THE UNIVERSITY OF CHICAGO

ON THE POWER OF QUANTUM MERLIN-ARTHUR CLASSES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

BY
ROOZBEH BASSIRIANJAHROMI

CHICAGO, ILLINOIS
JUNE 2025

ABSTRACT

The complexity class NP has a natural definition that has been widely studied in various contexts in classical complexity theory. For the quantum version of NP, however, one can make several choices while defining a class of problems that can be verified in quantum polynomial time. QCMA, QMA and QMA(2) are examples of these classes – see Gharibian [2023c] for a survey. Although these three classes have been studied extensively in the past two decades, little is known about their comparative power beyond QCMA $\subseteq$ QMA $\subseteq$ QMA(2). Here we study the relation between these classes using different techniques: either through oracle separations or by studying restricted variants of these classes.

First, to study the power of classical proofs, we examine how different design choices for an oracle influence the complexity of quantum property testing problems that are defined with respect to that oracle. Specifically, we represent a regular graph with even degree as an invertible function f, and explore various oracle models for accessing f. First, we introduce a one-query QMA protocol to test whether the graph encoded by f contains a small disconnected component. Then, using representation theory, we demonstrate that a classical witness cannot assist a quantum verifier in efficiently solving this problem when using an in-place oracle. Interestingly, a minor alteration to the conventional oracle setup can prevent even a quantum verifier with unlimited witness access from solving the problem efficiently.

Second, in an attempt to understand the power of QMA(2), we explore a direction proposed by Jeronimo and Wu [2023a] – where it was shown that QMA(2) with non-negative amplitudes has the ability to solve NEXP-hard problems, and hence amplification for that class implies QMA(2) = NEXP. To give negative evidence for this direction, we explore QMA with non-negative amplitudes. When only the completeness condition is altered, this modified class remains equivalent to standard QMA. However, if both completeness and soundness are changed, the resulting class becomes significantly more powerful. We demonstrate that QMA^+ with one fixed completeness-soundness gap equals NEXP, while with a

different constant gap, it remains equivalent to QMA. This result limits the possibility of proving $\mathsf{QMA}(2) = \mathsf{NEXP}$ by amplification, since any amplification technique must fail for QMA^{+}.

Lastly, we use first these techniques to establish that QMA, when the verifier is allowed to perform a single non-collapsing measurement, is equivalent to NEXP–answering an open question posed by Aaronson. Central to several findings influenced by Blier and Tapp Blier and Tapp [2010] is a non-physical property testing problem, which involves determining whether a quantum state is close to a specific basis element. This viewpoint leads us to a variation of QMA in which a single quantum proof is strictly less powerful than two unentangled proofs—under the assumption that $\mathsf{EXP} \neq \mathsf{NEXP}$. This provides a novel path towards proving that $\mathsf{QMA}(2) = \mathsf{NEXP}$, while avoiding a key limitation of the previous proposed method. In our modification, each proof is required to exhibit a type of multipartite unentanglement: specifically, once one register is traced out, a small subset of qubits remains separable from the rest of the quantum state.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am deeply grateful to the many incredible mentors who guided me during my time in graduate school, especially my advisor, Bill Fefferman. I am also thankful to my collaborators and friends for all the knowledge and insights they shared with me. Lastly, I would like to express my heartfelt thanks to my family and close friends for their unwavering support and patience throughout this journey.

CHAPTER 1

INTRODUCTION

Computational complexity explores the fundamental resources needed to perform a given task. Tasks that require similar resource amounts are grouped into complexity classes, and a major goal in the field is to understand how these classes relate to one another. Some progress has been made in showing that certain classes are equivalent—such as $\mathsf{IP} = \mathsf{PSPACE}$ Shamir [1992], the PCP theorem Arora et al. [1998b], and $\mathsf{MIP}^* = \mathsf{RE}$ Ji et al. [2021a]. However, demonstrating that two complexity classes are different remains much more difficult; for instance, we still cannot prove $\mathsf{P} \neq \mathsf{PSPACE}$, let alone the famous $\mathsf{P} \neq \mathsf{NP}$.

