which should indicate that the q variables with order listed in $\mathrm{Adj}_C(j)$ queried by C_j have value $v_{j,1}, v_{j,2}, \ldots, v_{j,q}$, respectively. This way of encoding is very convenient for evaluating whether each constraint is satisfied or not. But requires some work to make sure that the values v_j are consistent: Different constraints will share variables and the value of any variable across different constraints should be the same. Recall that, in the previous section when we discuss the regularization step for our CSP $\mathfrak{C}$ with variable set $V = [N]$ and constraints $C_1, \ldots, C_R$, from which we obtain a new CSP $\mathfrak{C}'$ such that each variable appears in exactly d number of the new constraints. Furthermore, a new variable in $\mathfrak{C}'$ will be a tuple composed of a variable $i \in V$ and a constraint C_j that queries i. Therefore, our way of encoding in (7.1), in a sense, is to write the superpositions of the new variables along with their values in the regularized CSP.

Let $n_1, n_2, \ldots, n_T$ be the cardinalities of $\mathrm{Adj}_V(i_1), \mathrm{Adj}_V(i_2), \ldots \mathrm{Adj}_V(i_T)$ where $i_1, i_2, \ldots, i_T$ are arbitrary variables from $V_1, V_2, \ldots, V_T$, respectively. Next, we describe our protocol for the CSP instance that we have.

Algorithm 7.10: Protocol for doubly explicit CSP

Let ε be some small enough constant, and k some large enough constant.

Prover provides:

(i) T primes $p_1, p_2, \ldots, p_T$, such that $p_i \in [\lfloor n_i^{1/3} \rfloor - 4\lfloor n_i^{1/3} \rfloor^{2/3}, \lfloor n_i^{1/3} \rfloor]$.

(ii) $\Psi := 2k$ copies of the state

$$\sum_{j \in [R]} |j\rangle|v_j\rangle, \qquad \forall j \in [R],\ v_j \in \Sigma^q.$$

(iii) $\Phi := 2k$ copies of the state

$$\sum_{j \in [R]} |j\rangle \sum_{v \in \Sigma^q : v \neq v_j} \frac{|v\rangle}{\sqrt{|\Sigma|^q - 1}}.$$

Verifier:

(i) Test if $p_1, p_2, \ldots, p_T$ are primes satisfying the size constraints, *reject* if not.

(ii) Symmetry test on Ψ and Φ.

(iii) Sparsity test Π on (Ψ, Φ) with target sparsity $|\Sigma|^{-q}$ and precision ε

(iv) Validity test on Ψ.

(v) Constraints test Ψ.

The constraints test will be used to check the new constraints of our CSP after the regularization. But before we formally describe the constraints test, we make some preparations. Let $H = \mathbb{C}^R \otimes \mathbb{C}^{|\Sigma|^q} \otimes \mathbb{C}^N \otimes \mathbb{C}^{|\Sigma|}$. The first register is the constraint register. The second register is used to encode the values of the q variables queried by the constraint stored in the first register. The third register is the variable register to store the variable name. The last register is used to store the value of the variable in the third register. Now we define three quantum channels that will be used to manipulate our state in the constraints test.

- $\mathcal{A}$, the operator that converts a given state from (7.1) to an actual superposition of the new variables from $\mathfrak{C}'$ together with their values.
- $\mathcal{M}_k$ for $k \in [d]$, the operator that "implements" the kth one after we decompose the d-regular expander graphs into matchings.
- $\mathcal{B}$, the operator that given $|j\rangle|v_j\rangle$, evaluates if C_j outputs 1 if the values of the variables it queries are given by the string v_j.

Precisely, let $\mathcal{B}$ acting on $\mathbb{C}^R \otimes \mathbb{C}^{q|\Sigma|} \otimes \mathbb{C}^2$ be such that

$$\mathcal{B} : |j\rangle|v\rangle|0\rangle \mapsto |j\rangle|v\rangle|C_j(v)\rangle.$$

Recall that the constraints of $\mathfrak{C}'$ consist of that from $\mathfrak{C}$ and the consistency constraints induced by the expander graphs and self-loops we add. As $\mathcal{B}$ checks if the value v satisfies the constraints C_j, it takes care of the first kind of constraints of $\mathfrak{C}'$.

