The Power of Unentangled Quantum Proofs with Non-negative Amplitudes

Fernando Granha Jeronimo* Pei Wu†

Abstract

Quantum entanglement is a fundamental property of quantum mechanics and plays a crucial role 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 $\text{QMA}^+(2)$. 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 $\subseteq \text{QMA}^+_{\log}(2)$ with a constant gap. By virtue of the new *constant* gap, we are able to "scale up" this result to $\text{QMA}^+(2)$, obtaining the full characterization $\text{QMA}^+(2) = \text{NEXP}$ 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).

Finally, as a result of the above characterization of NEXP, we further show that if $\text{QMA}^+(2)$ admits strong gap amplification for the completeness and soundness gap, then QMA(2) = NEXP, an intriguing possibility.

*`IAS & Simons Institute`. `granha@ias.edu`. This material is based upon work supported by the National Science Foundation under Grant No. CCF-1900460. The work is done while the author was at IAS.

†`Weizmann Institute of Science`. `pwu@ias.edu`. This material is based upon work supported by the National Science Foundation under Grant No. CCF-1900460. The work is done while the author was at IAS.

Contents

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 [HHHH09, NC10, Wat18, Oru19]. 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 [EPR35] and the violation of Bell's (style) inequalities [Bel64, CHSH69]. 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 [Vid03]. Entanglement is also crucial in a variety of protocols such as quantum key distribution [BB14], teleportation [BBC+93], interactive proof systems [JNV+20], 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 [AB09]. The quantum setting allows for various and vast generalizations of these classical notions [VW16]. 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 [Wat00]. 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) [KMY03] (see Definition 2.3). 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 [HM13].

Determining the complexity of QMA(2) is a major open problem in quantum complexity. Contrary to many other quantum proof systems (e.g., QIP [JJUW11] and MIP* [JNV+20]), 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 semi-definite programming hierarchies [DPS04, BKS17, HNW17], quantum de Finetti theorems [KM09, BH13, BCY11], and carefully designed nets [BH15, SW12], 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 [BT09] 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 $\mathrm{QMA}_{\log}(2)$. Since $\mathrm{QMA}_{\log}(1) \subseteq \mathrm{BQP}$ due to Marriott and Watrous [MW05], Blier and Tapp's work 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 QMA(2) = NEXP 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 [Bei10, GNN12, CF13]. However, all these results 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. [ABD$^+$08]. 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 [HM13], 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 $\mathrm{NP} \subseteq \mathrm{QMA}(2)$.

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 Definition 2.5) 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 [RS10, BBH$^+$12]. 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 [Kho02, KR03, KKMO04, Rag08, KO09, Kho10]. 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

[1] In terms of edge expansion.

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 [RST10].

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 to be NP-hard [DKK+18a, KMS18, BKS19]. 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.*

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 above characterization is shown 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

[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.

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.

At this moment, a natural question remains is what is the relationship between $\mathrm{QMA}^+(2)$ and QMA(2). We prove that they can simulate each other to some extent. In particular, a QMA(2) protocol can always be simulated by a $\mathrm{QMA}^+(2)$ protocol without any loss in the completeness and soundness. This direction is not surprising at this point, as $\mathrm{QMA}^+(2) =$ NEXP. On the other direction, we show that a QMA(2) protocol can also simulate the $\mathrm{QMA}^+(2)$ protocol at the cost of worsening the soundness by a multiplicative factor ofat most 4. An immediate corollary is the following:

Corollary 1.5. *If* $\mathrm{NEXP} \subseteq \mathrm{QMA}^+(2)$ *with a completeness and soundness gap at least* $3/4 + 1/\mathrm{poly}(n)$*. Then*

$$\mathrm{QMA}(2) = \mathrm{NEXP}.$$

Therefore, a strong enough gap amplification for $\mathrm{QMA}^+(2)$ would solve the long-standing open problem of characterizing QMA(2).

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 $\mathrm{QMA}^+(2) = \mathrm{NEXP}$. In the last section, Section 8, we discuss the relationship between the complexity class of $\mathrm{QMA}^+(2)$ and QMA(2), pointing out a potential direction towards the "QMA(2) = NEXP?" problem.

2 Preliminaries

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

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

Let Σ be an arbitrary non-empty alphabet. For any strings $s \in \Sigma^*$, $|s|$ denotes its length. For and $I \subset \mathbb{N}$, we denote the substring of s with index in I by $s|_I$. Thus, $s|_{\{i_1, i_2, \ldots, i_m\}} = s_{i_1} s_{i_2} \cdots s_{i_m}$. For two strings s, t, we use $s \prec t$ to mean that t is a prefix of s.

We adopt the Dirac notation for vectors representing quantum states, e.g., $|\psi\rangle, |\phi\rangle$, etc. In this paper, all the vectors of the form $|\psi\rangle$ are unit vectors. Given any pure state $|\psi\rangle$, we adopt the convention that its density operator is denoted by the Greek letter without the "ket", e.g. $\psi = |\psi\rangle\langle\psi|$. Given any set $H \subseteq \mathcal{H}$ for some Hilbert space $\mathcal{H}$, $\mathrm{conv}\,(H)$ is the convex hull of H. One particularly interesting set of states to us is the *separable* states. We adopt the following notation for the set of density operators regarding separable states,

$$\mathrm{SEP}(d, r) \coloneqq \mathrm{conv}\left(\psi_1 \otimes \cdots \otimes \psi_r \mid |\psi_1\rangle, \ldots, |\psi_r\rangle \in \mathbb{C}^d\right).$$

A related notion is that of separable measurement, whose formal definition is given below.

Definition 2.1 (Separable measurement). *A measurement $M = (M_0, M_1)$ is separable if in the yes case, the corresponding Hermitian matrix M_1 can be represented as a conical combination of two operators acting on the first and second parts, i.e., for some distribution μ over the tensor product of PSD matrices α and β on the corresponding space,*

$$M_1 = \int \alpha \otimes \beta \,\mathrm{d}\mu.$$

We record the following well-known fact. An interested reader is referred to [Har13] for a formal proof.

Fact 2.2 (Folklore). *The swap test is separable.*

Matrix Analysis Given any matrix $M \in \mathbb{C}^{n \times n}$, $M^\dagger$ is its conjugate transpose. Let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ denote its singular values. Then the trace norm $\|\cdot\|_1$, Frobenius norm $\|\cdot\|_F$ are defined as below

$$\|M\|_1 = \sum_i \sigma_i, \qquad \|M\|_F = \sqrt{\sum_i \sigma_i^2}.$$

The Frobenius norm also equals the square root of sum of squared lengths, i.e., $\|M\|_F = \sqrt{\sum_{i,j} |M(i,j)|^2}$.

For a positive semi-definite (PSD) matrix M, $\|M\|_F = \sqrt{\operatorname{Tr} M^2}$. For two PSD matrices, there is one (of many) analogous matrix Cauchy-Schwarz inequality.

$$\operatorname{Tr}(\sigma\rho) \leq \|\sigma\|_F \cdot \|\rho\|_F. \tag{2.1}$$

We adopt the notation $\succeq$ to denote the partial order that $\sigma \succeq \rho$ if $\sigma - \rho$ is positive semi-definite.

2.1 Quantum Merlin-Arthur with Multiple Provers

The class QMA(k) can be formally defined in more generality as follows.

Definition 2.3 ($\mathrm{QMA}_\ell(k,c,s)$). *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 $\mathcal{L}_{\text{yes}}, \mathcal{L}_{\text{no}} \subseteq \{0,1\}^*$ 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 $x \in \mathcal{L}_{\text{yes}}$, 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 $x \in \mathcal{L}_{\text{no}}$, 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

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

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 2.4 (Harrow and Montanaro [HM13]). *For any $\ell, k, 0 \leq s < c \leq 1$,*

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

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

The class $\mathrm{QMA}(k)^+$ is formally defined in more generality as follows.

Definition 2.5 ($\mathrm{QMA}^+_\ell(k,c,s)$). *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 $\mathcal{L}_{\text{yes}}, \mathcal{L}_{\text{no}} \subseteq \{0,1\}^*$ 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 $x \in \mathcal{L}_{\text{yes}}$, 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.* $\Pr[V(x, |\psi_1\rangle \otimes \cdots \otimes |\psi_{k(n)}\rangle) \text{ accepts}] \geq c(n)$.
>
> **Soundness:** *If $x \in \mathcal{L}_{\text{no}}$, then for every 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* $\Pr[V(x, |\psi_1\rangle \otimes \cdots \otimes |\psi_{k(n)}\rangle) \text{ accepts}] \leq s(n)$.

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 2.4 generalizes to $\mathrm{QMA}^{+}(k)$ as well. Because the product test works for general states, it in particular works for states with non-negative amplitudes. Furthermore, the closest product state to a state with non-negative amplitudes also has non-negative amplitudes.

Theorem 2.6. *For any $\ell, k, 0 \le s < c \le 1$,*

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

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

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 ψ and ϕ, denoted $\mathrm{D}(\psi,\phi)$, is

$$\frac{1}{2}\|\psi-\phi\|_1 = \frac{1}{2}\operatorname{Tr}\sqrt{(\psi-\phi)^\dagger(\psi-\phi)}. \tag{2.2}$$

We also use the notation $\mathrm{D}(|\psi\rangle,|\phi\rangle)$ if we want to emphasize that ψ and ϕ are pure states. The following fact provides an alternative definition for trace distance between pure states.

Fact 2.7. *The trace distance between $|\phi\rangle$ and $|\psi\rangle$ is given by $\mathrm{D}(|\phi\rangle,|\psi\rangle) = \sqrt{1-|\langle\phi\,|\,\psi\rangle|^2}$.*

The trace distance remains small under the tensor product.

Fact 2.8. *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*

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

Proof. By the alternative definition of the trace distance,

$$\begin{aligned}\mathrm{D}(|\psi_0\rangle\otimes|\psi_1\rangle,|\phi_0\rangle\otimes|\phi_1\rangle)^2 &= (1-|\langle\psi_0,\phi_0\rangle|^2|\langle\psi_1,\phi_1\rangle|^2)\\ &\le (1-|\langle\psi_0,\phi_0\rangle|^2 + 1-|\langle\psi_1,\phi_1\rangle|^2)\\ &= \mathrm{D}(|\psi_0\rangle,|\phi_0\rangle)^2 + \mathrm{D}(|\psi_1\rangle,|\phi_1\rangle)^2,\end{aligned}$$

where the second step can be easily verified as $-a^2b^2+b^2 \le 1-a^2$ for any $a,b \in [0,1]$. □

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

Fact 2.9. *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 2.10 (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*

$$\frac{1}{2} + \frac{|\langle\phi\,|\,\psi\rangle|^2}{2}.$$

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

Remark 2.11. *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 for states with nonnegative amplitudes.

Claim 2.12. *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 $\langle u, v\rangle^2 \geq 1 - \varepsilon$. Then*
$$\left|\langle u, z\rangle^2 - \langle v, z\rangle^2\right| \leq 3\sqrt{\varepsilon}.$$

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

Proof. The first item is bounded as below

$$\begin{aligned}
|\langle u, z\rangle^2 - \langle v, z\rangle^2| &= |\langle u - v, z\rangle| \cdot |\langle u, z\rangle + \langle v, z\rangle| \\
&\leq 2\|u - v\| \\
&\leq 2\sqrt{2 - 2\langle u, v\rangle} \\
&\leq 2\sqrt{2 - 2\sqrt{1 - \varepsilon}} \\
&\leq 3\sqrt{\varepsilon},
\end{aligned}$$

where the last step can be verified by elementary calculus.

Next, we prove the second item as follows

$$\begin{aligned}
\langle u, v\rangle^2 &= \left(\frac{2 - \|u - v\|^2}{2}\right)^2 \\
&\geq \left(\frac{2 - \|u - z\|^2 - \|v - z\|^2}{2}\right)^2 \\
&= (\langle u, z\rangle + \langle v, z\rangle - 1)^2 \\
&\geq (2\sqrt{1 - \varepsilon} - 1)^2 \\
&= 5 - 4\varepsilon - 4\sqrt{1 - \varepsilon} \\
&\geq 1 - 2\varepsilon,
\end{aligned}$$

where the last step holds because $\sqrt{1 - \varepsilon} \leq 1 - \varepsilon/2$. $\square$

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

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

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

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

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 [HM13], these protocols can be converted into two-proof protocols with a constant multiplicative increase in the proof size.

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

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

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 relies 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

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

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

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

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

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

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

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 [BL06] 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 vice versa 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

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

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

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

verifier becomes exponential and this is also intolerable to 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 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

[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.

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 [ABD+08]. 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).

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.*

Proof. Let S and T be the representative set of Ψ and Φ, respectively. Simply note that

$$|S \cap T| \geq (1 - \varepsilon - \gamma)k,$$

and for any $i, j \in S \cap T$,

$$\mathrm{D}(|\psi_i\rangle \otimes |\phi_i\rangle, |\psi_j\rangle \otimes |\phi_j\rangle)^2 \leq \mathrm{D}(|\psi_i\rangle, |\phi_i\rangle)^2 + \mathrm{D}(|\psi_j\rangle, |\phi_j\rangle)^2 \leq \varepsilon + \gamma,$$

where the first inequality is due to Fact 2.8. □

As commented earlier that we should treat an ε-tilted state as a single state conceptually. Now we make this comment more formal. When we apply some quantum algorithm $\mathcal{A}$ to Ψ, we mean apply $\mathcal{A}$ to all the states in Ψ. For any $f : \mathbb{C}^n \to \mathbb{C}$, when we evaluate f on Ψ, we mean the expected value of f on all states in Ψ, i.e.,

$$f(\Psi) = \mathop{\mathbb{E}}_{|\psi\rangle \in \Psi}[f(|\psi\rangle)].$$

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

Proposition 4.3. *For any quantum algorithm $\mathcal{A}$, let $\mathcal{A}(|\psi\rangle)$ denote the probability that $\mathcal{A}$ accepts $|\psi\rangle$. Let Ψ be an ε-tilted state, and $|\psi\rangle$ any representative state of Ψ. Then*

$$|\mathcal{A}(|\psi\rangle) - \mathcal{A}(\Psi)| \leq 3\sqrt{\varepsilon}. \tag{4.1}$$

Furthermore, when apply $\mathcal{A}$ to Ψ, let α be the fraction of accepted executions of $\mathcal{A}$. Then

$$\Pr[|\alpha - \mathcal{A}(\Psi)| \geq \sqrt{\varepsilon}] \leq \exp(-\varepsilon|\Psi|/2), \tag{4.2}$$

and therefore,

$$\Pr[|\alpha - \mathcal{A}(|\psi\rangle)| \geq 4\sqrt{\varepsilon}] \leq \exp(-\varepsilon|\Psi|/2). \tag{4.3}$$

Proof. Let $\Psi = \{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle\}$ and S be the representative set for Ψ. Then

$$\mathcal{A}(\Psi) = \frac{1}{k}\sum_{i=1}^{k} \mathcal{A}(|\psi_i\rangle) = \frac{|S|}{k} \mathop{\mathbb{E}}_{i\in S} \mathcal{A}(|\psi_i\rangle) + \frac{k-|S|}{k} \mathop{\mathbb{E}}_{i\notin S} \mathcal{A}(|\psi_i\rangle).$$

It follows that

$$(1-\varepsilon) \mathop{\mathbb{E}}_{i\in S} \mathcal{A}(|\psi_i\rangle) \leq \mathcal{A}(\Psi) \leq (1-\varepsilon) \mathop{\mathbb{E}}_{i\in S} \mathcal{A}(|\psi_i\rangle) + \varepsilon.$$

Therefore,

$$\left| \mathop{\mathbb{E}}_{i\in S} \mathcal{A}(|\psi_i\rangle) - \mathcal{A}(\Psi) \right| \leq \varepsilon. \tag{4.4}$$

By Fact 2.9 and the definition of ε-tilted state, for any $j \in S$,

$$\left| \mathop{\mathbb{E}}_{i\in S} \mathcal{A}(|\psi_i\rangle) - \mathcal{A}(|\psi_j\rangle) \right| \leq \sqrt{\varepsilon}. \tag{4.5}$$

Combining (4.4) and (4.5), we obtain (4.1). The furthermore part follows by Chernoff bound. □

By (4.3), it suffices to understand the typical behavior of the representative state in an ε-tilted state.

4.2 Symmetry Test

The symmetry test is described below.

<u>Symmetry Test</u>
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 4.4 (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.

Let $\mathcal{N}(i) := \{a_j : \mathrm{D}(a_i, a_j) \leq \sqrt{\varepsilon}/2\}$ be the set of vectors that are close to a_i, and $\mathcal{B} := \{i : |\mathcal{N}(i)| \leq k/2\}$ the set of vector a_i who is far from at least half of the other vectors. Finally for a random matching define

$$\ell(\pi) = |\{i : \pi(i) \notin \mathcal{N}(i)|,$$

twice the number of distant pairs in the matching.

Claim 4.5. *Suppose $|\mathcal{B}| = \gamma k$ for any constant $\gamma \in (0, 1]$. Then*

$$\Pr_{\pi}\left[\ell(\pi) \geq \frac{\gamma k}{18}\right] \geq 1 - \exp(-\Theta(\gamma k)).$$

Proof. Without loss of generality, let $\mathcal{B} = \{a_1, a_2, \ldots, a_m\}$. Assume $m \leq 2k/3$, in another word, $\gamma \leq 2/3$. Consider the matching procedure: For i from 1 to $k/2$, find one vector in $\mathcal{B}$ if there is one that hasn't been matched yet, and pair it with a random unmatched vector; if all vectors in $\mathcal{B}$ have been matched, pair two random unmatched vectors. Let X_i be the indicator function that at time i, the paired vectors are $\sqrt{\varepsilon}/2$ far away in trace distance. Then,

$$\begin{aligned}
\sum_{i=1}^{\lceil m/2\rceil} \mathbb{E}[X_i] &\geq \frac{1}{2} + \left(\frac{k/2-2}{k}\right) + \cdots + \left(\frac{k/2 - 2\lceil m/2\rceil + 2}{k}\right) \\
&= \frac{1}{2k}\left(k - 2\lceil m/2\rceil + 2\right)\lceil m/2\rceil \\
&\geq \frac{1}{4k}(k-m)m \\
&\geq \frac{1}{4}\gamma(1-\gamma)k.
\end{aligned}$$

Since $S_j = \sum_{i=1}^{j} X_i - \mathbb{E}[X_i]$ is a martingale, we have

$$\Pr\left[\sum_{i=1}^{\lceil m/2\rceil}(X_i - \mathbb{E}[X_i]) \leq -t\right] \leq \exp\left(-\frac{t^2}{m+1}\right).$$

Set $t = \gamma(1-\gamma)k/12$, our claim holds. When $\gamma > 2/3$, the claim can be verified by comparing it with the case of $\gamma = 2/3$. □

Lemma 4.6. *Suppose $|\mathcal{B}| \geq \gamma k$, for any constant $\gamma \in (0, 1]$. Then the probability that the symmetry test passes with probability at most* $\exp(-\Theta(\varepsilon\gamma k))$.