One approach to addressing this challenge involves using an oracle – a black-box computational tool that can solve a specific (potentially very complex) problem in one step. Oracles often make it easier to prove results like $\mathsf{P} \neq \mathsf{NP}$ in a relativized world. Moreover, they clarify which proof techniques cannot resolve the problem in the unrelativized setting. Beyond separating complexity classes, oracles and the broader concept of query complexity also appear naturally in areas like cryptography Katz and Lindell [2020] and learning theory Kushilevitz and Mansour [1993].

Still, even when oracles are involved, proving class separations can remain extremely challenging. For example, Aharonov and Naveh introduced the class QCMA – a subset of QMA where the proof (or witness) is classical, and questioned whether QCMA is strictly contained in QMA Aharonov and Naveh [2002a]. Aaronson and Kuperberg proposed an oracle that separate these classes, but their result relied on a "quantum oracle" Aaronson and Kuperberg [2007]. Later works Fefferman and Kimmel [2018], Natarajan and Nirkhe [2022] removed the explicitly quantum aspects, but still relied on oracles with internal randomness or other non-standard features.

In Chapter 2, we study quantum property testing problems where the input is provided via oracles drawn from various models. Specifically, we encode the edges of a graph as a reversible function f, and explore how presenting f as a standard, in-place, or random-

ized oracle affects computational power. Under reasonable limits on the quantum verifier's workspace, we find:

1. In several oracle models, a quantum witness allows a quantum verifier to efficiently determine whether the graph encoded in f contains a small disconnected component.

2. If f is given as a randomized in-place oracle, no classical witness enables efficient verification by a quantum verifier.

3. When f is presented as a randomized phase oracle, no witness – regardless of its type or size – can assist a quantum verifier in solving the problem efficiently.

These results emphasize that the quantum complexity of oracle-based tasks – especially in the case of QMA and QCMA – is deeply affected by the specific oracle model used.

Next, we turn to another variant of quantum NP, $\mathsf{QMA}(2)$. Two key sources of power of quantum computation are entanglement, which gives rise to nonlocal correlations, and relative phase, which enables interference effects like destructive cancellation. Using tools from complexity theory, we examine how these properties influence computational power.

Extending the results of Jeronimo and Wu [2023b], we study the class QMA^{+}, where quantum witnesses are constrained to have no relative phase. Surprisingly, we find that this restriction alone does not weaken the class, in fact, there exist constants $1 > c > s > 0$ such that $\mathsf{QMA}^{+}_{c,s}$ equals NEXP.

To understand this result, recall that a $\mathsf{QMA}_{c,s}$ protocol is a verification procedure where a quantum verifier ("Arthur") interacts with a potentially deceptive but computationally unbounded prover ("Merlin"). Merlin provides a quantum proof, which Arthur checks with high probability if the input is valid (completeness) and rejects with high probability if it is invalid (soundness). The standard QMA class corresponds to protocols with $(c, s) = (2/3, 1/3)$.

Intuitively, one might expect that limiting Merlin's capabilities – like forcing him to send simpler or less expressive quantum states – would reduce the class's power. However, in many

cases, restricting Merlin's proof in completeness (when the statement is true) has surprisingly little effect. For instance, even if the proof is a subset state – a special case with neither relative phase nor variable amplitudes – QMA remains unchanged Grilo et al. [2014]. This robustness stems from promise gap amplification, which boosts the completeness-soundness gap so that even "imperfect" witnesses succeed if the subset of witnesses form an epsilon net.

In contrast, limiting Merlin's power on the soundness side (when the statement is false) can significantly enhance the verifier's ability to detect deception. For example, if Merlin is forced to provide a quantum state without relative phase, Arthur can assess its *sparsity*, a value that is obscured by interference in general quantum states. In such settings, sparse states have low overlap with the uniform superposition $|+\rangle^{\otimes n}$, making them easier to distinguish.