Define the operator $\mathcal{A}$ acting on $H = \mathbb{C}^R \otimes \mathbb{C}^{|\Sigma|^q} \otimes \mathbb{C}^N \otimes \mathbb{C}^{|\Sigma|}$ such that

$$\mathcal{A} : |j\rangle|v\rangle|0\rangle|0\rangle \mapsto \frac{1}{\sqrt{q}} \sum_{\iota=1}^{q} |j\rangle|v\rangle|i_\iota\rangle|v_\iota\rangle,$$

where $x_{i_1}, x_{i_2}, \ldots, x_{i_q}$ are the variables listed in $\mathrm{Adj}_C(j)$. In words, given the constraints j, and the values v to the variables that j queries, we put the third and fourth register (the variable register) into the superposition of the variables in $\mathrm{Adj}_C(j)$ together with their value based on v.

Next, we define $\mathcal{M}$ formally. Recall that for any variable $i \in [N]$, after regularization, the set of variables constructed from i includes

$$V_i' = \{(i, j) : j \le n_i'\},$$
$$V_i'' = \{(i, j) : n_i' < j \le n_i\}.$$

The new constraints include an expander G_i on V' and self-loops on V_i''. We can decompose G_i into d matchings, and for variables in V_i'', they are matched with themselves. For any $k \in [d]$, let $\mathcal{M}_k$ be the operator such that:

$$\mathcal{M}_k : |j\rangle|v\rangle|i\rangle|v'\rangle \mapsto |j'\rangle|v\rangle|i\rangle|v'\rangle,$$

where

$$j' = \begin{cases} \mathrm{Adj}_V^{\text{local}\to\text{global}}(i, \Pi_{G_i}(\iota, k)), & \iota \le n_i', \\ j, & \text{otherwise}, \end{cases} \tag{7.2}$$

$$\iota = \mathrm{Adj}_V^{\text{global}\to\text{local}}(i, j).$$

That is, suppose we take the kth matching to permute the variables in $\mathfrak{C}'$, then j' in (7.2) determines that $(i, j) \in \mathfrak{C}'$ should be switched to $(i, j') \in \mathfrak{C}'$. But the expander graphs are labeled by $\{1, 2, \ldots, n_i'\}$, corresponding to indices of $\mathrm{Adj}_V(i)$, to obtain the actual constraint $C_{j'}$, we need to convert from local index to global index, and later convert it back.

$\mathcal{A}$ together with $\mathcal{M}_k$ takes care of the consistency constraints just like how we do it for UG games. Take a pair of equal states $|\psi\rangle$ and $|\phi\rangle$ supposed to be valid. Apply $\mathcal{A}$ to both states. But apply $\mathcal{M}_k$ only to $|\phi\rangle$. Now the two states are equal if the original states encode a consistent value for all constraints, except we should ignore the second register. To get rid of the second register, we make a measurement. In particular, let

$$|\mu\rangle = \frac{1}{|\Sigma|^{-q}} \sum_{v \in \Sigma^q} |v\rangle.$$

Consider the measurement $M = \{\Pi_{|\mu\rangle\langle\mu|}, 1 - \Pi_{|\mu\rangle\langle\mu|}\}$. It's easy to see that after the measurement, with probability $p = |\Sigma|^{-q}$, the second register is set to $|\mu\rangle$ and thus disentangled from the other registers. Since we have a larger number of provers, with p fraction of proofs left is enough.

Note that $\mathcal{A}, \mathcal{M}, \mathcal{B}$ are all valid quantum operations.

Claim 7.11. *$\mathcal{A}, \mathcal{B}, \mathcal{M}_k$ can be implemented by* BQP *circuits.*

With the above preparation, we now describe the constraints test.

Constraints test

Input: Ψ_0, Ψ_1, each is a set of k states for some large constant k. Pair the states in Ψ and Φ.