Proof. Fix any permutation π, the symmetry test passes with probability at most $(1 - \varepsilon/8)^{\ell(\pi)}$. Therefore using Claim 4.5, we have

$$\begin{aligned}
&\Pr[\text{Symmetry test passes}] \\
&\quad\leq \Pr\left[\ell(\pi) < \frac{\gamma k}{18}\right] + (1-\varepsilon/8)^{\gamma k/18} \\
&\quad\leq \exp(-\Theta(\gamma k)) + \exp(-\Theta(\varepsilon\gamma k)).
\end{aligned}$$

□

At the point, Theorem 4.4 is a straightforward corollary of the above lemma.

Proof of Theorem 4.4. Let $\mathcal{G} = [k] \setminus \mathcal{B}$. Note that for any $i, j \in \mathcal{G}$,

$$\mathcal{N}(i) \cap \mathcal{N}(j) \neq \varnothing.$$

Thus $\mathrm{D}(a_i, a_j) \leq \sqrt{\varepsilon}$ by triangle inequality. Thus, $|\mathcal{G}| \geq (1-\varepsilon)k$ implies that Ψ is an ε-tilted state. By contraposition, if Ψ is not an ε-tilted state, then $|\mathcal{B}| > \varepsilon k$. It follows that, by Lemma 4.6, the symmetry test passes with probability at most $\exp(-\Theta(\varepsilon^2 k))$. □

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.

<u>Sparsity test I (with precision ε)</u>
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 $\lambda \leq 1/2 + \sqrt{\varepsilon}$, where α, β and λ are the fractions of accepted SwapTests in (i), (ii), and (iii), respectively.
Output: α, β, λ.

Theorem 4.7 (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,

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

(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 $\gamma > 0$,*

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

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

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

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

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

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

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

We first prove the following lemma useful in the soundness part.

Lemma 4.8. *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) $\langle u, v\rangle^2 \leq \varepsilon$,
(ii) $|\langle u, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 - \delta| \leq \varepsilon$,
(iii) $|\langle v, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 - (1-\delta)| \leq \varepsilon$.

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

$$\|u|_S\|^2 \geq 1-\gamma. \tag{4.6}$$

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

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

we have

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

Proof. Let

$$U = \left\{i : u_i \geq \sqrt{\frac{\gamma}{n}}\right\}, V = \left\{i : v_i \geq \sqrt{\frac{\gamma}{n}}\right\},$$

for some γ to be determined later. U will be the set S in the statement. Note that by our definition of U, V,

$$\|u|_{\overline{U}}\|^2, \|v|_{\overline{V}}\|^2 \leq \gamma, \tag{4.7}$$
$$\|u|_U\|^2, \|v|_V\|^2 \geq 1-\gamma. \tag{4.8}$$

We claim that

$$|U| \geq (\delta - \varepsilon)n, \tag{4.9}$$
$$|V| \geq (1-\delta-\varepsilon)n, \tag{4.10}$$
$$|U \cap V| \leq \frac{\sqrt{\varepsilon}}{\gamma}n. \tag{4.11}$$

We verify (4.9), and (4.10) will follow the same reasoning. Note

$$\delta - \varepsilon \leq \left\langle u|_U, \frac{\mathbf{1}_{[n]}}{\sqrt{n}}\right\rangle^2 \leq \|u|_U\|^2 \frac{|U|}{n},$$

where the first inequality is given; the second step uses Cauchy-Schwarz. Rearranging the terms, we get (4.9). Next, we obtain (4.11),

$$\sqrt{\varepsilon} \geq \sum_{i \in U \cap V} u_i v_i \geq |U \cap V| \frac{\gamma}{n},$$

where the first step uses (i), and second step follows the definition of U and V. In view of (4.11), we are done by rearranging the terms. By (4.10)-(4.11), we can conclude

$$\begin{aligned} |U| &\leq |U \cup V| - |V| + |U \cap V| \\ &\leq n - (1 - \delta - \varepsilon)n + \frac{\sqrt{\varepsilon}}{\gamma} n \\ &\leq \left(\delta + \frac{2\sqrt{\varepsilon}}{\gamma}\right) n. \end{aligned} \tag{4.12}$$

This finishes the proof of the first part of the lemma. For the furthermore part, calculate:

$$\begin{aligned} \left\langle u, \frac{\mathbf{1}_U}{\sqrt{|U|}} \right\rangle &= \frac{1}{\sqrt{|U|}} \langle u|_{U}, \mathbf{1}_{[n]} \rangle \\ &= \frac{1}{\sqrt{|U|}} (\langle u, \mathbf{1}_{[n]} \rangle - \langle u|_{\overline{U}}, \mathbf{1}_{[n]} \rangle) \\ &\geq \sqrt{\frac{n}{|U|}} (\sqrt{\delta - \varepsilon} - \sqrt{\gamma}) \\ &\geq \sqrt{\frac{\delta - \varepsilon}{\delta + 2\sqrt{\varepsilon}/\gamma}} - \sqrt{\frac{\gamma}{\delta + 2\sqrt{\varepsilon}/\gamma}} \\ &\geq \sqrt{1 - \frac{2\sqrt{\varepsilon}/\gamma + \varepsilon}{\delta + 2\sqrt{\varepsilon}/\gamma}} - \sqrt{\frac{\gamma}{\delta}}, \end{aligned}$$

where the third step uses (ii) given in the lemma statement, and (4.7) with Cauchy-Schwarz inequality; the fourth step uses (4.12). Set $\kappa^6 = \varepsilon/\delta^4, \gamma = \kappa^2 \delta$, then

$$\left\langle u, \frac{\mathbf{1}|_U}{\sqrt{|U|}} \right\rangle \geq 1 - O(\kappa). \qquad \square$$

Equipped with the above lemma, we move on to prove Theorem 4.7.

Proof of Theorem 4.7. The completeness part is a straightforward application of Chernoff bound. So we focus on the soundness part. Let R and T be the representative set of Ψ and Φ, respectively. When Ψ, Φ are ε-tilted states, then $\Psi_0, \Psi_1, \Phi_0, \Phi_1$ are 2ε-tilted states, and $\Psi_1 \otimes \Phi_1$ is a 4ε-tilted state by Proposition 4.2. By Proposition 4.3, we have for any $i \in R$, and $j \in T$,

$$\Pr\left[\left|\langle u_i, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 + 1 - 2\alpha\right| > 12\sqrt{\varepsilon}\right] \leq \exp(-\varepsilon k), \tag{4.13}$$
$$\Pr\left[\left|\langle v_j, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 + 1 - 2\beta\right| > 12\sqrt{\varepsilon}\right] \leq \exp(-\varepsilon k), \tag{4.14}$$
$$\Pr\left[\left|\langle u_i, v_j\rangle^2 + 1 - 2\lambda\right| > 16\sqrt{\varepsilon}\right] \leq \exp(-2\varepsilon k). \tag{4.15}$$

Set $\delta = 2\alpha - 1$. Note that the test passes only if $|(2\alpha - 1) + (2\beta - 1) - 1| \leq 2\sqrt{\varepsilon}$. Together with (4.13) and (4.14), it implies that

$$|\langle u_i, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 - \delta| \leq 12\sqrt{\varepsilon}, \tag{4.16}$$
$$|\langle v_j, \mathbf{1}_{[n]}/\sqrt{n}\rangle^2 - (1-\delta)| \leq 14\sqrt{\varepsilon}. \tag{4.17}$$

Therefore, if either (4.16) or (4.17) fails, the protocol accepts with probability at most $\exp(-\varepsilon k)$.

Moreover, the test passes only if $2\lambda - 1 \leq 2\sqrt{\varepsilon}$. Thus when the following does not hold the test fails with probability at least $1 - \exp(-2\varepsilon k)$.

$$\langle u_i, v_j\rangle^2 \leq 18\sqrt{\varepsilon}. \tag{4.18}$$

Now suppose (4.16), (4.17) and (4.18) are true for some $i \in R$ and $j \in T$. By Lemma 4.8, we have:

(i) For any γ, there is subset $S \subseteq [n]$ such that $|S| \leq ((2\alpha-1)+9\varepsilon^{1/4}/\gamma)n$, and $\|u_i|_S\|^2 \geq 1-\gamma$.

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

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

and

$$\langle u_i, \mathbf{1}_S/\sqrt{|S|}\rangle \geq 1 - O\left(\frac{\varepsilon^{1/12}}{(2\alpha-1)^{2/3}}\right).$$

Since for any representative state $u \in \Psi$, $\mathrm{D}(u, u_i) \leq \sqrt{\varepsilon}$, the above two items implies (i) and (ii) in the theorem statements. Therefore, if either (i) or (ii) in the theorem statements does not hold, then one of (4.16), (4.17) and (4.18) is not true, failing the sparsity test with probability at least $1 - \exp(-\varepsilon k)$. □

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 4.9 (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:*

$$\mathrm{D}(\Psi, \mathcal{S}_\gamma) \leq O\left(\frac{\varepsilon^{1/24}}{\gamma^{1/3}}\right). \tag{4.19}$$

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

Proof. To prove the first part, it suffices to show that assuming Theorem 4.7 (ii) holds then (4.19) holds. Suppose $2\alpha - 1 = (1+\varepsilon')\gamma$. Then we assume that $|\varepsilon'| \leq \sqrt{\varepsilon}/\gamma$, since otherwise the sparsity test rejects immediately. Note that ε' is a very tiny number in absolute value. By Theorem 4.7 (ii), there is constant c, C such that for any representative state $|\psi\rangle \in \Psi$,

$$\mathrm{D}(|\psi\rangle, \mathcal{S}_{\gamma'}) \leq \frac{C\varepsilon^{1/24}}{((1+\varepsilon')\gamma)^{1/3}}, \tag{4.20}$$

where

$$|\gamma' - (2\alpha - 1)| \leq c\varepsilon^{1/12}(1+\varepsilon')^{1/3}\gamma^{1/3}.$$

Therefore,

$$\begin{aligned} |\gamma - \gamma'| &\leq |\gamma - (2\alpha-1)| + |\gamma' - (2\alpha - 1)| \\ &\leq \varepsilon'\gamma + c\varepsilon^{1/12}(1+\varepsilon')^{1/3}\gamma^{1/3} \\ &\leq c'\varepsilon^{1/12}\gamma^{1/3}, \end{aligned} \tag{4.21}$$

where the last step holds due to that we set $\varepsilon < \gamma^{4/5}$, and c' is some constant. Note that for any $S \subseteq T \subseteq [n]$, we have,

$$\mathrm{D}\left(\frac{\mathbf{1}_S}{\sqrt{|S|}}, \frac{\mathbf{1}_T}{\sqrt{|T|}}\right) = \sqrt{1 - \left(\frac{|S|}{\sqrt{|S||T|}}\right)^2} = \sqrt{\frac{|T| - |S|}{|T|}}. \tag{4.22}$$

By (4.20)-(4.22) and triangle inequality, for some absolute constant C',

$$\begin{aligned} \mathrm{D}(|\psi\rangle, \mathcal{S}_{\gamma}) &\leq \frac{C\varepsilon^{1/24}}{(1+\varepsilon')^{1/3}\gamma^{1/3}} + \sqrt{|\gamma - \gamma'|/\gamma}, \\ &\leq C'\varepsilon^{1/24}\gamma^{-1/3}. \end{aligned} \tag{4.23}$$

The second part of the theorem is simply the completeness case from Theorem 4.7. $\square$

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 4.10 (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 $1 - \exp(-\Theta(qd^2k))$.

Proof. Fix an arbitrary representative state $|\psi\rangle$, let $|\phi\rangle \in \mathcal{S}_{1/q}$ be such that

$$\mathrm{D}(|\psi\rangle, |\phi\rangle) = \mathrm{D}(|\psi\rangle, \mathcal{S}_{1/q}) \leq d.$$

If $\mathrm{D}(|\psi\rangle, \mathcal{V}) \geq \sqrt{2qd} + d$, by triangle inequality

$$\mathrm{D}(|\phi\rangle, \mathcal{V}) \geq \mathrm{D}(|\psi\rangle, \mathcal{V}) - \mathrm{D}(|\psi\rangle, |\phi\rangle) \geq \sqrt{6qd}. \tag{4.24}$$

Say $S \subseteq [n] \times [q]$ of size n is such that

$$|\phi\rangle = \frac{1}{\sqrt{n}} \sum_{(i,v)\in S} |i\rangle|v\rangle.$$

For each $i \in [n]$, let $c_i := |\{(i, v) \in [n] \times [q] : (i, v) \in S\}|$. Let $Z := \{i : c_i = 0\}$. Then,

$$\mathrm{D}(|\phi\rangle, \mathcal{V}) = \sqrt{1 - \left(\frac{n - |Z|}{n}\right)^2} \implies \frac{|Z|}{n} \geq \frac{1}{2}\mathrm{D}(|\phi\rangle, \mathcal{V})^2.$$

When measuring the second register of $|\phi\rangle$ after the discrete Fourier transform, the probability $\tilde{p}$ that we observe $|0\rangle$ can be calculated as below,

$$\begin{aligned}
\tilde{p} &= \frac{\sum_{i\in[n]} c_i^2}{nq} \geq \frac{n - |Z|}{nq} \left(\frac{n}{n - |Z|}\right)^2 \\
&= \frac{1}{q} \cdot \frac{n}{n - |Z|} \geq \frac{1}{q}\left(1 + \frac{|Z|}{n}\right) \\
&\geq \frac{1}{q}(1 + \mathrm{D}(|\phi\rangle, \mathcal{V})^2/2),
\end{aligned} \tag{4.25}$$

where the second step follows by convexity. By (4.25) and (4.24),

$$\tilde{p} \geq (1 + 3qd)\,\frac{1}{q}.$$

Now let p be the probability that we observe 0 measuring the second register of $|\psi\rangle$ after applying Fourier transform, then by Fact 2.9,

$$p \geq (1 + 3qd)\frac{1}{q} - d \geq (1 + 2qd)\frac{1}{q}.$$

The first part of our lemma holds by Chernoff bound.

Now suppose that Ψ is a 0-tilted state from $\mathcal{V}$. Let $|\psi\rangle$ be the representative state of Ψ and let $\Pi|\psi\rangle$ denote the projection of $|\psi\rangle$ onto the subspace

$$\mathbb{C}^n \otimes \left(\frac{1}{\sqrt{q}} \sum_{v \in \Sigma} |v\rangle \right).$$

Thus $\|\Pi|\psi\rangle\|^2$ is the probability of observing $|0\rangle$, after applying the Fourier transform to and measuring the second register of $|\psi\rangle$. For any $|\psi\rangle \in \mathcal{V}$,

$$\|\Pi|\psi\rangle\|^2 = \frac{1}{q}.$$

It thus follows that in the validity test, we observe less than $1/q + d$ fraction of $|0\rangle$ with probability at least $1 - \exp(-\Theta(qd^2k))$. □

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

The small-set expansion problem arises in the context of the unique games conjecture [RS10, BBH+12]. The formal definition is given below.[11]

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 $\mathrm{QMA}^+_{O_\delta(\log(n))}(k,c,s)$ for some constant k, 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 $\mathrm{QMA}_{\log}(k)$ protocol described in Algorithm 5.4 is complete and sound for (η,δ)-SSE. More precisely, the theorem follows immediately from the following lemmas proven in Sections 5.1 and 5.2, respectively.

Lemma 5.6 (Completeness)**.** *The protocol in Algorithm 5.4 accepts any yes instance with probability at least $1-\eta$.*

Lemma 5.7 (Soundness)**.** *The protocol in Algorithm 5.4 accepts any no instance with probability at most $5/6 + O(\sqrt{\eta}\log(1/\eta))$.*

[11] This definition is slightly stronger than the original definition in [RS10].

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.,

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

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

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

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 I 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 $i, j \in \{1, 2, \dots, 2k\}$.

Since G is a d regular graph, its adjacency matrix A can be written as a sum of d permutation matrices $P_1, \dots, P_d$. 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.

5.1 Completeness Analysis

We now analyze the completeness of the protocol by proving the following lemma.

Lemma 5.6 (Completeness)**.** *The protocol in Algorithm 5.4 accepts any yes instance with probability at least* $1 - \eta$.

Proof. Suppose that G is the input graph of a yes instance where S is a non-expanding set of measure at most δ. We expect $4k$ unentangled quantum proofs of the form

$$|\psi_j\rangle = \frac{1}{\sqrt{|S|}} \sum_{i\in S} |i\rangle, \qquad \forall j \in \{1, 2, \dots, 2k\},$$
$$|\phi_j\rangle = \frac{1}{\sqrt{n-|S|}} \sum_{i\notin S} |i\rangle, \qquad \forall j \in \{1, 2, \dots, 2k\}.$$

The two symmetry tests accept with probability 1 since they are running on sets of equal states. The sparsity test accepts with probability at least $1-\eta$ by Theorem 4.7. It only remains to analyze the expansion test. Recall that A is the adjacency matrix of the graph instance, the assumption that $\Phi_G(S) \leq \eta$ can be expressed as

$$\frac{1}{d} \langle A\psi_1 \,|\, \psi_1 \rangle \;\geq\; 1-\eta\,.$$

Then, using Jensen's inequality we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{r\in[d]} \left[|\langle P_r\psi_1 \,|\, \psi_1\rangle|^2 \right] \;&\geq\; \left(\mathop{\mathbb{E}}_{r\in[d]} \left[|\langle P_r\psi_1 \,|\, \psi_1\rangle| \right] \right)^2 \\
&=\; \left\langle \mathop{\mathbb{E}}_{r\in[d]} [P_r]\,\psi_1 \,\middle|\, \psi_1 \right\rangle^2 \\
&=\; \left\langle \frac{1}{d} A\psi_1 \,\middle|\, \psi_1 \right\rangle^2 \;=\; (1-\eta)^2\,.
\end{aligned}$$

In this case, the swap test on $(P_r|\psi_1\rangle, |\psi_2\rangle)$ accepts with probability at least $1/2+(1-\eta)^2/2 \geq 1-\eta$. Therefore, the entire protocol accepts with probability at least $1-\eta$ as claimed. $\square$

5.2 Soundness Analysis

We will establish the soundness of the protocol by showing the following lemma.