A related line of research looks at restricting entanglement in QMA, leading to the class QMA(2), where two Merlins provide unentangled proofs Kobayashi et al. [2003]. While QMA(2) is known to lie between QMA and NEXP, its precise power remains an open question despite significant attention Blier and Tapp [2010], Aaronson et al. [2008], Chen and Drucker [2010], Pereszlényi [2012b], Harrow and Montanaro [2013], Gharibian et al. [2018], She and Yuen [2022].

This raises a natural question: What happens if we restrict both entanglement and relative phase? In Jeronimo and Wu [2023b], the classes QMA^+c, s and $\mathsf{QMA}^+(2)c, s$ are defined to capture exactly this. Perhaps unexpectedly, it was shown that for certain constants $c > s > 0$, $\mathsf{QMA}^+(2)c, s = \mathsf{NEXP}$ – driven by the verifier's ability to estimate the sparsity of a given quantum state. This result also opens the door to possibly proving $\mathsf{QMA}(2) = \mathsf{NEXP}$ by amplification, since other constants c', s' exist for which $\mathsf{QMA}^+(2)c', s' = \mathsf{QMA}(2)$.

In Chapter 3, we take this further and show that restricting only the relative phase, without any entanglement constraints, still gives the full power of NEXP. That is, there are constants $c > s > 0$ such that $\mathsf{QMA}^+c, s = \mathsf{NEXP}$. This has striking implications: assuming $\mathsf{EXP} \neq \mathsf{NEXP}$, it follows that QMA^+ cannot be amplified using traditional gap amplification

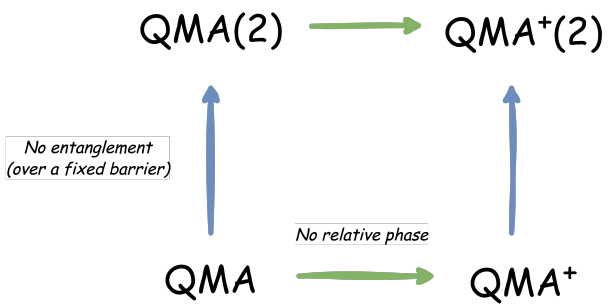

Figure 1.1: Relationship between QMA^+ and $\mathsf{QMA}(2)$. Jeronimo and Wu [2023b] show that for some constants $1 > c > s > 0$, $\mathsf{QMA}^+(2)_{c,s} = \mathsf{NEXP}$. We show that the same is true for QMA^+. Restricting relative phase does not restrict entanglement across a fixed barrier: for example, consider the GHZ state $|0\rangle^{\otimes n} + |1\rangle^{\otimes n}$, or more generally states where the Schmidt vectors have no relative phase.

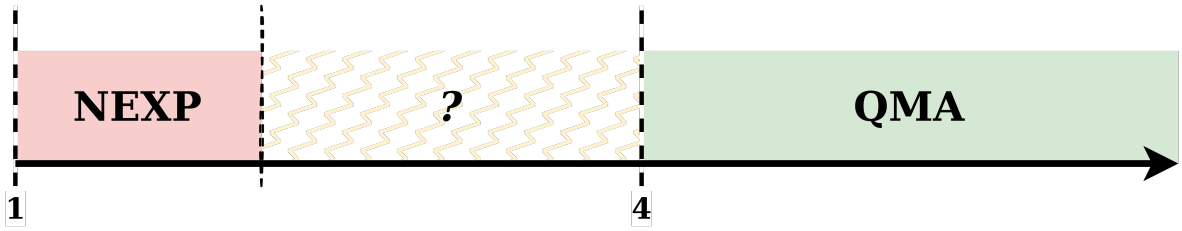

Figure 1.2: Plot of $\mathsf{QMA}^+_{c,s}$ for increasing $\frac{c}{s}$. By our main result, there exist constants c, s where $1 < \frac{c}{s} < 4$ and $\mathsf{QMA}^+_{c,s} = \mathsf{NEXP}$, but by Corollary 90, $\mathsf{QMA}^+_{c,s} = \mathsf{QMA}$ when $\frac{c}{s} > 4$. Gap amplification of QMA^+ would imply $\mathsf{QMA} = \mathsf{NEXP}$.

techniques, since at other values of c', s' we have $\mathsf{QMA}^+c', s' = \mathsf{QMA} \subseteq \mathsf{EXP}$.