For each pair $|\psi\rangle$ and $|\phi\rangle$, with probability $2d/(2d+1)$ take the consistency check, with the remaining probability take the inner constraints test

(i) Consistency check
- Apply $\mathcal{A}$ to $|\phi\rangle$ and $|\psi\rangle$.
- Apply $\mathcal{M}_k$ to $|\phi\rangle$ for a uniformly random $k \in [d]$.
- Measure the second register of $|\psi\rangle, |\phi\rangle$ based on M, if either measurement does not output $|\mu\rangle$, ignore this pair.
- SwapTest on $|\psi\rangle$ and $|\phi\rangle$.

(ii) Inner constraints test
- With probability $1 - |\Sigma|^{-2q}$, ignore this pair.
- Apply $\mathcal{B}$ to $|\psi\rangle$
- Measure the third register, *Accept* if 1 is observed.

Accepts if more than θ fraction of the pairs (that are not ignored) get accepted, where

$$\theta = 1 - \frac{1 - \delta}{4(2d+1)}.$$

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Grant No. CCF-1900460. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

We thank Vijay Bhattiprolu for the discussions during the initial stages of this project. We thank STOC reviewers and Avi Wigderson for their valuable feedback on our earlier draft.

REFERENCES

[1] Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. 2008. The Power of Unentanglement. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*. 223–236. https://doi.org/10.1109/CCC.2008.5

[2] Noga Alon. 2021. Explicit Expanders of Every Degree and Size. *Combinatorica* (Feb. 2021). https://doi.org/10.1007/s00493-020-4429-x

[3] Sanjeev Arora and Boaz Barak. 2009. *Computational Complexity: A Modern Approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804090

[4] Boaz Barak, Fernando G.S.L. Brandão, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. 2012. Hypercontractivity, Sum-of-Squares Proofs, and Their Applications. In *Proceedings of the 44th ACM Symposium on Theory of Computing*. https://doi.org/10.1145/2213977.2214006

[5] Boaz Barak, Pravesh Kothari, and David Steurer. 2019. Small-Set Expansion in Shortcode Graph and the 2-to-2 Conjecture. In *ITCS 2019*. https://doi.org/10.4230/LIPIcs.ITCS.2019.9

[6] Boaz Barak, Pravesh K. Kothari, and David Steurer. 2017. Quantum Entanglement, Sum of Squares, and the Log Rank Conjecture. In *Proceedings of the 49th ACM Symposium on Theory of Computing*. ACM, 975–988. https://doi.org/10.1145/3055399.3055488

[7] Salman Beigi. 2010. NP VS QMAlog(2). *Quantum Info. Comput.* (2010). https://doi.org/10.5555/2011438.2011448

[8] J. S. Bell. 1964. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika* 1 (Nov 1964), 6 pages. Issue 3. https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195

[9] Charles H. Bennett and Gilles Brassard. 2014. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science* (2014). https://doi.org/10.1016/j.tcs.2014.05.025

[10] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. 1993. Teleporting an unknown quantum state via dual

classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* 70 (Mar 1993). Issue 13. https://doi.org/10.1103/physrevlett.70.1895

[11] Yonatan Bilu and Nathan Linial. 2006. Lifts, Discrepancy and Nearly Optimal Spectral Gap. *Combinatorica* 26, 5 (Oct. 2006), 495–519. https://doi.org/10.1007/s00493-006-0029-7

[12] Hugue Blier and Alain Tapp. 2009. All Languages in NP Have Very Short Quantum Proofs. In *2009 Third International Conference on Quantum, Nano and Micro Technologies*. 34–37. https://doi.org/10.1109/icqnm.2009.21

[13] Fernando G.S.L. Brandão and Aram W. Harrow. 2013. Quantum de Finetti Theorems under Local Measurements with Applications. In *Proceedings of the 45th ACM Symposium on Theory of Computing*. https://doi.org/10.1145/2488608.2488718