Lemma 5.7 (Soundness)**.** *The protocol in Algorithm 5.4 accepts any no instance with probability at most* $5/6 + O(\sqrt{\eta}\log(1/\eta))$.

First, we record a simple fact about expander graphs.

Fact 5.8. *Suppose that the graph G is (η,δ)-SSE. Then G is also $((c+1)\eta, (1+c\eta)\delta)$-SSE, for any $c \geq 0$.*

Proof. For any $\delta n < |S| \leq (1+c\eta)\delta n$, let $T \subseteq S$ be such that $|T| = \delta n$. Then

$$|E(S,S)| \leq |E(T,T)| + d|S\setminus T| \leq \eta d|T| + d|S|\frac{|S\setminus T|}{|S|} \leq (1+c\eta)d|S|. \qquad \square$$

We will also need the following analytic version of the SSE property.

Definition 5.9 (Analytic SSE)**.** *Let $\eta,\delta \in (0,1)$. We say that a graph $G=(V,E)$ with normalized adjacency matrix A is (η,δ)-analytic SSE if for every $v \in \mathbb{R}^V$ of ℓ_2-norm 1 and support of measure at most δ it holds that*

$$|\langle Av, v\rangle| \leq \eta\,.$$

This analytic property is implied by the SSE property as we show in the following proposition (proved in Section 5.3).

Proposition 5.10. *If G is (η,δ)-SSE, then G is $(O(\sqrt{\eta}(\log(1/\eta)+1)), \delta)$-analytic SSE.*

Assuming Proposition 5.10, we now proceed to the proof of Lemma 5.7.

Proof of Lemma 5.7. Assume that $|\Psi\rangle = \{|\psi_1\rangle, \ldots, |\psi_{2k}\rangle\}$ and $|\Phi\rangle = \{|\phi_1\rangle, \ldots, |\phi_{2k}\rangle\}$ are ε-tilted states. We call this event $\mathcal{E}_1$. If $\mathcal{E}_1$ does not hold, the symmetry test accepts with probability at most $\sqrt{\eta}$ for $k = \Omega(\varepsilon^{-2}\log(1/\eta))$.

Now we further assume that there is $S \subseteq [n]$, such that

$$|S| \leq (1+6\eta)\delta n,$$

and let Π_S be the projection into the subspace corresponding to S, then for any representative state $|\psi_i\rangle$,

$$\|\Pi_S|\psi_i\rangle\|^2 \geq 1 - 1.1\eta.$$

This is the second event $\mathcal{E}_2$. By our choice of parameters and the fact that if $2\alpha - 1 > (1+\eta)\delta$ the sparsity test fails immediately, we can assume that

$$(2\alpha - 1) + 9\varepsilon^{1/4}/\eta \leq (1+6\eta)\delta,$$
$$20\sqrt{\varepsilon} \leq \eta.$$

Therefore, by Theorem 4.7 (i), the sparsity test accepts with probability at most $\sqrt{\eta}$ by our choice of parameters if $\mathcal{E}_2$ does not hold.

Conditioning on $\mathcal{E}_1$ and $\mathcal{E}_2$, we analyze the probability that the expansion test passes. Let's say the two proofs we get for the expansion test are $|\psi_1\rangle, |\psi_2\rangle$. With probability at least $(1-2\varepsilon)^2 \geq 1 - 4\varepsilon$, both are representative, thus satisfying that their mass projected on to the coordinates of S is at least $1 - \eta$. We call this event $\mathcal{E}_3$. Let

$$|\pi_1\rangle = \frac{\Pi_S|\psi_1\rangle}{\|\Pi_S|\psi_1\rangle\|},\ |\pi_2\rangle = \frac{\Pi_S|\psi_2\rangle}{\|\Pi_S|\psi_2\rangle\|}.$$

It follows that

$$\langle \pi_1 \mid \psi_1\rangle^2 = \|\Pi|\psi_1\rangle\| \geq 1 - 1.1\eta. \tag{5.1}$$

Let $\delta_0 = (1+6\eta)\delta$. By Proposition 5.10, the *analytic* $(O(\sqrt{\eta}(\log(1/\eta)+1)), \delta_0)$-SSE property follows from the (η, δ)-SSE assumption and Fact 5.8. To determine the expected acceptance probability of the swap test, we first bound the average value of $|\langle P_r\pi_1 \,|\, \pi_1\rangle|$ over the random choice of r obtaining

$$\begin{aligned}
\mathop{\mathbb{E}}_{r\in[d]}\left[|\langle P_r\pi_1 \,|\, \pi_1\rangle|\right] &= \left\langle \mathop{\mathbb{E}}_{r\in[d]}\left[P_r\right]\pi_1 \,\middle|\, \pi_1 \right\rangle \\
&= \frac{1}{d}\left\langle A\pi_1 \,|\, \pi_1\right\rangle \\
&\leq O(\sqrt{\eta}(\log(1/\eta)+1))\,,
\end{aligned}$$

where the first step holds because the entries of $|\psi_1\rangle$, $|\psi_2\rangle$ and P_r, for every r, are non-negative real numbers. Now, it follows that

$$\begin{aligned}
\mathop{\mathbb{E}}_{r\in[d]}\left[|\langle P_r\psi_1 \,|\, \psi_2\rangle|^2\right] &\leq \mathop{\mathbb{E}}_{r\in[d]}\left[|\left\langle P_r\psi_1 \,|\, \psi_1\right\rangle|^2\right] + 3\sqrt{\varepsilon} \\
&\leq \mathop{\mathbb{E}}_{r\in[d]}\left[|\left\langle P_r\pi_1 \,|\, \pi_1\right\rangle|^2\right] + 3\sqrt{\varepsilon} + \sqrt{1.1\eta} \\
&\leq \mathop{\mathbb{E}}_{r\in[d]}\left[\langle P_r\pi_1 \,|\, \pi_1\rangle\right] + 3\sqrt{\varepsilon} + \sqrt{1.1\eta} \\
&= O(\sqrt{\eta}\log(1/\eta)),
\end{aligned}$$

where the first step follows Claim 2.12 (i); the second step is due to Fact 2.9 and the bound $\mathrm{D}(|\pi_1\rangle, |\psi_1\rangle) \leq \sqrt{1.1\eta}$ that follows (5.1). Hence, the swap test on $(P_r|\psi_1\rangle, |\psi_2\rangle)$ accepts with probability at most $1/2 + O(\sqrt{\eta}\log(1/\eta))$.

To conclude, if $\mathcal{E}_1$ (or $\mathcal{E}_2$) does not hold with probability at least $1/3 \times (1 - \sqrt{\eta})$, the protocol chooses the symmetry test (or sparsity test) and rejects. If both $\mathcal{E}_1$ and $\mathcal{E}_2$ hold, then the protocol chooses the expansion test with probability 1/3 and rejects with probability at least $(1/2 - O(\sqrt{\eta}\log(1/\eta)))$ conditioning on $\mathcal{E}_3$, which happens with probability at least $1 - 4\varepsilon$. Hence the protocol accepts with probability at most $5/6 + O(\sqrt{\eta}\log(1/\eta))$. □

5.3 The Analytic SSE Property

In this section, we will establish the *analytic* SSE property from the usual SSE property.

Proposition 5.10. *If G is (η, δ)-SSE, then G is $(O(\sqrt{\eta}(\log(1/\eta) + 1)), \delta)$-analytic SSE.*

In a seminal work on 2-lifts of graphs [BL06], Bilu and Linial found conditions under which bounding the quadratic form $\langle Au, u\rangle$ of a matrix A for *arbitrary* vector u follows from bounds on much simpler "flat" indicator vectors $\langle A\mathbf{1}_S, \mathbf{1}_T\rangle$. Our goal is to use a version of their result adapted for vectors of small support as arising in our application. More precisely, we will need an inequality of the form $|\langle A\mathbf{1}_S, \mathbf{1}_T\rangle| \leq \eta(|S| + |T|)$ for every disjoint $S, T \subseteq V$ of support at most δ. We first show that this inequality is indeed satisfied by the adjacency matrix of SSE graph.

Lemma 5.11. *Suppose $G = (V, E)$ is a d-regular (η, δ)-SSE with adjacency matrix A (not normalized). If $S, T \subseteq V$ are disjoint sets with $|S| + |T| \leq \delta\,|V|$, then*

$$\langle A\mathbf{1}_S, \mathbf{1}_T\rangle \leq 2\sqrt{\eta}d\sqrt{|S|\,|T|}.$$

Proof. Let S, T be as in the assumption of the claim. Without loss of generality, assume that $|S| \leq |T|$. If $|S| \leq \eta\,|T|$, then we can use the trivial bound by the fact that A is the adjacency matrix of a d-regular graph,

$$\langle A\mathbf{1}_S, \mathbf{1}_T\rangle \leq d|S| \leq \sqrt{\eta}\sqrt{|S|\,|T|}.$$

Now consider the case $\eta\,|T| < |S| \leq |T|$. Set $S' = S \sqcup T$. Towards a contradiction, suppose that $\langle A\mathbf{1}_S, \mathbf{1}_T\rangle > 2\sqrt{\eta}d\sqrt{|S|\,|T|}$. In turn, this assumption implies that

$$\langle A\mathbf{1}_S, \mathbf{1}_T\rangle > 2\sqrt{\eta}d\sqrt{|S|\,|T|} \geq 2\eta d\,|T| \geq \eta d\,(|S| + |T|) = \eta d\,|S'|\,. \tag{5.2}$$

Using the above bound on the number of edges between S and T together with the SSE assumption on G, we obtain

$$\begin{aligned}(1-\eta)d\,|S'| &\leq \langle A\mathbf{1}_{S'}, \mathbf{1}_{\overline{S'}}\rangle \\ &\leq d\,|S'| - \langle A\mathbf{1}_S, \mathbf{1}_T\rangle \\ &< d\,|S'| - \eta d\,|S'| && \text{(By (5.2))} \\ &\leq (1-\eta)d\,|S'|\,,\end{aligned}$$

contradicting the (η, δ)-SSE property.

□

We now show a "sparse support" analogue of a lemma in Bilu and Linial [BL06, Lemma 3.3] bounding the quadratic form of the adjacency matrix for arbitrary sparse vectors assuming that the quadratic form is bounded for "flat" sparse indicator vectors. This sparse analogue follows by checking that their proof suitably "respects" the sparse support conditions we need.

Lemma 5.12 (Sparse Analogue of [BL06, Lemma 3.3])**.** *Let $A \in \mathbb{R}^{V\times V}$ be a real symmetric matrix with non-negative entries, ℓ_1-norm of each row at most d and diagonal entries zero. Let $\delta \in (0,1)$. If there exists $\alpha \in (0,1)$ such that for every disjoint sets $S, T \subseteq V$ with $|S \sqcup T| \leq \delta\,|V|$ we have*

$$\langle A\mathbf{1}_S, \mathbf{1}_T\rangle \leq \alpha d\sqrt{|S|\,|T|}, \tag{5.3}$$

then for every $u \in \mathbb{R}^V$ with $|\mathrm{supp}(u)| \leq \delta\,|V|$ we have

$$\langle Au, u\rangle \leq O(\alpha(\log(1/\alpha)+1))d\,\|u\|^2\,. \tag{5.4}$$

Proof. The assumption of (5.3) on disjoint sets S and T is strong enough to imply a similar bound with an additional factor of 2 when $S = T$ as follows.

Claim 5.13. *Suppose that A is a symmetric matrix with diagonal entries equal to zero. The assumption from (5.3) implies that for every $R \subseteq V$ with $|R| \leq \delta\,|V|$*

$$\langle A\mathbf{1}_R, \mathbf{1}_R\rangle \leq 2\alpha d\,|R|\;.$$

Proof. Let $r = |R|$. If $r = 1$, we have $\langle A\mathbf{1}_R, \mathbf{1}_R\rangle = 0$ since A has diagonal entries equal to zero. Now assume $r \geq 2$. On one hand, we have

$$\sum_{\substack{R'\subseteq R\\ |R'|=\lceil r/2\rceil}} \left|\left\langle A\mathbf{1}_{R'}, \mathbf{1}_{R\setminus R'}\right\rangle\right| \leq \binom{r}{\lceil r/2\rceil}\alpha d\sqrt{|R'|\,|R\setminus R'|} \leq \binom{r}{\lceil r/2\rceil}\alpha d\,|R|\,/2.$$

On the other hand, for distinct $x, y \in R$, the value $A_{x,y}$ appears $\binom{r-2}{\lceil r/2\rceil-1}$ in the LHS above. Since A has diagonal entries equal to zero, this gives

$$\binom{r-2}{\lceil r/2-1\rceil}\left|\langle A\mathbf{1}_R, \mathbf{1}_R\rangle\right| = \sum_{\substack{R'\subseteq R\\ |R'|=\lceil r/2\rceil}} \left|\left\langle A\mathbf{1}_{R'}, \mathbf{1}_{R\setminus R'}\right\rangle\right|.$$

From the two previous displays and the bound on the following binomial ratio

$$\binom{r}{\lceil r/2\rceil}\Big/\binom{r-2}{\lceil r/2\rceil-1} = \frac{r(r-1)}{\lceil r/2\rceil\lfloor r/2\rfloor}\ \leq\ 4,$$

we conclude the proof. $\square$

Arbitrary vectors can be approximated to have entries that are powers of two with the following nice properties.

Claim 5.14. *Suppose $A \in \mathbb{R}^{V\times V}$ has diagonal entries equal to zero. Let $u \in \mathbb{R}^V$ with $\|u\|_\infty \leq 1/2$. Then, there exists $u' \in \{\pm 1/2^i \mid i \in \mathbb{N}^+\}^V$ such that*

(i) $|\langle Au, u\rangle| \leq |\langle Au', u'\rangle|$,

(ii) $\|u'\| \leq 2\|u\|$,

(iii) $\mathrm{supp}(u') \subseteq \mathrm{supp}(u)$.

Proof. For every $i \in V$, we define $\eta_i \in [0, 1/4]$ such that $u_i = (1-2\eta_i)\,\mathrm{sgn}(u_i) 2^{\lceil \log(|u_i|) \rceil}$. Note that $(1-2\eta_i) \in [1/2, 1]$. We define a random vector $\mathbf{Z}' \in \mathbb{R}^V$ by setting $\mathbf{Z}'_i = 0$ if $i \notin \mathrm{supp}(u)$, and otherwise by setting

$$\mathbf{Z}'_i = \begin{cases} +\,\mathrm{sgn}(u_i) 2^{\lceil \log(|u_i|) \rceil} & \text{w.p. } 1-\eta_i, \\ -\,\mathrm{sgn}(u_i) 2^{\lceil \log(|u_i|) \rceil} & \text{w.p. } \eta_i. \end{cases}$$

Note that by construction, we have $\mathbb{E}[\mathbf{Z}'_i] = u_i$. Using linearity of expectation and the assumption that A has diagonal entries equal to zero, we have

$$\langle Au, u\rangle = \sum_{i \neq j} A_{i,j} u_i u_j = \sum_{i \neq j} A_{i,j} \mathbb{E}[\mathbf{Z}'_i]\mathbb{E}[\mathbf{Z}'_j] = \sum_{i \neq j} A_{i,j} \mathbb{E}[\mathbf{Z}'_i \mathbf{Z}'_j] = \mathbb{E}\left[\left\langle A\mathbf{Z}', \mathbf{Z}'\right\rangle\right] .$$

This implies that there is a choice of u' satisfying

$$|\langle Au, u\rangle| = \left|\mathbb{E}\left[\left\langle A\mathbf{Z}', \mathbf{Z}'\right\rangle\right]\right| \leq \mathbb{E}\left[\left|\left\langle A\mathbf{Z}', \mathbf{Z}'\right\rangle\right|\right] \leq \left|\left\langle Au', u'\right\rangle\right| .$$

To conclude note that a term-by-term inequality gives

$$\left\|u'\right\|_2^2 = \sum_i (u'_i)^2 \leq 4 \sum_i u_i^2 = 4\|u\|_2^2 ,$$

concluding the proof. □

Let $u \in \mathbb{R}^V$ be an arbitrary vector with $|\mathrm{supp}(u)| \leq \delta |V|$. We want to give an upper bound on $|\langle Au, u\rangle|$ as in (5.4). To prove this bound, we can assume $\|u\|_\infty \leq 1/2$ without loss of generality. Using Claim 5.14, we obtain $u' \in \{\pm 1/2^i \mid i \in \mathbb{N}^+\}^V$. Let $S_i \coloneqq \{j \in V \mid |u'_j| = 2^{-i}\}$. Set $t = \log(1/\alpha)$. Since the entries of A are non-negative, we have

$$\begin{aligned} \left|\left\langle Au', u'\right\rangle\right| &\leq \sum_{x,y \in V} A_{x,y} \left|u'_x\right| \left|u'_y\right| \\ &= \sum_{i,j \in \mathbb{N}^+} \frac{1}{2^{i+j}} \left\langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_j}\right\rangle \\ &= \underbrace{\sum_i \frac{1}{2^{2i}} \langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_i}\rangle}_{(a)} + \underbrace{\sum_i \sum_{i<j\leq i+t} \frac{1}{2^{i+j}} \left\langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_j}\right\rangle}_{(b)} \\ &\qquad + \underbrace{\sum_i \sum_{j > i+t} \frac{1}{2^{i+j}} \left\langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_j}\right\rangle}_{(c)} . \end{aligned}$$

Using the assumption in (5.3), by Claim 5.13 term (a) becomes

$$\sum_i \frac{1}{2^{2i}} \langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_i}\rangle \leq 4\alpha d \sum_i \frac{1}{2^{2i}} |S_i| = 4\alpha d \left\|u'\right\|_2^2 .$$

Note that $S_i \cap S_j = \emptyset$ when $i \neq j$. Using the assumption in (5.3), term (b) becomes

$$\begin{aligned}
\sum_i \sum_{i<j\le i+t} \frac{1}{2^{i+j}} \left\langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_j} \right\rangle &\le \sum_i \sum_{i<j\le i+t} \frac{1}{2^{i+j}} \alpha d \sqrt{|S_i|\,|S_j|} \\
&\le \sum_i \sum_{i<j\le i+t} \alpha d \left(\frac{1}{2^{2i}} |S_i| + \frac{1}{2^{2j}} |S_j| \right) \\
&\le 2\alpha \log(1/\alpha) d \sum_i \frac{1}{2^{2i}} |S_i| \\
&= 2\alpha \log(1/\alpha) d \left\| u' \right\|_2^2 ,
\end{aligned}$$

where the second step applies the Cauchy-Schwartz inequality.

Note that the ℓ_1 bound of d on the row and column sums of A trivially implies that $\left\langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_j} \right\rangle \le d\,|S_i|$. By the choice of t and this trivial bound, term (c) becomes

$$\begin{aligned}
\sum_i \sum_{j>i+t} \frac{1}{2^{i+j}} \left\langle A\mathbf{1}_{S_i}, \mathbf{1}_{S_j} \right\rangle &\le \sum_i \sum_{j>i+t} \frac{1}{2^{i+j}} d\,|S_i| \\
&\le \alpha d \sum_i \sum_{j>i} \frac{1}{2^{i+j}} |S_i| \\
&\le 2\alpha d \sum_i \frac{1}{2^{2i}} |S_i| = 2\alpha d \left\| u' \right\|_2^2 .
\end{aligned}$$

Putting the bounds on (a), (b) and (c) together, we obtain

$$|\langle Au, u\rangle| \le \left|\left\langle Au', u' \right\rangle\right| \le 6\alpha(\log(1/\alpha)+1)d \left\|u'\right\|_2^2 \le 12\alpha(\log(1/\alpha)+1)d \left\|u\right\|_2^2 ,$$

concluding the proof. □

As a consequence of the above lemma and Lemma 5.11, we obtain our main result of this section, namely, that SSE graphs are analytic SSE as follows.

Proposition 5.10. *If G is (η, δ)-SSE, then G is $(O(\sqrt{\eta}(\log(1/\eta)+1)), \delta)$-analytic SSE.*

Proof of Proposition 5.10. Since G is a (η, δ)-SSE, using Lemma 5.11 we have for every disjoint sets $S, T \subseteq V(G)$ with $|S \sqcup T| \le \delta\,|V|$

$$\langle A\mathbf{1}_S, \mathbf{1}_T \rangle \le \alpha d \sqrt{|S|\,|T|} ,$$

where $\alpha = 2\sqrt{\eta}$. By Lemma 5.12, this implies that G is $(O(\sqrt{\eta}\log(1/\eta)), \delta)$-analytic SSE concluding the proof. □

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

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.,

$$\frac{1}{nd}|\{(a,b) \in E : f_{(a,b)}(\ell(a)) = \ell(b)\}|.$$ [12]

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

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) $\mathsf{val}(\mathfrak{I}) \geq 1-\delta$.
(No) $\mathsf{val}(\mathfrak{I}) \leq \eta$.

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 6.6) 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

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

We prove Theorem 6.3 by establishing the following two lemmas in the next subsection.

[12] 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.

Lemma 6.4 (Completeness of UG protocol)**.** *For any unique games instance $\mathfrak{I}$, if* $\mathsf{val}(\mathfrak{I}) \geq 1-\delta$*. Then there is a proof with $k = O_{\delta,\eta}(1)$ unentangled states, each of size $O_{\delta,\eta}(\log n)$, such that Algorithm 6.6 accepts with probability at least* 0.99*.*

Lemma 6.5 (Soundness of UG protocol)**.** *For any unique games instance $\mathfrak{I}$, if* $\mathsf{val}(\mathfrak{I}) \leq \eta$*. Then for any proof with $k = O_{\delta,\eta}(1)$ unentangled states, each of size $O_{\delta,\eta}(\log n)$, such that Algorithm 6.6 accepts with probability at most* $7/8$*.*

Algorithm 6.6: $(1-\delta,\eta)$-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 that realize $\mathsf{val}(\mathfrak{I})$, 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.

6.1 Analysis

We first prove Lemma 6.4, the completeness. In particular, we show that for whichever test the protocol chooses, it accepts with probability at least 0.99 when $\mathsf{val}(\mathfrak{I}) \geq 1-\delta$.

For faithful proofs, the symmetry test passes with probability 1, and the sparsity test accepts with probability at least, by Theorem 4.9, $1-\exp(-\Theta(\varepsilon k))$. The validity test accepts with probability at least $1-\exp(-\Theta(q\nu^2 k))$ by Theorem 4.10. The way we choose our parameters guarantees that the accept probability is at least 0.99.

Finally, when the UG instance has a value of at least $1-\delta$, then there is valid labeling

$|\psi\rangle \in \mathcal{V}$, such that

$$\mathop{\mathbb{E}}_{r\in[d]} \langle \psi \,|\, \Pi_r \psi \rangle \geq 1 - \delta.$$

Analogous to our analysis in Section 5, we have

$$\mathop{\mathbb{E}}_{r\in[d]}[\langle \psi \,|\, \Pi_r \psi \rangle^2] \geq \left(\mathop{\mathbb{E}}_{r\in[d]}[\langle \psi \,|\, \Pi_r \psi \rangle] \right)^2 \geq (1-\delta)^2.$$

Therefore, each swap test in the labeling test accepts with probability at least $1/2 + (1-\delta)^2/2 \geq 1-\delta$. By Chernoff bound, with probability at least $1 - \exp(-\Theta(\lambda^2 k)) \geq 0.99$ for our choice of parameters, the labeling test accepts.

Now, we have proved the completeness. Next, we prove Lemma 6.5, the soundness, for which the following analysis on the labeling test will complete the last missing piece.

Lemma 6.7 (Labeling test)**.** *Suppose* $\mathsf{val}(\mathfrak{I}) \leq \eta$*. Given* ε*-tilted states* Ψ *such that any representative state* $|\psi\rangle$ *satisfies* $\mathrm{D}(|\psi\rangle, \mathcal{V})$ *and* ε *being 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* $\exp(-\Theta(\lambda^2 k))$*.*

Proof. For any valid labelings $|\tilde{\psi}\rangle \in \mathcal{V}$,

$$\mathsf{val}(\mathfrak{I}) \geq \mathop{\mathbb{E}}_{r\in[d]} \langle \tilde{\psi}, \Pi_r \tilde{\psi} \rangle \geq \mathop{\mathbb{E}}_{r\in[d]} [\langle \tilde{\psi}, \Pi_r \tilde{\psi} \rangle^2].$$

Therefore the probability that SwapTest accepts $|\tilde{\psi}\rangle$ is at most $1/2 + \eta/2$. Let $|\psi\rangle, |\phi\rangle$ be two representative states from Ψ. Suppose that for some $|\tilde{\psi}\rangle \in \mathcal{V}$, $\mathrm{D}(|\psi\rangle, |\tilde{\psi}\rangle) \leq D$. By Fact 2.8,

$$\mathrm{D}(|\psi\rangle \otimes |\phi\rangle, |\tilde{\psi}\rangle \otimes |\tilde{\psi}\rangle) \leq \sqrt{D^2 + (D + \sqrt{\varepsilon})^2} \leq 2(D + \sqrt{\varepsilon}).$$

It then follows by Fact 2.9 that the labeling test accepts $|\psi_1\rangle \otimes |\psi_2\rangle$ for two representative states in Ψ with probability at most $1/2 + \eta/2 + 2(D + \sqrt{\varepsilon})$. When we partition Ψ into two subsets Ψ_1 and Ψ_2, then with probability at least $1 - 2\varepsilon$ the states we pick from Ψ_1 and Ψ_2 are both representative states of Ψ. By Chernoff bound, with probability at most $\exp(-\Theta((\lambda - 2D - 2\sqrt{\varepsilon} - 3\varepsilon)^2 k)) = \exp(-\Theta(\lambda^2 k))$, the SwapTests accept more than $\theta - 3\varepsilon$ fraction within the $1 - 2\varepsilon$ good pairs. Since $2\varepsilon \leq 3\varepsilon(1 - 2\varepsilon)$ for sufficiently small ε, in total, the swap tests accept more than θ fraction of the pairs with probability at most $\exp(-\Theta(\lambda^2 k))$. $\square$

With all the above preparations, we are now ready prove the soundness lemma.

Proof of Lemma 6.5. Consider the following events.

$\mathcal{E}_1:$ Ψ and Γ are ε-tilted states;
$\mathcal{E}_2:$ $\mathrm{D}(\Psi, \mathcal{S}_{1/q}) \leq O(\varepsilon^{1/24} q^{1/3})$;
$\mathcal{E}_3:$ $\mathrm{D}(\Psi, \mathcal{V}) \leq O(\varepsilon^{1/48} q^{2/3})$.

If $\mathcal{E}_1$ is not true, then the symmetry test accepts with probability at most $\exp(-\Theta(\varepsilon^2 k)) < 0.01$ by Theorem 4.4 for $k = \Omega(1/\varepsilon^2)$. Thus the probability that the protocol accepts is at most $3/4 + 0.01 < 7/8$.

Conditioning on $\mathcal{E}_1$, if $\mathcal{E}_2$ does not hold, then the sparsity test accepts with probability at most $\exp(-\Theta(\varepsilon k)) < 0.01$ for $k = \Omega(1/\varepsilon)$ by Theorem 4.9. In total, the protocol accepts with a probability less than $7/8$.

Conditioning on $\mathcal{E}_1$ and $\mathcal{E}_2$, by Theorem 4.10, if $\mathcal{E}_3$ does not hold, then the validity test accepts with probability at most $\exp(-\Theta(q^{5/3}\varepsilon^{1/12}k)) < 0.01$. Therefore, the protocol accepts with probability less than $7/8$ again.

Finally, conditioning on $\mathcal{E}_1$ and $\mathcal{E}_3$, by Lemma 6.7, the labeling test accepts with probability at most $\exp(-\Theta(\lambda^2 k))$ if $\varepsilon^{1/48}q^{2/3} = O(\lambda)$ and $\varepsilon = O(\lambda^2)$. By our choice of parameters, the protocol accepts with probability at most $7/8$. □

6.2 Regularization—NP $\subseteq$ QMA$^+_{\log}(2)$

Due to the works [KMS17, KMS18, DKK+18b, DKK+18a], it is known that the $(1/2, \eta)$-GapUG problem is NP-hard. An optimistic reader would happily conclude that NP $\subseteq$ 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 6.8 (Regularization [Din07])**.** *For any general unique games instance $\mathfrak{I}$, there is a new unique games instance $\mathfrak{I}'$ that is polynomial time constructible such that*

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

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

The regularization process follows closely that of Dinur's treatment [Din07]. 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.[13] 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 6.8. Therefore, for the regular graph, $(1 - \frac{1}{2(d+1)}, 1 - \frac{1-\eta}{d+1})$-GapUG problem is NP-hard.

We verify the above claim. First note that in the new graph G', the number of edges blows up by a factor of $d+1$. This is because

$$|E'| = |V'|(d+1)/2 = |E|(d+1).$$

[13] 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 [HLW06].

Now for (6.1), a faithful prover will assign the label of a vertex v in G to the vertices of the form (v,e). Then the number of unsatisfied constraints is unchanged, but the fraction decreases by a factor of $d+1$.

For (6.2), let ℓ' be the labeling that the adversarial prover chooses. Let ℓ be the labeling on V induced by ℓ' such that for any $v \in G$, $\ell(v)$ is chosen to be the majority labeling of $\{(v,e) : e \sim v\}$ (break ties arbitrarily). For any $e = (u,v) \in E$ that is not satisfied by ℓ, either both $\ell'((u,e)) = \ell(u)$ and $\ell'((v,e)) = \ell(v)$, then the edge $((u,e),(v,e))$ is not satisfied. Or, one of the vertices $(u,e),(v,e)$ is not labeled by the majority label. The following lemma proves that within any cluster, the number of unsatisfied constraints is at least the number of vertices with minority labels. Therefore, the total number of unsatisfied constraints in G' with ℓ' is at least that of G with labeling ℓ.

Lemma 6.9. *Suppose the d-regular graph $G = (V,E)$ has Cheeger constant at least 2 and ℓ be some labeling $\ell : V \to \Sigma$. Let q denote the majority label on V, and let* $\mathrm{uneq}(G)$ *denote the number of edges (u,v) in G such that $\ell(v) \neq \ell(u)$. Then*

$$\mathrm{uneq}(G) \geq |\{v \in V : \ell(v) \neq q\}|.$$

Proof. The vertex set is partitioned by the labeling ℓ into, say, m subsets $V_1, V_2, \ldots, V_m$. Let $n_1 \geq n_2 \geq \ldots \geq n_m$ be the number of vertices in each subset.

If $n_1 \geq n/2$, then statement holds by the expansion property of G:

$$\mathrm{uneq}(G) \geq E(V_1, \overline{V}_1) \geq 2|\overline{V}_1|.$$

If $n_1 < n/2$, we bound the number of edges within each subset:

$$\begin{aligned}\frac{1}{2}\sum_{i\in\{1,\ldots,m\}} E(V_i,V_i) &\leq \sum_{i\in\{1,\ldots,m\}} (d|V_i| - 2|V_i|)/2 \\ &= \frac{dn}{2} - n,\end{aligned}$$

where the first inequality uses the expansion property of G as $|V_i| < n/2$. Therefore $\mathrm{uneq}(G) \geq n \geq |\overline{V}_1|$. □

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

Theorem 6.10. *With constant completeness and soundness gap,* $\mathrm{NP} \subseteq \mathrm{QMA}^+_{\log}(2)$.

One can work with various other approaches to prove the above theorem. For example, one can work with the 3COLOR problem, or work with the PCP characterization of NP. Looking ahead, to take advantage of the PCP characterization will be the approach we take to show $\mathrm{NEXP} = \mathrm{QMA}^+(2)$.

7 NEXP = QMA$^+$(2)

In this section, we scale up our previous result to NEXP = QMA$^+$(2). The direction that QMA$^+$(2) $\subseteq$ NEXP follows the same trivial argument that QMA(2) $\subseteq$ 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 $\{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_R\}$, and the arity of each constraint is exactly q. The value of $\mathfrak{C}$, denoted $\mathsf{val}(\mathfrak{C})$, 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)** $\mathsf{val}(\mathfrak{C}) = 1$ or **(No)** $\mathsf{val}(\mathfrak{C}) \le \delta$.*

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 $\mathcal{C}_j$ 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 $\mathcal{C}_j$ in time $\mathrm{poly}\log(NR)$. That includes deciding which variables are queried by $\mathcal{C}_j$, and based on the values of the relevant variables compute $\mathcal{C}_j$. 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) *$\mathrm{Adj}_C^{\text{global}\to\text{local}} : [R] \times [N] \to [q]$, such that $\mathrm{Adj}_C^{\text{global}\to\text{local}}(j, i) = \iota$ if i is ιth variable that $\mathcal{C}_j$ queries.*[14]

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

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

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

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 $\mathrm{Adj}_V^{\text{local}\to\text{global}}(i, \cdot)$, and vice versa.

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

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

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 $\mathsf{val}(\mathfrak{C}) = 1$ or $\mathsf{val}(\mathfrak{C}) < \delta$ 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 Appendix A.

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.*

In the next three subsections, we prove Theorem 7.5.

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 6.8. 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 [Lub11, Alo21])**.** *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 $\operatorname{poly}\log(n)$.

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 [Che10])**.** *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 $\mathsf{val}(\mathfrak{C}) = 1$, then $\mathsf{val}(\mathfrak{C}') = 1$. If $\mathsf{val}(\mathfrak{C}) = \delta < 1$, then the total number of unsatisfied constraints in $\mathfrak{C}'$ is at least $(1 - \delta - q\eta)R$.*

Proof. The analysis is similar to that of Theorem 6.8. If $\mathsf{val}(\mathfrak{C}) = 1$, then just assign the same label to all variables in V_i' and V_i'' based on the correct label for $\mathfrak{C}$. If $\mathsf{val}(\mathfrak{C}) < 1$, whenever some constraints $\mathcal{C}_i \in \mathfrak{C}$ is not satisfied by the majority labeling for the queried variables, then either (1) $\mathcal{C}_i$ is still not satisfied in $\mathfrak{C}'$ for the corresponding constraint or (2) at least one of the queried variables is not colored by the majority label. The difference in the current case from that of Theorem 6.8 is that all the variables from V_i'' can have arbitrary values without hurting any equality constraints. Since $|V_i''| \leq \eta n_i$ for any $i \in [N]$, in total

$$\left| \bigcup_{i \in [N]} V_i'' \right| \leq \eta \sum_{i \in [N]} n_i = \eta q R.$$

Therefore the total number of unsatisfied constraints is at least $(1 - \delta - q\eta)R$. $\square$

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 $\mathcal{C}_j$ 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 $\mathcal{C}_1, \ldots, \mathcal{C}_R$, 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 $\mathcal{C}_j$ 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 strongly uniform and 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 II on (Ψ, Φ) with target sparsity $|\Sigma|^{-q}$ and precision ε

(iv) Validity test on Ψ.

(v) Constraints test Ψ.

To remove any ambiguity, when taking the validity test, the valid set will be

$$\mathcal{V} := \left\{ \sum_{j\in[R]} |j\rangle|v_j\rangle : v_j \in \Sigma^q \right\}.$$

Since Σ is of constant size, and q is a constant, Σ^q is still of constant size.

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 $\mathcal{C}_j$ 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

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

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 $\mathcal{C}_j$, 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

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

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

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 $\mathcal{C}_{j'}$, 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.