[14] Fernando G. S. L. Brandão, Matthias Christandl, and Jon Yard. 2011. Faithful Squashed Entanglement. *Communications in Mathematical Physics* (2011). https://doi.org/10.1007/s00220-011-1302-1

[15] Fernando G. S. L. Brandao and Aram W. Harrow. 2015. Estimating operator norms using covering nets. arXiv:1509.05065

[16] Yuan-You Fu-Rui Cheng. 2010. Explicit Estimate on Primes Between Consecutive Cubes. *Rocky Mountain Journal of Mathematics* 40, 1 (2010), 117 – 153. https://doi.org/10.1216/RMJ-2010-40-1-117

[17] Alessandro Chiesa and Michael A. Forbes. 2013. Improved Soundness for QMA with Multiple Provers. *Chic. J. Theor. Comput. Sci.* (2013). https://doi.org/10.4086/cjtcs.2013.001

[18] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. 1969. Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.* 23 (Oct 1969). Issue 15. https://doi.org/10.1103/physrevlett.24.549

[19] Irit Dinur. 2007. The PCP Theorem by Gap Amplification. *J. ACM* 54, 3 (jun 2007), 12–es. https://doi.org/10.1145/1236457.1236459

[20] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. 2018. On Non-Optimally Expanding Sets in Grassmann Graphs. In *Proceedings of the 50th ACM Symposium on Theory of Computing*. https://doi.org/10.1145/3188745.3188806

[21] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. 2018. Towards a Proof of the 2-to-1 Games Conjecture?. In *Proceedings of the 50th ACM Symposium on Theory of Computing* (Los Angeles, CA, USA) *(STOC 2018)*. Association for Computing Machinery, New York, NY, USA, 376–389. https://doi.org/10.1145/3188745.3188804

[22] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. 2004. Complete family of separability criteria. *Physical Review A* 69 (2004). https://doi.org/10.1103/physreva.69.022308

[23] A. Einstein, B. Podolsky, and N. Rosen. 1935. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* 47 (May 1935). Issue 10. https://doi.org/10.1007/978-3-322-91080-6_6

[24] François Le Gall, Shota Nakagawa, and Harumichi Nishimura. 2012. On QMA Protocols with Two Short Quantum Proofs. *Quantum Info. Comput.* (2012). https://doi.org/10.26421/qic12.7-8-4

[25] Aram W. Harrow and Ashley Montanaro. 2013. Testing Product States, Quantum Merlin-Arthur Games and Tensor Optimization. *J. ACM* 60, 1, Article 3 (feb 2013), 43 pages. https://doi.org/10.1145/2432622.2432625

[26] Aram W. Harrow, Anand Natarajan, and Xiaodi Wu. 2017. An Improved Semidefinite Programming Hierarchy for Testing Entanglement. *Communications in Mathematical Physics* (2017). https://doi.org/10.1007/s00220-017-2859-0

[27] Shlomo Hoory, Nathan Linial, and Avi Wigderson. 2006. Expander Graphs and Their Applications. *Bull. Amer. Math. Soc.* 43, 04 (Aug. 2006), 439–562. https://doi.org/10.1090/S0273-0979-06-01126-8

[28] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. 2009. Quantum entanglement. *Rev. Mod. Phys.* 81 (Jun 2009), 865–942. Issue 2. https://doi.org/10.1103/RevModPhys.81.865

[29] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. 2011. QIP = PSPACE. *J. ACM* (dec 2011). https://doi.org/10.1145/1806689.1806768

[30] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. 2020. MIP*=RE. https://doi.org/10.1145/3485628

[31] Subhash Khot. 2002. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing*. 767–775. https://doi.org/10.1145/509907.510017

[32] Subhash Khot. 2010. Inapproximability of NP-complete Problems, Discrete Fourier Analysis, and Geometry. In *Proceedings of the International Congress of Mathematicians*. https://doi.org/10.1142/9789814324359_0163

[33] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. 2004. Optimal inapproximability results for MAX-CUT and other two-variable CSPs?. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*. 146–154. https://doi.org/10.1109/focs.2004.49