The next claim certifies 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.*

Proof. First, consider the implementation of $\mathcal{A}$. Let $H' = H \otimes \mathbb{C}^q$, where the new register will be some working space. Take the following sequence of manipulations:

(i) Get a superposition on the last register:

$$|j\rangle|v\rangle|0\rangle|0\rangle|0\rangle \mapsto \frac{1}{\sqrt{q}} \sum_{\iota=1}^{q} |j\rangle|v\rangle|0\rangle|0\rangle|\iota\rangle.$$

(ii) From the second and the last register, compute v_ι and set the fourth register accordingly:

$$|j\rangle|v\rangle|0\rangle|0\rangle|\iota\rangle \mapsto |j\rangle|v\rangle|0\rangle|v_\iota\rangle|\iota\rangle.$$

(iii) Compute $\mathrm{Adj}_C^{\text{local}\to\text{global}}(j, \iota)$, and put it in the third register:

$$|j\rangle|v\rangle|0\rangle|v_\iota\rangle|\iota\rangle \mapsto |j\rangle|v\rangle|\mathrm{Adj}_C^{\text{local}\to\text{global}}(j, \iota)\rangle|v_\iota\rangle|\iota\rangle.$$

(iv) Set the last register to 0:

$$|j\rangle|v\rangle|i\rangle|v_\iota\rangle|\iota\rangle \mapsto |j\rangle|v\rangle|i\rangle|v_\iota\rangle|0\rangle.$$

The final step is valid because it is the inverse of the following operation

$$|j\rangle|v\rangle|i\rangle|v_\iota\rangle|0\rangle \mapsto |j\rangle|v\rangle|i\rangle|v_\iota\rangle|\mathrm{Adj}_C^{\text{global}\to\text{local}}(j, i)\rangle.$$

Since $\mathrm{Adj}_C^{\text{global}\to\text{local}}$ and $\mathrm{Adj}_C^{\text{local}\to\text{global}}$ can be computed efficiently classically due to the explicitness of $\mathfrak{C}$, the above steps are efficient.

The situation for $\mathcal{M}_k$ is similar. Consider $H' = H \otimes \mathbb{C}^R$. Do the following:

(i) Based on constraint j and variable i, and k, compute j' as in (7.2), put j' in the working space.

$$|j\rangle|v\rangle|i\rangle|v_i\rangle|0\rangle \mapsto |j\rangle|v\rangle|i\rangle|v_i\rangle|j'\rangle.$$

(ii) Set the first register to be 0.

$$|j\rangle|v\rangle|i\rangle|v_i\rangle|j'\rangle \mapsto |0\rangle|v\rangle|i\rangle|v_i\rangle|j'\rangle.$$

(iii) Swap the contents of the first and the last registers.

The second step is a valid step because it is the inverse of the first operation (acting on a different order of the registers). Since $\mathrm{Adj}_V^{\mathrm{local}\to\mathrm{global}}$, $\mathrm{Adj}_V^{\mathrm{global}\to\mathrm{local}}$ and Π_{G_i} are efficient classically due to the explicitness of our CSP system and expander graphs provided in Theorem 7.7, $\mathcal{M}_k$ is efficient.

$\mathcal{B}$ can be implemented efficiently because each constraint can be verified in polynomial time classically. □

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

<u>Constraints test</u>

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

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

7.3 Analysis

Lemma 7.12 (constraints test)**.** *Suppose* $\mathsf{val}(\mathfrak{C}) = 1$*, then a faithful prover passes the constraints test with probability* 1*. On the other hand, if* $\mathsf{val}(\mathfrak{C}) \leq \delta$*, then on any two valid pair of states* $|\psi\rangle$ *and* $|\phi\rangle$*, the constraints test rejects with probability at least* $(1-\delta)/(2(2d+1))$*.*

Proof. Let s_1 be the fraction of $|j\rangle|v\rangle$ such that $\mathcal{C}_j(v) = 0$, and let s_2 be the fraction of unsatisfied edges coming from the expander graphs implicitly used in the consistency test, then by Claim 7.9, we have:

$$s_1 + s_2 \geq (1 - \delta - q\eta)R.$$

The consistency test partitions all pairs of the same variable in different constraints into d matching. Let $\lambda_1, \lambda_2, \ldots, \lambda_d$ denotes the fraction of inconsistency pairs in matching

$1, 2, \ldots, d$, respectively. Analogous to the previous analysis, the probability the SwapTest accepts is

$$\mathop{\mathbb{E}}_{i\in[d]}\left[\frac{1+(1-\lambda_i)^2}{2}\right] \le 1 - \frac{1}{2}\mathop{\mathbb{E}}_{i\in[d]}\lambda_i = 1 - \frac{s_2}{2dR}.$$

In the inner constraints test, 1 is observed with probability $1 - s_1/R$. Therefore, in total the reject probability is at least

$$\frac{1}{2d+1}\cdot\frac{s_1}{R} + \frac{2d}{2d+1}\cdot\frac{s_2}{2dR} \ge \frac{1-\delta-q\eta}{2d+1}.$$

By picking the suitable n_0, we make sure $q\eta < (1-\delta)/2$, thus the reject probability is at least $(1-\delta)/(4d+2)$. □

Proof of Theorem 7.5. The completeness in Theorem 7.5 is completely analoguous to the analysis in Theorem 6.3. The soundness in Theorem 7.5 is also similar to the previous analysis. If Ψ supplied by the prover is not an ε-tilted state or is far from $\mathcal{V}$, then the symmetry test, sparsity test, and validity test will catch it. Therefore, we can assume that essentially all states in Ψ are close to some state $|\psi\rangle \in \mathcal{V}$. By choosing ε small enough and the size of Ψ sufficiently large, by Lemma 7.12 and Chernoff bound, the fraction of accepted states in the constraints test will be less than θ with high probability in the constraints test. □

8 QMA(2) v.s. QMA$^+$(2)

Our discussion so far focuses on the somewhat "artificial" complexity class QMA$^+$(2). In section, we aim to demystify this complexity class. In particular, show the relationship between QMA(2) and QMA$^+$(2), and provide a concrete reason why the QMA$^+$(2) = NEXP is an interesting result and what it says about the long-standing open problem QMA(2) v.s. NEXP.

8.1 Simulations between QMA(2) and QMA$^+$(2)

We start by showing that QMA(2) and QMA$^+$(2) can actually simulate each other to some extent. In particular, we will prove the following two lemmas.

Lemma 8.1. *For any constant* $c, s \in [0,1]$,

$$\mathrm{QMA}^+(2, c, s) \subseteq \mathrm{QMA}(2, c, 4s).$$

Lemma 8.2. *For any constant* $c, s \in [0,1]$ *with* $c - s > 1/\mathrm{poly}(n)$,

$$\mathrm{QMA}(2, c, s) \subseteq \mathrm{QMA}^+(2, 1 - e^{-O(\mathrm{poly}(n))}, e^{-O(\mathrm{poly}(n))}).$$

For the purpose of settling the problem "QMA(2) v.s. NEXP", Lemma 8.1 is probably more interesting. In fact, an immediate consequence of Lemma 8.1 is the following.

Corollary 8.3. QMA$^+(2, c, s) \subseteq$ QMA(2) *for any* $c - s \ge 3/4 + 1/\mathrm{poly}(n)$. *In particular, suppose* NEXP $\subseteq$ QMA$^+$(2) *with a gap at least* $3/4 + 1/\mathrm{poly}(n)$, *then*

$$\mathrm{QMA}(2) = \mathrm{NEXP}.$$

Proof. Suppose a given $\mathrm{QMA}^+(2,c,s)$ protocol has completeness c and soundness s. By our assumption,

$$c \geq s + 3/4 + 1/\mathrm{poly}(n), \tag{8.1}$$
$$s \leq 1/4 - 1/\mathrm{poly}(n). \tag{8.2}$$

By Lemma 8.1, there is a QMA(2) protocol that solves the same problem with completeness $c' = c$ and soundness $s' = 4s$. Then

$$c' - s' = c - 4s \geq 3/4 + 1/\mathrm{poly}(n) - 3s \geq 1/\mathrm{poly}(n), \tag{8.3}$$

where the first and second inequalities use (8.1) and (8.2), respectively. Since QMA(2) admits strong gap amplification [HM13], it shows that $\mathrm{QMA}^+(2,c,s) \subseteq \mathrm{QMA}(2)$. This finishes the proof. □

Next, we show a straightforward proof of Lemma 8.1. The fact that such proof works may be somewhat surprising.

Proof of Lemma 8.1. Given any $\mathrm{QMA}^+(2)$ protocol $\mathcal{P}$ with completeness c and soundness s. We simply reuse the verfication procedure of $\mathcal{P}$. In the completeness case, the benign provers will supply proofs with nonnegative amplitudes. Thus the completeness remains the same.

In the soundness case. On input x, let M be the operator representing the verification measurement of $\mathcal{P}$ for accepting (i.e. $I - M$ correponds to rejecting). Let $|\psi\rangle$ be an arbitrary quantum state, we compute the probability that $\mathcal{P}$ accept, $\langle\psi|M|\psi\rangle$. First, consider a decomposition of $|\psi\rangle$ as follows:

$$|\psi\rangle = \sqrt{\alpha_1}|\psi_1\rangle + \sqrt{\alpha_2}|\psi_2\rangle + \sqrt{\alpha_3}|\psi_3\rangle + \sqrt{\alpha_4}|\psi_4\rangle,$$

where $|\psi_1\rangle$ is the normalized vector consisting exactly the components of positive real amplitudes of $|\psi\rangle$; $|\psi_2\rangle$ is the normalized vector consisting exactly the components of positive imaginary amplitudes of $|\psi\rangle$; and so on. For example, if $|\psi\rangle = 1/\sqrt{3}(|1\rangle + |2\rangle + i|3\rangle)$, then $|\psi_1\rangle = 1/\sqrt{2}(|1\rangle + |2\rangle)$ and $|\psi_2\rangle = i|3\rangle$. Calculating

$$\begin{aligned}\langle\psi|M|\psi\rangle &= \sum_{i,j} \sqrt{\alpha_i\alpha_j}\langle\psi_i|M|\psi_j\rangle \\ &\leq \sum_{i,j} \sqrt{\alpha_i\alpha_j}s = \left(\sum_i \sqrt{\alpha_i}\right)^2 s \\ &\leq 4s,\end{aligned}$$

where the final step is due to Cauchy-Schwarz and that $\sum \alpha_i = 1$. The reason why the second inequality holds is as follows: In the case $i = j$, $|\psi_i\rangle$ is a quantum state without relative phase, equivalent to a quantum state with nonnegative amplitudes. Therefore $\langle\psi_i|M|\psi_i\rangle \leq s$ by the soundness of $\mathcal{P}$. In the case when $i \neq j$, then because M is PSD,

$$\begin{aligned}&(\langle\psi_i| - \langle\psi_j|)M(|\psi_i\rangle - |\psi_j\rangle) \geq 0 \\ &\quad\Rightarrow \langle\psi_i|M|\psi_j\rangle + \langle\psi_j|M|\psi_i\rangle \leq \langle\psi_i|M|\psi_i\rangle + \langle\psi_j|M|\psi_j\rangle \leq 2s.\end{aligned}$$
□

Finally, we settle Lemma 8.2, which is much more intuitive as a general quantum state can always be re-encoded into a quantum state with nonnegative amplitudes.

Proof of Lemma 8.2. Given some QMA$(2, c, s)$ protocol, which decides the measurement $\{M_x\}$. Since strong gap amplification for QMA(2) is known [HM13], assume, without loss of generality, that $c > 3/4$ and $s < 1/2$. The QMA$^+(2)$ protocol will reuse the same measurement $\{M_x\}$. On input x, suppose that a faithful prover for QMA(2) sends

$$|\psi_x\rangle = \sum_r (\alpha_r + \beta_r \cdot i)|r\rangle.$$

Now the correponding QMA$^+(2)$ prover will be asked to send the following:

$$|\phi_x\rangle = \sum |\alpha_r||r\rangle|\operatorname{sgn}\alpha_i\rangle|\mathbb{R}\rangle + |\beta_r||r\rangle|\operatorname{sgn}\beta_i\rangle|\mathbb{C}\rangle.$$

Arthur, on receiving a proof $|\phi\rangle$, applies the following transforms to the second and third registers, respectively:

$$|+\rangle \mapsto \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \qquad |-\rangle \mapsto \frac{-|0\rangle + |1\rangle}{\sqrt{2}}.$$
$$|\mathbb{R}\rangle \mapsto \frac{|0\rangle + i \cdot |1\rangle}{\sqrt{2}}, \qquad |\mathbb{C}\rangle \mapsto \frac{i \cdot |0\rangle - |1\rangle}{\sqrt{2}}.$$

Arthur then measures the second and third registers. If after the measurement, 00 is observed, apply M_x to $|\phi\rangle$. Otherwise, Arthur can output a random value.

In the completeness case, the 00 is observed with probability 1/4 and in this case the correct answer is outputted with probability c by the completeness of the QMA(2) protocol. In total, Arthur outputs a correct value with a probability of at least

$$\frac{c}{4} + \frac{3}{4}\frac{1}{2} > \frac{9}{16}.$$

In the soundness case, if 00 is observed, then Arthur is correct with probability at most s by the soundness of the QMA(2) protocol. Otherwise, Arthur is correct with probability 1/2. In any case, Arthur is correct with probability at most 1/2. In conclusion,

$$\text{QMA}(2, c, s) \subseteq \text{QMA}^+\left(2, \frac{1}{2} + \frac{1}{16}, \frac{1}{2}\right).$$

Note that in the above argument, the measurement used in QMA$^+(2)$ is adapted from QMA(2), whose soundness makes no assumption of the proofs that Arthur receives. In particular, the gap amplification for general QMA(2) can be applied. □

Note that similar arguments for Lemmas 8.1 and 8.2 can also be adapted to QMA or QMA(k).

8.2 Product Test and Gap Amplification for QMA(2)

The previous discussion begs for a better understanding of gap amplification for QMA$^+(2)$. Since many randomized/quantum computation models admit strong gap amplification, e.g.

QMA and QMA(2). This gives hope that $\mathrm{QMA}^+(2)$ may also admit strong gap amplification. The story, however, will be much more complicated in view of the recent work of Bassirian, Fefferman and Marwaha [BFM23], building on the STOC version of this paper [JW23], which curiously showed that $\mathrm{QMA}^+(1) = \mathrm{NEXP}$ also with a constant gap.[15] Since in the large constant gap regime of $\mathrm{QMA}^+(1)$, we have $\mathrm{QMA}^+(1) = \mathrm{QMA} \subseteq \mathrm{PP}$ for the same reason as discussed in Section 8.1, their result rules out the strong gap amplification for $\mathrm{QMA}^+(1)$ unless $\mathrm{NEXP} \subseteq \mathrm{PP}$. Moreover, it also suggests that techniques to amplify the gap for $\mathrm{QMA}^+(2)$ should crucially use the unentanglement assumption.

This subsection is focused on product test and its applications to gap amplification for QMA(2), which can actually be applied to $\mathrm{QMA}^+(2)$ as to be studied in the next part. In this subsection, though, our discussion will be based on Harrow and Montanaro's work [HM13] with a bit more careful treatment of the completeness and soundness gap. We include this discussion because it's central to sections to come. Experts should feel free to skip.

Given some Hilbert space $H = H_1 \otimes H_2 \otimes \ldots \otimes H_k$, each H_i is of dimension d_i. A fundamental task is to test whether a given state $|\psi\rangle \in H$ is unentangled between the k subsystems. In another word, the task is to test whether $|\psi\rangle$ is a product state, i.e., for some $|\psi_i\rangle \in H_i$, $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_k\rangle$. The following test has been proposed for this purpose when there are many copies of $|\psi\rangle$.

Algorithm 8.4: Prodcut Test

Input: $|\psi\rangle, |\phi\rangle \in H_1 \otimes H_2 \otimes \cdots \otimes H_k$.[16]

For $i = 1$ to k: Apply Swap Test to the ith subsystem of $|\psi\rangle$ and $|\phi\rangle$.

Accept if all the swap test accepts.

Let $\mathrm{P}(|\psi\rangle, |\phi\rangle)$ be the probability that $|\psi\rangle$ and $|\phi\rangle$ pass the product test. Harrow and Montanaro first gave a formal analysis of the product test, and recently Soleimanifar and Wright improved the analysis.

Lemma 8.5 (Harrow, Montanaro [HM13])**.**

$$\mathrm{P}(|\psi\rangle, |\phi\rangle) \leq \frac{1}{2}(\mathrm{P}(|\psi\rangle, |\psi\rangle) + \mathrm{P}(|\phi\rangle, |\phi\rangle)).$$

Theorem 8.6 (Soleimanifar, Wright [SW22])**.** *For any state $|\psi\rangle \in H_1 \otimes H_2 \otimes \cdots \otimes H_k$, let*

$$\max_{\phi_1 \in H_1, \phi_2 \in H_2, \ldots, \phi_k \in H_k} |\langle \psi \mid \phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_k \rangle|^2 = \omega.$$

Then, the following are two upper bounds on the probability that the product test passes. The first only works for ω at least $1/2$ (and is optimal in that regime); the second is general:

$$\begin{aligned} \mathrm{P}(|\psi\rangle, |\psi\rangle) &\leq 1 - \omega + \omega^2, && \omega \geq \frac{1}{2}; \\ \mathrm{P}(|\psi\rangle, |\psi\rangle) &\leq \frac{1}{3}\omega^2 + \frac{2}{3}, && 0 \leq \omega \leq 1. \end{aligned}$$

With the product test, one can do some gap amplification for QMA(2).

[15] It is not clear that their gap can be made as large as the one for $\mathrm{QMA}^+(2) = \mathrm{NEXP}$.

[16] Ideally, the verifier should receive two copies of $|\psi\rangle$. However, in a QMA(2) protocol, the verifier can receive anything.

Lemma 8.7 (Gap amplification for QMA(2), c.f. [HM13])**.** *For any $c - s > 1/\mathrm{poly}(n)$,*

$$\mathrm{QMA}(2, c, s) \subseteq \mathrm{QMA}\left(2,\ 1 - e^{-\mathrm{poly}(n)},\ 7/9 + e^{-\mathrm{poly}(n)}\right).$$

Proof. Repeat the protocol by asking for more provers. Based on the standard Chernoff argument, for any $c - s > 1/\mathrm{poly}(n)$, there is $k \in \mathrm{poly}(n)$ such that

$$\mathrm{QMA}(2, c, s) \subseteq \mathrm{QMA}(k, c', s'), \tag{8.4}$$

where $c' = 1 - e^{-\mathrm{poly}(n)}$ and $s' = e^{-\mathrm{poly}(n)}$.

Let the two provers simulate k provers by requiring each prover to provide the k proofs that are supposed to be sent in the QMA(k) protocol. Now with probability p, apply the product test; otherwise, apply the QMA(k) verification to the proofs provided by a random prover.

In the completeness case, the product test passes with probability 1. Therefore, the verification passes with probability still $1 - \exp(-\mathrm{poly}(n))$.

In the soundness case, let $|\psi_1\rangle, |\psi_2\rangle$ be the proofs provided by the first prover and second prover, respectively. Suppose that $|\phi_1\rangle, |\phi_2\rangle$ are the product states with the max overlap with $|\psi_1\rangle$ and $|\psi_2\rangle$, respectively,

$$|\langle\psi_1, \phi_1\rangle|^2 = \omega_1, \quad |\langle\psi_2, \phi_2\rangle|^2 = \omega_2. \tag{8.5}$$

By the soundness of the QMA(k) protocol, Fact 2.9, and (8.5), in the case that the verification is applied to a random proof $|\psi_i\rangle$ for $i = 1, 2$, it passes with probability at most $s' + \sqrt{1 - \omega_i}$. Therefore, the total probability of acceptance, i.e., the soundness of the new QMA(2) protocol, can be bounded as below

$$\begin{aligned}
&p\mathrm{P}(|\psi_1\rangle,|\psi_2\rangle) + (1-p)\left(s' + \frac{1}{2}(\sqrt{1-\omega_1} + \sqrt{1-\omega_2})\right) \\
&\qquad \leq \mathop{\mathbb{E}}_{i\in\{1,2\}}[p\mathrm{P}(|\psi_i\rangle, |\psi_i\rangle) + (1-p)\sqrt{1-\omega_i}] + (1-p)s' \\
&\qquad \leq \mathop{\mathbb{E}}_{i\in\{1,2\}}[f(p, \omega_i)] + (1-p)s',
\end{aligned} \tag{8.6}$$

where the first step uses Lemma 8.5; in the second step, by Theorem 8.6, f can be chosen as below

$$f(p, \omega_i) = p \cdot \frac{\omega_i^2 + 2}{3} + (1-p)\sqrt{1-\omega_i}.$$

Set $p = 2/3$. It can be computed that

$$\frac{\mathrm{d}}{\mathrm{d}x} f(p, x) = \frac{4x}{9} - \frac{1}{6\sqrt{1-x}},$$

and the critical points are

$$x_0 = \frac{3}{4}, \quad x_1 = \frac{1+\sqrt{13}}{8}.$$

Turns out that the global maximum of f is obtained in the boundary,

$$\max_{x\in[0,1]} f(2/3, x) = f(2/3, 0) = 7/9.$$

That finishes the proof. ∎

The above amplification can be further boosted by a "sequential" repetition. The crucial observation is that the product test is a *separable* measurement.

Definition 8.8 (Separable measurement). *A measurement $M = (M_0, M_1)$ is separable if in the yes case, the corresponding Hermitian matrix M_1 can be represented as a conical combination of two operators acting on the first and second parts, i.e., for some distribution μ over the tensor product of PSD matrices α and β on the corresponding space,*

$$M_1 = \int \alpha \otimes \beta \, \mathrm{d}\mu.$$

Fact 8.9 (Folklore). *The product test is separable.*

Proof. It suffices to show that the swap test is separable. The swap test on $\mathbb{C}^d \otimes \mathbb{C}^d$ is the projection onto the symmetric subspace, i.e., the space consisting of states that are invariant under permutations. This projection can be expressed as (up to some multiplicative constant C)

$$\mathop{\mathbb{E}}_{\theta \in \mathbb{C}^d}[|\theta, \theta\rangle\langle\theta, \theta|] = \int_{\theta \in \mathbb{C}^d} |\theta, \theta\rangle\langle\theta, \theta| \, \mathrm{d}\theta. \tag{8.7}$$

□

A detailed discussion about the fact that projection onto the symmetric subspace has the representation of (8.7) can be found in [Har13].

Definition 8.10. $\mathrm{QMA}^{\mathrm{SEP}}(2, c, s)$ *is like the standard* $\mathrm{QMA}(2, c, s)$ *with the restriction that the measurement $M = (M_0, M_1)$ used by Arthur needs to be separable between the provers.*

So Lemma 8.7 can be rephrased.

Lemma 8.11 (Gap amplification for QMA(2) rephrased, c.f. [HM13]). *For any $c - s > 1/\mathrm{poly}(n)$,*

$$\mathrm{QMA}(2, c, s) \subseteq \mathrm{QMA}^{\mathrm{SEP}}\left(2,\ 1 - e^{-\mathrm{poly}(n)},\ 7/9 + e^{-\mathrm{poly}(n)}\right).$$

Proof. In the proof of Lemma 8.7, the measurement is a linear combination of the product test and the verification procedure applied to the proof provided by either the first prover or the second prover. The product test is a separable measurement and since the verification is applied to either one part, it's also separable. □

The nice property of $\mathrm{QMA}^{\mathrm{SEP}}(2)$ is that it admits sequential repetition without going through the path that increases the number of provers and then reduces the number of provers again by the product test.

Lemma 8.12 (Gap amplification for $\mathrm{QMA}^{\mathrm{SEP}}(2)$ [HM13]). *For any $c, s \in [0, 1]$, and any $\ell = \mathrm{poly}(n)$,*

$$\mathrm{QMA}^{\mathrm{SEP}}(2, c, s) \subseteq \mathrm{QMA}^{\mathrm{SEP}}(2, c^\ell, s^\ell).$$

Proof. Do sequential repetition without increasing the number of provers, i.e., each prover will provide many copies of the original proof. Apply the measurement M from the $\mathrm{QMA}^{\mathrm{SEP}}(2, c, s)$ to each copy. Accept only if all the measurements accept.

In the completeness case, it's clear that each copy is accepted independently with a probability at least c. Therefore, with probability c^ℓ all measurements accept.

In the soundness case, use the fact that the measurement M is separable. The reduced state after applying M by tracing out the registers being measured, pure or mixed, is separable between the two provers given that the M accepts. Therefore, by the soundness, each measurement is accepted with probability at most s, and the probability that all measurements accept is at most s^ℓ. $\square$

Combining all the above pieces, Harrow and Montanaro obtained a stronger gap amplification for QMA(2).

Theorem 8.13 (Sequential repetition for QMA(2) [HM13]). *For any* $c - s > \mathrm{poly}(n)$*,*

$$\mathrm{QMA}(2, c, s) \subseteq \mathrm{QMA}(2,\ 1 - e^{-\mathrm{poly}(n)},\ e^{-\mathrm{poly}(n)}).$$

Proof. Apply Lemma 8.11 and then Lemma 8.12 with suitable parameters. $\square$

8.3 A Mild Gap Amplification for QMA$^+(2)$

In the previous sections, we discussed the connection between $\mathrm{QMA}^+(2)$ and QMA(2). The point is that, suppose we can do a strong gap amplification for $\mathrm{QMA}^+(2)$, then it would imply that

$$\mathrm{NEXP} \subseteq \mathrm{QMA}^+(2, c, s) \overset{?}{\subseteq} \mathrm{QMA}^+(2, 0.01, 0.99) \subseteq \mathrm{QMA}(2)!$$

This motivates us to review the gap amplification for QMA(2).

In this section, we discuss a little bit about the gap amplification for $\mathrm{QMA}^+(2)$. Although we don't know how to do gap amplification to achieve arbitrary completeness soundness gap (c-s gap). But as already somewhat alluded, gap amplification for $\mathrm{QMA}^+(2)$ in several regimes do exist. In particular, whenever the c-s gap is smaller than $2/9$, it can be amplified to $2/9$; and whenever the c-s gap is larger than $3/4 + 1/\mathrm{poly}(n)$, then it can be further amplified to $1 - \exp(-\mathrm{poly}(n))$.

Lemma 8.14 (Gap amplification for $\mathrm{QMA}^+(2)$ in the small gap regime). *For any* $c - s \geq 1/\mathrm{poly}(n)$*,*

$$\mathrm{QMA}^+(2, c, s) \subseteq \mathrm{QMA}^+\left(2,\ 1 - e^{-\mathrm{poly}(n)},\ 7/9 + e^{-\mathrm{poly}(n)}\right).$$

Proof. This is very much analogous to Lemma 8.7. Just note that for any state $|\psi\rangle$, the closest product state to $|\psi\rangle$ also has nonnegative amplitudes. Because, for any state $|\phi\rangle = \sum \alpha_i |i\rangle$, consider $|\phi'\rangle = \sum |\alpha_i||i\rangle$. Then $|\langle \phi' \mid \psi\rangle|^2 \geq |\langle \phi \mid \psi\rangle|^2$. $\square$

Gap amplification also works when $c - s > 3/4 + 1/\mathrm{poly}(n)$.

Lemma 8.15 (Gap amplification for $\mathrm{QMA}^+(2)$ in the large gap regime). *For any* $c - s \geq 7/8$*,*

$$\mathrm{QMA}^+(2, c, s) \subseteq \mathrm{QMA}^+(2,\ 1 - e^{-\mathrm{poly}(n)},\ e^{-\mathrm{poly}(n)}).$$

Proof. Apply Corollary 8.3 and Lemma 8.2,

$$\begin{aligned}\mathrm{QMA}^+(2,c,s) &\subseteq \mathrm{QMA}\left(2,c,s+3/4+1/\mathrm{poly}(n)\right)\\ &\subseteq \mathrm{QMA}^+(2,\ 1-e^{-\mathrm{poly}(n)},\ e^{-\mathrm{poly}(n)}).\end{aligned}$$

□

In a follow-up work, we developed some tools so that we get stronger gap amplification than that achieved here based on "sequential repetition" and product test. In particular, we showed that $\mathrm{QMA}^+(3) = \mathrm{NEXP}$ with an almost optimal completeness soundness gap of $1/2 - 1/\mathrm{poly}(n)$. In particular, if this gap can be improved slightly to $1/2 + 1/\mathrm{poly}(n)$, then $\mathrm{QMA}^{\mathbb{R}}(3) = \mathrm{NEXP}$.[17] In this case if $\mathrm{QMA}^{\mathbb{R}}(3) \neq \mathrm{NEXP}$. This shows a sharp phase transition within $\mathrm{QMA}^{\mathbb{R}}(3)$.

The key tool developed there is a disentangler type channel which is of independent interest and we expect to see some more applications. In particular, we prove that: Let $d, \ell, k \in \mathbb{N}^+$. There is an efficient channel $\Lambda\colon \mathcal{D}(\mathbb{C}^d)^{\otimes \ell} \otimes \mathcal{D}(\mathbb{C}^d)^{\otimes \ell} \to \mathcal{D}(\mathbb{C}^d)^{\otimes k}$ such that for every input states $\rho_1, \rho_2 \in \mathcal{D}(\mathbb{C}^d)^{\otimes \ell}$ there is $\int |\psi\rangle\langle\psi|^{\otimes k} d\mu$ satisfying

$$\left\|\Lambda(\rho_1 \otimes \rho_2) - \int |\psi\rangle\langle\psi|^{\otimes k} d\mu\right\|_1 \ \leq\ \left(\frac{2k^{\Theta(1)}}{\ell}\right)^{\Theta(1)},$$

and product states of the form $\rho_1 = \rho_2 = |\psi\rangle\langle\psi|^{\otimes \ell}$ are mapped to $|\psi\rangle\langle\psi|^{\otimes k}$.

This should be compared with the quantum de Finetti theorem [KR05], whose error have polynomial dependence on the dimension. This polynomial dependence is also conjectured to be inherent in a more general setting [ABD+08]. In our tool, we completely removed the dependencel on local dimension! The cost is that we require the input to be bi-partite unentangled states.

Acknowledgement

We thank Vijay Bhattiprolu for the discussions during the initial stages of this project. We thank STOC reviewers, Scott Aaronson, Penghui Yao, and Avi Wigderson for their valuable feedback on our early draft.

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach.* Cambridge University Press, 2009. 1

[ABD+08] Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. The power of unentanglement. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 223–236, 2008. 2, 14, 51

[17] $\mathrm{QMA}^{\mathbb{R}}(3)$ is the kinda of quantum proofs over reals. Quantum proofs over reals is much more natural especially we know that in many models quantum computation over reals is equivalent to over complex numbers. Furthermore, even if in the case of multi-prover systems as $\mathrm{QMA}(k)$, the fields of reals or complex numbers can change things dramatically. The hardness implication of $\mathrm{QMA}^{\mathbb{R}}(3) = \mathrm{NEXP}$ is still a substantial result for optimization problems.

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. Preliminary version in *Proc. of FOCS'92*. 55, 60

[Alo21] Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, February 2021. 38

[BB14] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science*, 2014. 1

[BBC+93] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70, Mar 1993. 1

[BBH+12] Boaz Barak, Fernando G.S.L. Brandão, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th ACM Symposium on Theory of Computing*, 2012. 2, 24

[BCY11] Fernando G. S. L. Brandão, Matthias Christandl, and Jon Yard. Faithful squashed entanglement. *Communications in Mathematical Physics*, 2011. 1

[Bei10] Salman Beigi. NP vs QMAlog(2). *Quantum Info. Comput.*, 2010. 2

[Bel64] J. S. Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1, Nov 1964. 1

[BFM23] Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. Quantum merlin-arthur and proofs without relative phase, 2023. 47

[BH13] Fernando G.S.L. Brandão and Aram W. Harrow. Quantum de finetti theorems under local measurements with applications. In *Proceedings of the 45th ACM Symposium on Theory of Computing*, 2013. 1

[BH15] Fernando G. S. L. Brandao and Aram W. Harrow. Estimating operator norms using covering nets, 2015. 1

[BKS17] Boaz Barak, Pravesh K. Kothari, and David Steurer. Quantum entanglement, sum of squares, and the log rank conjecture. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, pages 975–988. ACM, 2017. 1

[BKS19] Boaz Barak, Pravesh Kothari, and David Steurer. Small-set expansion in short-code graph and the 2-to-2 conjecture. In *ITCS 2019*, 2019. 3

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, October 2006. 11, 28, 29

[BT09] Hugue Blier and Alain Tapp. All languages in NP have very short quantum proofs. In *2009 Third International Conference on Quantum, Nano and Micro Technologies*, pages 34–37, 2009. 1

[CF13] Alessandro Chiesa and Michael A. Forbes. Improved soundness for QMA with multiple provers. *Chic. J. Theor. Comput. Sci.*, 2013. 2

[Che10] Yuan-You Fu-Rui Cheng. Explicit Estimate on Primes Between Consecutive Cubes. *Rocky Mountain Journal of Mathematics*, 40(1):117 – 153, 2010. 39

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23, Oct 1969. 1

[Din07] Irit Dinur. The pcp theorem by gap amplification. *J. ACM*, 54(3):12–es, jun 2007. 35

[DKK+18a] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in grassmann graphs. In *Proceedings of the 50th ACM Symposium on Theory of Computing*, 2018. 3, 35

[DKK+18b] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In *Proceedings of the 50th ACM Symposium on Theory of Computing*, STOC 2018, page 376–389, New York, NY, USA, 2018. Association for Computing Machinery. 35

[DPS04] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69, 2004. 1

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47, May 1935. 1

[GNN12] François Le Gall, Shota Nakagawa, and Harumichi Nishimura. On QMA protocols with two short quantum proofs. *Quantum Info. Comput.*, 2012. 2

[Har04] Prahladh Harsha. *Robust PCPs of proximity and shorter PCPs.* PhD thesis, Massachusetts Institute of Technology, 2004. 55, 57

[Har13] Aram W Harrow. The church of the symmetric subspace. *arXiv preprint arXiv:1308.6595*, 2013. 5, 49

[HHHH09] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009. 1

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, August 2006. 35

[HM13] Aram W. Harrow and Ashley Montanaro. Testing product states, quantum merlin-arthur games and tensor optimization. *J. ACM*, 60(1), feb 2013. 1, 2, 6, 9, 45, 46, 47, 48, 49, 50

[HNW17] Aram W. Harrow, Anand Natarajan, and Xiaodi Wu. An improved semidefinite programming hierarchy for testing entanglement. *Communications in Mathematical Physics*, 2017. 1

[JJUW11] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. Qip = pspace. *J. ACM*, dec 2011. 1

[JNV+20] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP*=RE, 2020. 1

[JW23] Fernando Granha Jeronimo and Pei Wu. The Power of Unentangled Quantum Proofs with Non-negative Amplitudes. In *Proceedings of the 55th ACM Symposium on Theory of Computing*, 2023. 47

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002. 2

[Kho10] Subhash Khot. Inapproximability of np-complete problems, discrete Fourier analysis, and geometry. In *Proceedings of the International Congress of Mathematicians*, 2010. 2

[KKMO04] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004. 2

[KM09] Robert König and Graeme Mitchison. A most compendious and facile quantum de finetti theorem. *Journal of Mathematical Physics*, 50(1), 2009. 1

[KMS17] Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and grassmann graphs. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, page 576–589, New York, NY, USA, 2017. Association for Computing Machinery. 35

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *Proceedings of the 59th IEEE Symposium on Foundations of Computer Science*, 2018. 3, 35

[KMY03] Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. Quantum merlin-arthur proof systems: Are multiple merlins more helpful to arthur? In *Algorithms and Computation*, 2003. 1

[KO09] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for max-cut-gain. *Theory of Computing*, 5(4):83–117, 2009. 2

[KR03] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2-\varepsilon$. In *Proceedings of the 18th IEEE Conference on Computational Complexity*, 2003. 2

[KR05] Robert König and Renato Renner. A de Finetti representation for finite symmetric quantum states. *Journal of Mathematical Physics*, 46(12), 2005. 51

[Lub11] Alexander Lubotzky. Finite simple groups of lie type as expanders. *Journal of the European Mathematical Society*, 013(5):1331–1341, 2011. 38

[MW05] Chris Marriott and John Watrous. Quantum arthur–merlin games. *Computational Complexity*, 2005. 1

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. 1

[Oru19] Roman Orus. Tensor networks for complex quantum systems. *Nature Reviews Physics*, 2019. 1

[PY86] Christos H. Papadimitriou and Mihalis Yannakakis. A note on succinct representations of graphs. *Information and Control*, 71(3):181–185, 1986. 56

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 245–254, 2008. 2

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, 2010. 2, 24

[RST10] Prasad Raghavendra, David Steurer, and Prasad Tetali. Approximations for the isoperimetric and spectral profile of graphs and related parameters. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, 2010. 3

[SW12] Yaoyun Shi and Xiaodi Wu. Epsilon-net method for optimizations over separable states. In *Proceedings of the 39th International Colloquium on Automata, Languages and Programming*, 2012. 1

[SW22] Mehdi Soleimanifar and John Wright. Testing matrix product states. In *Proceedings of the 33rd ACM-SIAM Symposium on Discrete Algorithms*, pages 1679–1701, 2022. 47

[Vid03] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91, Oct 2003. 1

[VW16] Thomas Vidick and John Watrous. Quantum proofs. *Foundations and Trends® in Theoretical Computer Science*, 2016. 1

[Wat00] John Watrous. Succinct quantum proofs for properties of finite groups. In *FOCS*, pages 537–546. IEEE Computer Society, 2000. 1

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. 1

A Doubly Explicit PCP for NEXP

In this section, we describe a PCP for NEXP, which is doubly explicit and satisfies the strong uniformity property. This will imply the following theorem immediately.