[34] Subhash Khot, Dor Minzer, and Muli Safra. 2017. On Independent Sets, 2-to-2 Games, and Grassmann Graphs. In *Proceedings of the 49th ACM Symposium on Theory of Computing* (Montreal, Canada). Association for Computing Machinery, New York, NY, USA, 576–589. https://doi.org/10.1145/3055399.3055432

[35] Subhash Khot, Dor Minzer, and Muli Safra. 2018. Pseudorandom Sets in Grassmann Graph Have Near-Perfect Expansion. In *Proceedings of the 59th IEEE Symposium on Foundations of Computer Science*. https://doi.org/10.1109/focs.2018.00062

[36] Subhash Khot and Ryan O'Donnell. 2009. SDP Gaps and UGC-hardness for Max-Cut-Gain. *Theory of Computing* 5, 4 (2009), 83–117. https://doi.org/10.1109/focs.2006.67

[37] Subhash Khot and Oded Regev. 2003. Vertex Cover Might be Hard to Approximate to within $2-\varepsilon$. In *Proceedings of the 18th IEEE Conference on Computational Complexity*. https://doi.org/10.1109/ccc.2003.1214437

[38] Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. 2003. Quantum Merlin-Arthur Proof Systems: Are Multiple Merlins More Helpful to Arthur?. In *Algorithms and Computation*. https://doi.org/10.1007/978-3-540-24587-2_21

[39] Robert König and Graeme Mitchison. 2009. A most compendious and facile quantum de Finetti theorem. *J. Math. Phys.* 50, 1 (2009). https://doi.org/10.1063/1.3049751

[40] Alexander Lubotzky. 2011. Finite simple groups of Lie type as expanders. *Journal of the European Mathematical Society* 013, 5 (2011), 1331–1341. http://eudml.org/doc/277517

[41] Chris Marriott and John Watrous. 2005. Quantum Arthur–Merlin games. *Computational Complexity* (2005). https://doi.org/10.1007/s00037-005-0194-x

[42] Michael A. Nielsen and Isaac L. Chuang. 2010. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press. https://doi.org/10.5555/1972505

[43] Roman Orus. 2019. Tensor networks for complex quantum systems. *Nature Reviews Physics* (2019). https://doi.org/10.1038/s42254-019-0086-7

[44] Prasad Raghavendra. 2008. Optimal algorithms and inapproximability results for every CSP?. In *Proceedings of the 40th ACM Symposium on Theory of Computing*. 245–254. https://doi.org/10.1145/1374376.1374414

[45] Prasad Raghavendra and David Steurer. 2010. Graph Expansion and the Unique Games Conjecture. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*. https://doi.org/10.1145/1806689.1806792

[46] Prasad Raghavendra, David Steurer, and Prasad Tetali. 2010. Approximations for the Isoperimetric and Spectral Profile of Graphs and Related Parameters. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*. https://doi.org/10.1145/1806689.1806776

[47] Yaoyun Shi and Xiaodi Wu. 2012. Epsilon-Net Method for Optimizations over Separable States. In *Proceedings of the 39th International Colloquium on Automata, Languages and Programming*. https://doi.org/10.1016/j.tcs.2015.03.031

[48] Guifré Vidal. 2003. Efficient Classical Simulation of Slightly Entangled Quantum Computations. *Phys. Rev. Lett.* 91 (Oct 2003). https://doi.org/10.1103/physrevlett.91.147902

[49] Thomas Vidick and John Watrous. 2016. Quantum Proofs. *Foundations and Trends® in Theoretical Computer Science* (2016). https://doi.org/10.1561/9781680831276

[50] John Watrous. 2000. Succinct quantum proofs for properties of finite groups. In *FOCS*. IEEE Computer Society, 537–546. https://doi.org/10.1109/sfcs.2000.892141

[51] John Watrous. 2018. *The Theory of Quantum Information*. Cambridge University Press. https://doi.org/10.1017/9781316848142

Received 2022-11-07; accepted 2023-02-06