Theorem A.1 (Doubly explicit PCP for NEXP)**.** *There is some absolute constant $\kappa < 1$ and natural number q, such that it is* NEXP*-hard to decide $(1, \kappa)$-GapCSP for $(N = 2^{\mathrm{poly}(n)}, R = 2^{\mathrm{poly}(n)}, q, \{0, 1\})$-CSP systems that are doubly explicit.*

The above PCP is obtained by PCP composition. The outer PCP follows closely that of [Har04, Chapter 5], the inner PCP (of proximity) is the Hadamard code based PCP [ALM+98]. Our focus is the double explicitness, therefore the analysis on correctness will be omitted. The interested readers are referred to Harsha's thesis [Har04].

A.1 A NEXP-Complete Problem—Succinct SAT

The starting point is a NEXP-complete problem—the succinct SAT problem [PY86]. A succinct SAT instance is an encoding of some circuit $M : \{0,1\}^{3n} \times \{0,1\}^3 \to \{0,1\}$, $\mathbf{succSAT}(M) = 1$ if and only if

$$\exists x \in \{0,1\}^{2^n}, \forall (i_1, i_2, i_3, \sigma) \in \{0,1\}^{3n} \times \{0,1\}^3, \text{ s.t.}$$
$$\neg M(i_1, i_2, i_3, \sigma) \vee \left(\bigvee_{j=1}^{3} (\sigma_j \oplus x_{i_j}) \right).$$

$M(i_1, i_2, i_3, \sigma)$ determines whether there is a clause consists of variables $x_{i_1}, x_{i_2}, x_{i_3}$, and σ indicates in the clause whether the variable is negated. For example, $\sigma_1 = 1$ would indicate the corresponding literal being $\neg x_{i_1}$, while $\sigma_1 = 0$ would indicate the literal being x_{i_1}. The size of the circuit M should be at most $\mathrm{poly}(n)$.

Integrating Cook-Levin's reduction, one can conclude that there is a polynomial-size 3-CNF formula $\Phi : \{0,1\}^{3n} \times \{0,1\}^3 \times \{0,1\}^t$ for $t = n^{O(1)}$ constructing from M in polynomial time, such that

$$\begin{aligned} &\mathbf{succSAT}(M) = 1 \\ \iff &\exists x \in \{0,1\}^{2^n}, \forall (i_1, i_2, i_3, \sigma, w) \in \{0,1\}^{3n} \times \{0,1\}^3 \times \{0,1\}^t, \text{ s.t.} \\ &\qquad \neg \Phi(i_1, i_2, i_3, \sigma, w) \vee \left(\bigvee_{j=1}^{3} (\sigma_j \oplus x_{i_j}) \right). \end{aligned}$$

Abbreviate $(i_1, i_2, i_3, \sigma, w)$ by $y \in \{0,1\}^{3n+3+t}$, and let $A : \{0,1\}^n \to \{0,1\}$ be the polynomial of degree at most n that encodes input x, i.e., $A(i) = x_i$ for $i \in \{0,1\}^n$. Using standard arithmetization, there is a polynomial $P : \{0,1\}^{3n+3+t+3} \to \{0,1\}$ with $\deg P = O(\mathrm{size}|\Phi|) = \mathrm{poly}(n)$, such that

$$\neg \Phi(i_1, i_2, i_3, \sigma, w) \vee \left(\bigvee_{j=1}^{3} (\sigma_j \oplus A(i_j)) \right) \iff P(y, A(i_1), A(i_2), A(i_3)) = 0.$$

Given M, this polynomial P can be evaluated on any input in polynomial time.

A.2 A Robust Outer PCP for NEXP with $\mathrm{poly}(n)$ Queries

Based on the above discussion, a prover needs to provide $A : \{0,1\}^n \to \{0,1\}$ which is supposedly a polynomial of degree at most n, representing a satisfying assignment. To assist the verifier, the prover will in reality provide the extended version of $A : \mathbb{F}^n \to \mathbb{F}$ for some large finite field $\mathbb{F}$ with $|\mathbb{F}| = \mathrm{poly}(n)$. The verifier will carry a low-degree test on A to make sure that A is close to some polynomial of degree at most n. The low-degree test is described below.

<u>Low-degree test</u>
Input: Oracle $A : \mathbb{F}^n \to \mathbb{F}$
(i) Sample a random line by sampling random $a, b \in \mathbb{F}^n$.
(ii) Query $A(at + b)$ for all $t \in \mathbb{F}$.
Accept if $A(at + b)$ is a polynomial of t with degree at most n.

Conditioning on A being close to a low-degree polynomial, $P(y, A(i_1), A(i_2), A(i_3))$ is close to a polynomial $P_0 : \mathbb{F}^{3n+3+t} \to \mathbb{F}$ of degree at most $d = O(\deg A \cdot \deg P) = \mathrm{poly}(n)$. Let $m = 3n + 3 + t$. The goal is to test if P_0 vanishes on $\{0,1\}^m$. To accomplish this goal, the prover should provide the following auxiliary polynomials

$$Q_1, Q_2, \ldots, Q_m, P_1, P_2, \ldots, P_m : \mathbb{F}^m \to \mathbb{F}$$

satisfying that for $i \in [m]$

$$P_{i-1} \equiv Z_i Q_i + P_i,$$
$$P_m \equiv 0,$$

where Z_i is a polynomial such that $Z_i(x) = 0$ if and only if $x_i \in \{0,1\}$, for example,

$$Z_i(x) = (x_i - 1)x_i.$$

These auxiliary polynomials will be bundled together in the oracle $\Pi : \mathbb{F}^m \to \mathbb{F}^{2m}$, such that for any $x \in \mathbb{F}^m$, $\Pi(x)$ is supposed to equal $(P_1(x), P_2(x), \ldots, P_m(x), Q_1(x), \ldots, Q_m(x))$. Once the prover provide the auxiliary proof P_0 and Π, the verifier will take the following test that check whether P_0 vanishes on $\{0,1\}^n$.

Zero subcube test
Input: Oracle $P_0 : \mathbb{F}^m \to \mathbb{F}, \Pi : \mathbb{F}^m \to \mathbb{F}^{2m}$
(i) Sample a random line by sampling $a, b \in \mathbb{F}^m$
(ii) Query all points in the line $L_{a,b} = \{t \in \mathbb{F} : at + b\}$ on Π and P_0.
Reject if $P_{i-1} \neq Z_i Q_i + P_i$ for any $i \in [m]$ or $P_m \neq 0$ on any point in $L_{a,b}$.
Reject if $P_i(at + b)$ is not a polynomial on t with degree at most d, $Q_i(at + b)$ is not a polynomial of degree at most $d - 2$.
Accept, otherwise.

The combined PCP will be the following

Algorithm A.2:
[Robust PCP for **succSAT**]**Input**: $A : \mathbb{F}^n \to \mathbb{F}, \Pi : \mathbb{F}^m \to \mathbb{F}^{2m}, P_0 : \mathbb{F}^m \to \mathbb{F}$
Take one of the following tests uniformly at random.
(i) Low-degree test on A.
(ii) Zero subcube test on P_0 and Π.
(iii) Consistency test: Sample a random line L by sampling random $a, b \in \mathbb{F}^m$. *Reject* if $P_0(y) \neq P(y, A(i_1), A(i_2), A(i_3))$ for any point $y \in L$.
Accept if all tests accept.

Theorem A.3 (Robust PCP [Har04, Lemma 5.4.4])**.** *For some large enough field $\mathbb{F}$ with size $\mathrm{poly}(n)$. If the succinct SAT instance M is satisfiable, then the test accepts with probability 1. Otherwise, the test satisfies the robust soundness: If* **succSAT**$(M) = 0$*, then for some constant $\delta \in (0,1]$, with probability at least δ, the test rejects; Furthermore, the variables queried have values δ/C far away from any satisfying assignment for some absolute constant C.*

We establish the uniformity and the double explicitness property for the outer PCP. The uniformity is very straightforward from the specifications of the PCP protocol.

Claim A.4 (Uniformity of the outer PCP)**.** *For any variable v in the proof $A \circ \Pi \circ P_0$, the size of the* $\mathrm{Adj}_V(v)$ *depends only on which of the following parts v lies in: A, P_0, or Π.*

To clarify, variables in the above claim have large and different alphabets. For example, a variable in A has alphabet $\mathbb{F}$, a variable in Π would have alphabet $\mathbb{F}^{2m}$. Toward the end, we will switch to the binary representation. But this is not an issue since the size of each variable is known and at most polynomially large (since the alphabet is at most exponentially large). The index of variables using a large alphabet and the index of the bit variables can be computed efficiently.

Given the randomness used in Algorithm A.2,

$$r = (r_0, a, b) \in (\{0\} \times \mathbb{F}^n \times \mathbb{F}^n) \cup (\{1, 2\} \times \mathbb{F}^m \times \mathbb{F}^m),$$

it is very efficient to compute the variables to query since only some elementary operations are required to compute the points on the line determined by a, b. Moreover, given any variable, we can also compute the randomness with which the test queries the corresponding variable. To see this, we first record a related simple fact.

Claim A.5. *Given some $n \in \mathbb{Z}$ and finite field $\mathbb{F}$ with size polynomial in n. For any $p \in \mathbb{F}^n$, let*

$$\mathcal{L}_{n,p} = \{(a, b) \in \mathbb{F}^n \times \mathbb{F}^n : at + b = p \text{ for some } t \in \mathbb{F}\}$$

be the set of lines that pass point p. There is a natural order on $\mathcal{L}_{n,p}$ (i.e., the alphabetical order), such that the following can be computed in time $\mathrm{poly}(n)$*:*

(i) *Given any $(a, b) \in \mathcal{L}_{n,p}$, output the index of (a, b) in $\mathcal{L}_{n,p}$;*
(ii) *Given any index $\iota \in [|\mathcal{L}_{n,p}|]$, output the line (a, b) with index ι in $\mathcal{L}_{n,p}$.*

Proof. (i) For $(a, b) = (0, p)$, this is the line with the first index in $\mathcal{L}_{n,p}$. For any $a \neq 0 \in \mathbb{F}^n$ and $t \in \mathbb{F}$, there is a unique $b \in \mathbb{F}^n$ such that $at + b = p$. Therefore, given (a, b), it is easy to compute the number of lines (a', b) containing p with $a' < a$. Now for all $t \in \mathbb{F}$, we can list all the b' such that $at + b' = p$. This gives us the exact index of the given pair (a, b).

(ii) Given an index $\iota \in [|\mathcal{L}_{n,p}|]$. If $\iota = 1$, we can determine that $(a, b) = (0, p)$. Otherwise, determine a by setting a to be $\lfloor (\iota - 2)/|\mathbb{F}| \rfloor + 2$ in $\mathbb{F}^n$. Then run over $t \in \mathbb{F}$, we find the correct b. □

Claim A.6 (Double explicitness of the outer PCP)**.** *For any variable v in A, or in P_0 or in Π. There is a list* $\mathrm{Adj}_V(v)$ *of randomness r with which the outer* PCP *queries the variable v. The following are computable in time polynomial in n.*

(i) *Given any $r \in \mathrm{Adj}_V(v)$, output the index ι of r in $\mathrm{Adj}_V(v)$.*
(ii) *Given any $\iota \in [|\mathrm{Adj}_V(v)|]$, output the ιth randomness in $\mathrm{Adj}_V(v)$.*

Proof. We check all the variables and tests.

Case 1: For any point $p \in \mathbb{F}^m$, consider the variable $P_0(p)$. $P_0(p)$ can be queried in the zero subcube tests and the consistency test. In either case, $P_0(p)$ is queried when the sampled line passes p. Therefore, double explicitness holds by Claim A.5.

Case 2: For any point $p \in \mathbb{F}^m$, consider the variable $\Pi(p)$, which is queried only in the zero subcube test. Again, $\Pi(p)$ is queried only when the sampled line passes p, so double explicitness follows from Claim A.5.

Case 3: For any point $p \in \mathbb{F}^n$ corresponding to the variable $A(p)$. In this case, $A(p)$ can be queried in the low-degree test and the consistency test. In the low-degree test, the situation is completely covered by Claim A.5. Furthermore, we know that there are exactly $m_0 = |\mathbb{F}|^{2n}$ possible (a, b) that queries p. So we ignore the low-degree test, this offsets the index ι in $\mathrm{Adj}_V(v)$ for $r = (r_0, a, b)$ with $r_0 \neq 0$ by m_0. So in the remainder of the proof, we handle the consistency test.

(i) Given $r = (2, a, b)$ that queries $A(p)$, we want to determine the index of r. For string $y \in \mathbb{F}^m$, focus on the coordinates I_1, I_2, I_3 that correspond to variables i_1, i_2 and i_3, respectively. Fix some arbitrary $a \in \mathbb{F}^m$, count the b that satisfies any of the following

$$(a|_{I_1}, b|_{I_1}) \in \mathcal{L}_{n,p}, \tag{1}$$
$$(a|_{I_2}, b|_{I_2}) \in \mathcal{L}_{n,p}, \tag{2}$$
$$(a|_{I_3}, b|_{I_3}) \in \mathcal{L}_{n,p}. \tag{3}$$

Recall that $\mathcal{L}_{n,p}$ is the set of the lines that pass the point p in $\mathbb{F}^n$. For $I \in \{I_1, I_2, I_3\}$, if $a|_I \neq 0$, the number of $b|_I$ in $\mathcal{L}_{n,p}$ is $|\mathbb{F}|$; if $a|_I = 0$, then there is only one $b|_I$. In any case, there are at most polynomially many different assignments to $b|_{I_1}, b|_{I_2}, b|_{I_3}$ to satisfy (1), (2) or (3) depending only on how many 0s in $a|_{I_1}, a|_{I_2}, a|_{I_3}$. The other coordinates can be set arbitrarily. Therefore, even if we don't know a exactly, but only the number of 0s in I_1, I_2, I_3, we can still compute the number of bs such that (a, b) queries $A(p)$. Let

$$C_k(a) = \left\{ a' < a : \sum_{i \in [3]} \mathbb{1}\left[a'|_{I_i} = 0\right] = k \right\}.$$

For any fixed a, note that $C_k(a)$ can be computed efficiently. Since k decides

$$|\{b' \in \mathbb{F}^m : (a', b') \text{ queries } A(p)\}|,$$

for any $a' \in C_k(a)$, we can compute the total number of (a', b') in consistency check that queries $A(p)$ for $a' < a$.

Now fix some b, such that p lies in line (a, b) restricted to I_1, I_2 or I_3. Count $b' < b$ such that (a, b') queries $A(p)$. We can do this because we can count the following efficiently

$$B_k = \{b' < b : (a|_{I_k}, b'|_{I_k}) \in \mathcal{L}_{n,p}\}, \qquad k = 1, 2, 3.$$
$$B_{jk} = \{b' < b : (a|_{I_j}, b'|_{I_j}), (a|_{I_k}, b'|_{I_k}) \in \mathcal{L}_{n,p}\}, \qquad 1 \leq j < k \leq 3.$$
$$B_{123} = \{b' < b : (a|_{I_1}, b'|_{I_1}), (a|_{I_2}, b'|_{I_2}), (a|_{I_3}, b'|_{I_3}) \in \mathcal{L}_{n,p}\}.$$

The reason is that for a fixed a, the arbitrary combination of (1), (2) and (3) restricts b' on the corresponding locations (e.g. $b'|_{I_1}$, or $b'|_{I_1 \cup I_2}, \ldots$) with at most polynomially many assignments (in particular at most $|\mathbb{F}|^3$). For each of the assignments, it is easy to count the number of assignments on the unrestricted coordinates such that $b' < b$. Finally, using the inclusion-exclusion principle, we know exactly the number of (a, b') that queries $A(p)$ for $b' < b$. This tells us the index ι of (a, b) for variable $A(p)$.

(ii) Now given the index ι, we want to determine (a, b) in the randomness. We first fix the value of a. To do so, we start by fixing the coordinates before I_1, I_2, I_3. Then we decide if a_{I_1} is 0. If not we can decide the value of a_{I_1}, and so on. After we fix the value of a, we decide the value of $b|_{I_1}$. For any assignment σ to $b|_{I_1}$, we can count the total number of assignments of b such that (a, b) queries $A(p)$. This number only depends on whether $(a|_{I_1}, b|_{I_1}) \in \mathcal{L}_{n,p}$ under the given assignment σ. Since there are only polynomially many assignments σ making $(a|_{I_1}, b|_{I_1}) \in \mathcal{L}_{n,p}$, we can decide the value of $b|_{I_1}$. Analogously, we can decide the value of $b|_{I_2}$ and $b|_{I_3}$, and finally all the other coordinates.

□

A.3 The Hadamard Inner PCP

The standard approach to reduce the number of queries in a PCP system is to compose the outer PCP with a query-efficient inner PCP. In the case of NEXP, the task is much easier. Simply note that once the randomness r is fixed, then there is a polynomial-time Turing machine M_r that verifies if the variables to query, again depending on r, satisfies the corresponding test. This verification can also be "verified" using the following well-known Hadamard code based PCP.

Theorem A.7 (cf. [ALM+98]). *For any constant $\delta > 0$, there is a* PCP *of proximity for any* NP *problem with* $\mathrm{poly}(n)$ *number of random bits, query complexity* $O(1)$*, perfect completeness, and robust soundness δ: for any input δ-far from satisfying the circuit, the test rejects with probability at least $O(\delta)$.*

For the purpose of showing the double explicitness property, we briefly go over the construction of this PCP. For any Turing machine M runs in time $\mathrm{poly}(|x|)$ on input x, whether $M(x) = 1$ can be reduced to the problem of deciding the existence of a solution to a system of polynomially many quadratic equations in $\mathbb{F}_2$.

Theorem A.8 (NP-completeness of quadratic equations). *Given any Turing machine M that runs in time $t = \mathrm{poly}(m)$ on input x of length m. There is a polynomial time reduction $\mathcal{A}$ that runs in time* $\mathrm{poly}(m)$ *on x, and outputs $A \in \{0,1\}^{\ell \times n^2}, b \in \{0,1\}^{\ell}$ for $n, \ell = \mathrm{poly}(m)$, such that*

(i) *If $M(x) = 1$, then for some $x' \in \{0,1\}^n$ that $x' \succ x$ and $A(x' \otimes x') = b$,*
(ii) *If $M(x) = 0$, for any $x' \in \{0,1\}^n$ that $x' \succ x$, $A(x' \otimes x') \neq b$.*

Furthermore, the rows of A are linearly independent.

Here $x' \succ x$ means that x is a prefix of x'.

Proof. The correctness is standard. The focus is to show that A has linearly independent rows. Start from the Cook-Levin's reduction. Consider the computational tableau $T \in \{0, 1, \dot{0}, \dot{1}, \perp\}^{t \times t}$, where $\dot{0}$ and $\dot{1}$ denote that the header is pointing to the current cell and $\perp$ denotes the empty cell. We can encode $0, 1, \dot{0}, \dot{1}, \perp$ using the binary alphabet by for example, $000, 001, 010, 011, 111$, respectively. We interpret $\{0, 1\}$ as elements in $\mathbb{F}_2$. Therefore, each symbol is encoded using three variables. By Cook-Levin's reduction, there is a 3SAT formula Ψ on variables associated with T. The way we encode the symbols guarantees that the input x to M is a substring of the input x' to Ψ. By rearranging, we can make sure x is a prefix of x'. We make Ψ to have fan-in 2 by adding intermediate gates. In particular, for every internal gate, associate a new variable. Then for every gate z that takes two variables x and y as its input (when the input, say x, is negated, simply replace x with $1 - x$), add the equation based on the operation of z, as below:

(i) If $z = x \wedge y$, add the equation: $xy + z = 0$,

(ii) If $z = x \vee y$, add the equation: $z + x + y + xy = 0$.

For the top gate z, add equation $z = 1$. For variables associated with the first row in the tableau T, add the corresponding equation to ensure things like the header is pointing to the first cell; the cells after the input x is empty; etc. These equations are only enforced on the "inputs" to the formula Ψ, and for each such variable, there is only one such equation. Note that for any internal gate z in the formula, they only show up in two equations. One that verifies the inputs variables are consistent with z. The other verifies that when z is fed into an upper gate, the values are consistent. In the first case, there is always the term z. In the second case, there is always the term zy for some other variable y.

We show that the equations introduced above result in a matrix A with linearly independent rows. Take an arbitrary equation that corresponds to some gate z in the formula, where we introduced a term z. If z is not the top gate, then to eliminate the term z we must include the equation corresponding to the gate z' that takes z as an input, which will introduce the term zy for some y. This term zy is not removable. If z is the top gate, then the equation itself already introduces a term xy for some gate x and y, which is not removable. One remaining case is when the equation is $z = 1$ or $z = 0$ for some z, an input to the formula. In this case, to remove the variable z, the only way is to look for any internal gate that takes z as an input. However, once we take variables associated with internal gates, we are back to the first case. □

Algorithm A.9: Hadamard PCP for some polynomial-time Turing machine M

Convert M into a system of quadratic equation $A : \{0,1\}^{\ell \times n^2}, b \in \{0,1\}^{\ell}$. Let $x \in \{0,1\}^m$ be the input to M.

Prover provides the proof consists of $Y \in \mathbb{F}_2^n, Z \in \mathbb{F}_2^{n^2}$ such that for some solution $x' \in \{0,1\}^n$ to $A(x' \otimes x') = b$ that extends x, i.e., $x' \succ x$, and satisfy

$$Y(y) = \langle y, x' \rangle,$$
$$Z(z) = \langle z, x' \otimes x' \rangle.$$

Verifier checks the following

(i) (Linearity test for Y) Sample random $y, y' \in \{0,1\}^n$, test if $\langle y, x' \rangle + \langle y', x' \rangle = \langle y + y', x' \rangle$.

(ii) (Linearity test for Z) Sample random $z \in \{0,1\}^{n \times n}, z' \in \{0,1\}^{n \times n}$, test if $\langle z, x' \otimes x' \rangle + \langle z', x' \otimes x' \rangle = \langle z + z', x' \otimes x' \rangle$.

(iii) (Consistency test on Y and Z) Sample $w, w' \in \{0,1\}^n$, test if $\langle w, x' \rangle \langle w', x' \rangle = \langle w \otimes w', x' \otimes x' \rangle$.

(iv) (Equation test) Sample $u \in \{0,1\}^{\ell}$, test if $\langle A^T u, x' \otimes x' \rangle = \langle A^T u, b \rangle$.

(v) (Proximity test) Sample $i \in [m]$ and $v \in \{0,1\}^n$, test if $\langle v + e_i, x' \rangle + \langle v, x' \rangle = \langle e_i, x' \rangle$.

Accept only if all tests pass.

We make a few remarks here. First, for our purpose, we don't really worry about the optimal query complexity as long as the total number of queries is a constant number. Second, note that by repeating the test multiple of times, we can detect any proximity parameter. If the above PCP is doubly explicit, it remains doubly explicit repeating constant number of times. Finally, actually the prover only provides $x'_{m+1} x'_{m+2} \cdots x'_n$, since the first m bits are part of the input.

Next, we establish the uniformity and double explicitness of the inner PCP. For uniformity, we can classify the variables in the proof of the Hadamard PCP described above into constantly different types based on:

(i) The input x to M, in another word, $Y(e_i)$ for $i \in [m]$ form one type of variables.
(ii) For $Y(a)$ for $a \notin \{e_i : i \in [m]\}$, form another type of variables.
(iii) For $Z(a)$, depending on whether $a \in \{0,1\}^n \otimes \{0,1\}^n$, and whether $\exists u \in \{0,1\}^\ell$ such that $A^T u = a$, $\{Z(a) : a \in \{0,1\}^{n^2}\}$ are decomposed into four different types of variables.

Claim A.10 (Uniformity of inner PCP)**.** *There are six types of variables for the inner* PCP *as listed above. For any two variable v_1, v_2 that belong to the same type, $|\mathrm{Adj}_V(v_1)| = |\mathrm{Adj}_V(v_2)|$. Furthermore, $|\mathrm{Adj}_V(v_1)|$ can be computed efficiently.*

Proof. By inspection. □

Claim A.11 (Double explicitness of inner PCP)**.** *Fix any variable a which can be either some $Y(y)$ for $y \in \mathbb{F}^n$, or $Z(z)$ for $z \in \mathbb{F}^{n^2}$. Let $\mathrm{Adj}_V(a)$ be the list of randomness $r = (y, y', z, z', w, w', u, i, v)$ that queries a. The following are computable in time* $\mathrm{poly}(n)$*:*

(i) *Given any $r \in \mathrm{Adj}_V(a)$, output the index ι of r in $\mathrm{Adj}_V(a)$.*
(ii) *Given an index ι, output the ιth random string r in $\mathrm{Adj}_V(a)$.*

Proof. We carefully examine all the cases. Let $\mathcal{U} = \{0,1\}^{2n+2n^2+2n+\ell} \times [m] \times \{0,1\}^n$, given any $r \in \mathcal{U}$, decompose $r = r_1 r_2 \cdots r_8 r_9$, such that $(r_1, r_2, \ldots, r_9)$ corresponds to $(y, y', z, z', w, w', u, i, v)$ in the Hadamard PCP.

Case 1: Suppose that $a \in \mathbb{F}^n$ corresponds to the variable $Y(a)$. $Y(a)$ can be queried in linearity test for Y, consistency test and proximity test. Then

$$\begin{aligned}
&\mathrm{Adj}_V(a) = E_1(a) \cup E_2(a) \cup E_{12}(a) \cup E_5(a) \cup E_6(a) \cup E'_8(a) \cup E_9(a) \cup E_{89}(a),\\
&E_i(a) := \{r \in \mathcal{U} : r_i = a\}, \qquad\qquad i \in \{1,2,5,6,9\},\\
&E_{12}(a) := \{r \in \mathcal{U} : r_1 + r_2 = a\},\\
&E'_8(a) := \{r \in \mathcal{U} : e_{r_8} = a\},\\
&E_{89}(a) := \{r \in \mathcal{U} : r_9 + e_{r_8} = a\}.
\end{aligned}$$

Now given any proper prefix p for some $r \in \mathcal{U}$ such that

$$p \in \{\varepsilon, r_1, r_1 r_2, \ldots, r_1 r_2 r_3 r_4 r_5 r_6 r_7 r_8\},$$

where ε stands for the empty string. let

$$\mathrm{Adj}_V(a)|_{p,r} := \mathrm{Adj}_V(a) \cap \{ps \in \mathcal{U} : ps < r\}.$$

We want to compute the cardinality of $\mathrm{Adj}_V(a)|_{p,r}$. Suppose the prefix p already implies a query on $Y(a)$, then the suffix s can be anything that makes $ps < r$. If the prefix p does not imply a query on $Y(a)$, we consider all the following sets.

$$\begin{aligned}
&E_i(a,p,r) := \{r' < r : r'_i = a, p \prec r'\}, \qquad\qquad i \in \{1,2,5,6,9\},\\
&E_{12}(a,p,r) := \{r' < r : r'_1 + r'_2 = a, p \prec r'\},\\
&E'_8(a,p,r) := \{r' < r : e_{r'_8} = a, p \prec r'\},\\
&E_{89}(a,p,r) := \{r' < r : r'_9 + e_{r'_8} = a, p \prec r'\}.
\end{aligned}$$

We claim that we can compute the cardinality of the intersection for an arbitrary combination of the above sets. If this is indeed the case, the cardinality of $\mathrm{Adj}_V(a)|_{p,r}$ can be computed efficiently using the inclusion-exclusion principle. First, for $r' \in E_i$, r'_i is fixed to be a. For E'_8, $|E'_8|$ is nonzero only if $a = e_i$ for some $i \in [m]$. In that case, it fixes the value of r'_8. For E_{89}, there are at most m possible ways of setting r'_8 and r'_9. When we consider the intersection of an arbitrary combination of the above sets, we are restricting the corresponding coordinates to at most m possible assignments, which we can list efficiently. For each assignment, it is easy to count the number of assignments to the unrestricted coordinates that are consistent with p and smaller than r. Finally, we take E_{12} into account. If p already fixes r'_1, then it determines r'_2. Otherwise, for every possible r'_1, there is one corresponding r'_2. When taking intersections with other sets, the corresponding coordinates I are restricted to at most m possible assignments. We can exhaust the assignments to I, and for all $r'_1 < r_1$, the unrestricted coordinates can have arbitrary values. For the single special case $r'_1 = r_1$, depending on whether $r'_2 < r_2$ or $r'_2 = r_2$ or $r'_2 > r_2$, we can also count efficiently the number of assignments to the other coordinates such that $r' < r$.

The above discussion helps us establish the double explicitness for $Y(a)$. In particular, (i) given any $r \in \mathcal{U}$, by computing the cardinality of $\mathrm{Adj}_V(a)|_{p,r}$ for $p = \varepsilon$, we can compute the index ι of r in $\mathrm{Adj}_V(a)$. (ii) Suppose we are given the index ι. For any prefix p, we can efficiently compute $|\mathrm{Adj}_V(a)|_{p,r}|$, by setting $r = 1^{2n+2n^2+2n+\ell} \circ m \circ 1^n$. The cardinality only depends on whether p already queries $Y(a)$, the length of p, and a. Therefore, we can compute the ιth randomness in $\mathrm{Adj}_V(a)$ by gradually determine $r_1, r_2, \ldots, r_9$.

Case 2: Suppose $a \in \mathbb{F}^{n^2}$ corresponds to some $Z(a)$. $Z(a)$ can be queried in linearity test for Z, consistency test and equation test. Then

$$\begin{aligned}
&\mathrm{Adj}_V(a) = E_3(a) \cup E_4(a) \cup E_{34}(a) \cup E_{56}(a) \cup E_7(a),\\
&E_i(a) := \{r \in \mathcal{U} : r_i = a\}, \qquad\qquad i \in \{3,4,7\},\\
&E_{34}(a) := \{r \in \mathcal{U} : r_3 + r_4 = a\},\\
&E_{56}(a) := \{r \in \mathcal{U} : r_5 \otimes r_6 = a\},\\
&E_7(a) := \{r \in \mathcal{U} : A^T r = a\}.
\end{aligned}$$

Analogous to case 1, we also consider the version $\mathrm{Adj}_V(a)|_{p,r}, E_i(a,p,r), E_{ij}(a,p,r)$ that are consistent with some prefix p. The cardinality of $E_i(a,p,r)$ can be computed just like in case 1. The cardinality of E_{56} is nonzero only when a is a tensor product of some w, w', and a completely determines w and w'. For E_7, we need to solve the following linear equation such that

$$A^T u = a.$$

Since the rows of A are independent, there is at most one solution for the above equation. This can be found in polynomial time using, for example, Gaussian elimination. All in all, when considering the intersection of an arbitrary combination of the above sets, we are restricting a few coordinates to at most 1 possible assignment. It is easy to count the number of assignments on the other unrestricted coordinates that are consistent with the prefix p and smaller than r. To take E_{34} into account, this is completely analogous to what happens in case 1. Therefore, we can compute $\mathrm{Adj}_V(a)|_{p,r}$ efficiently.

Now it follows the same argument as in case 1, given any $r \in \mathcal{U}$, we can compute the index ι of r in $\mathrm{Adj}_V(a)$ and given any index ι we can compute the corresponding ιth randomness r. $\square$

A.4 The PCP Composition

The final PCP for the succinct SAT problem will be the composition of the outer PCP and the inner PCP. In particular, for any succinct SAT instance M, let $s = \text{size}(M)$. The prover should provide the proof Π^{outer} for the outer PCP. The outer PCP verifier samples the randomness $r \in \{0,1\}^{\text{poly}(s)}$. Depending on r, some polynomial-time verification M_r will be invoked to verify a set of variables I_r in Π^{outer}, denoted by $\Pi^{\text{outer}}|_{I_r}$. M_r can be converted into a quadratic equation instance (A_r, b_r) in time poly(s). The prover will provide for all possible randomness r, a proof Π_r^{inner}. Now the inner PCP will verify $\Pi^{\text{outer}}|_{I_r} \circ \Pi_r^{\text{inner}}$. Sample the randomness $r' \in \{0,1\}^{\text{poly}(s)}$ for the inner PCP. Based on r', there is a polynomial-time verification $M_{r'}^{\text{inner}}$ that verifies $\Pi^{\text{outer}}|_{I_r} \circ \Pi_r^{\text{inner}}$.

The prover will arrange the proofs as a concatenation of $\Pi^{\text{outer}} \circ \Pi_0^{\text{inner}} \circ \Pi_1^{\text{inner}} \circ \cdots$. Note that there are exactly $m_0 = |\mathbb{F}|^{2n}$ random strings for the low-degree tests in the outer PCP. These tests correspond to the same verification procedure, therefore the inner PCPs have the same structure. Following the low-degree tests are the zero tests corresponding to the next $m_1 = |\mathbb{F}|^{2m}$ random strings. Finally, the remaining are $m_2 = |\mathbb{F}|^{2m}$ consistency tests. We know exactly the size of $|\Pi_r^{\text{inner}}|$ for each r. Therefore for any variable v, it can be computed efficiently whether v lies in Π^{outer} or Π_r^{inner}, and in the latter case, we can computer r in polynomial time. So when we talk about a variable v, we suppose the information is provided.

Theorem A.12. *The composed* PCP *is doubly explicit.*

Proof. Fix some variable v, there are two cases. First, if $v \notin \Pi^{\text{outer}}$. This case is straightforward: v is queried only if the random string r for the outer PCP is correct. Then the double explicitness for v follows the double explicitness of the inner PCP.

Second, if $v \in \Pi^{\text{outer}}$. Now given r, r' the random strings for the outer and inner PCPs, respectively. From the double explicitness of the outer PCP, we know the index ι of the $r \in \text{Adj}_V^{\text{outer}}(v)$. From which, we can compute the cardinality of $\mathcal{R} = \{(s, s') \in \text{Adj}_V(v) : s < r\}$. This is because by ι we know exactly the cardinality of $\mathcal{R}_i = \{s \in \text{Adj}_V^{\text{outer}}(v) : s < r\} \cap T_i$ for $i \in \{1,2,3\}$, where T_1, T_2, T_3 are the sets of random strings for the outer PCP that invoke the low-degree tests, zero tests, and consistency tests, respectively. Due to the uniformity of the inner PCP, for any $s \in \mathcal{R}_i$, the size of the adjacency list of v for the inner PCP is the same. Denote n_i be the size of adjacency list of v for any $s \in \mathcal{R}_i$, we have

$$|\mathcal{R}| = \sum_{i=1}^{3} n_i \cdot |\mathcal{R}_i|.$$

Now by the double explicitness of the inner PCP, we get the index ι' of r'. The index of (r, r') for the composed PCP is therefore $|\mathcal{R}| + \iota'$. On the other hand, let some index ι be given. Since we can compute n_i, it is easy to fix r. Then the double explicitness of the inner PCP will determine r'.

The above argument establishes the double explicitness on adjacency list Adj_V. The explicitness of Adj_C is straightforward. Given the random string r, r', fully determined by r the outer PCP queries a line in one of three tests, the points on which are efficient to list. Look inside the corresponding inner PCP, by r' we can efficiently output the corresponding locations to query. Since we can efficiently output the list of variables that (r, r') queries, it shows the explicitness on the adjacency list Adj_C. $\square$

The uniformity of the inner PCP and outer PCP together implies the uniformity of the composed PCP.

Theorem A.13. *In the composed* PCP*, there are only a constant number of different types* $[N = 2^{\mathrm{poly}(s)}] = V_1 \cup V_2 \cup \cdots \cup V_k$ *of variables in the sense that the size of* $\mathrm{Adj}_V(v)$ *only depends on which type the variable* v *is.*

Proof. First, consider any variable $v \notin \Pi^{\mathrm{outer}}$. There are only 3 different kinds of inner PCP depending on whether the outer PCP invokes the low-degree test, zero subcube test, or consistency test. For each test, by the uniformity of the inner PCP, there are 5 different types of variables. In total, there are 15 different types of variables. Consider any variable $v \in \Pi^{\mathrm{outer}}$. (Note that when we discuss the outer PCP, we are using a large alphabet. Here, a variable has a binary value. So we split one variable from the outer PCP into polynomially many.) By the uniformity of the outer PCP, which and how many of the low-degree tests, zero tests, and consistency tests are only depending on whether v belongs to A, P_0 or Π. For any v that belongs to the same type, the uniformity of the inner PCP tells us that the total number of constraints that queries v is fixed. $\square$