Consequently, any attempt to prove that $\mathsf{QMA}(2)$ equals NEXP must crucially leverage the unentanglement constraint, as restricting relative phase alone is already enough to reach NEXP. See Figure 1.1 and Figure 1.2 for visual summaries of these relationships.

The previous sections explored how constraining quantum states – by limiting entanglement or relative phase – can increase computational power. A related and equally intriguing direction is to consider quantum measurements themselves. In particular, hidden-variable theories aim to explain quantum randomness as the result of underlying, inaccessible parameters. If these hidden variables became observable to a quantum computer, they could potentially unlock more powerful models of computation. Motivated by this idea, Aaronson and collaborators Aaronson [2005b], Aaronson et al. [2014] proposed complexity classes that extend BQP by incorporating features inspired by hidden-variable interpretations.

One such extension is the class CQP (Collapse-free Quantum Polynomial-time)Aaronson et al. [2014], which models a quantum computer that, after every gate or intermediate measurement, receives a readout of the current quantum state in the computational basis – without collapsing or disturbing it. Though this does not give polynomial-time solutions to all problems, it does provide access to classes like SZK, and improves unstructured search over Grover's algorithm in a black-box setting. With the addition of quantum advice, this model becomes immensely powerful: Aaronson showed that CQP/qpoly = ALL Aaronson [2018], a big contrast to the more constrained power of BQP/qpoly Aaronson [2005a], Aaronson and Drucker [2013].

This dramatic increase in computational power echoes themes in quantum verification, where an untrusted quantum proof is checked by a quantum verifier. A particularly subtle case is QMA(2) Kobayashi et al. [2003], where the verifier receives two unentangled quantum proofs. Building on this, Aaronson recently posed a compelling new question:

Question 1 (Aaronson [2023])**.** *What is the power of* QMA *where the verifier can make non-collapsing measurements?*

This model is trivially upper bounded by NEXP, but Aaronson conjectured it might actually be equal to NEXP.

In Chapter 4, we prove this conjecture by using similar tools: allowing even a single non-collapsing measurement in a QMA protocol yields a complexity class equal to NEXP. Our proof draws inspiration from the results in earlier chapters, and is unified by a common structure we call the superposition detection problem. Here, the verifier's task is to distinguish between states that are computational basis states and those that are in superposition (i.e., ϵ-far from any basis state). If a verifier can test for superposition on post measurement quantum states, it can simulate any problem in NEXP.

It is important to note that this detection task is not physically implementable with only a single copy of a quantum state – it even allows distinguishing ensembles with identical mixed-state density matrices (see Fact 111). Thus, any successful test must access hidden

information beyond the conventional quantum state, like that provided by non-collapsing measurements. Our results show that such a measurement enables perfect superposition detection, while QMA^+ achieves the same effect by limiting the set of allowed quantum witnesses. In both cases, this yields NEXP-level computational power.

Although superposition detection becomes possible with multiple copies of the state, generating these copies is highly nontrivial – it often requires techniques like postselection Aaronson [2004], inverse exponential precision Pereszlényi [2012b], or an infeasibly large number of quantum witnesses Aaronson et al. [2008], Chen and Drucker [2010], Chiesa and Forbes [2011]. These limitations mirror long-standing obstacles in attempts, starting with Blier and Tapp Blier and Tapp [2010], to prove that $\mathsf{QMA}(2) = \mathsf{NEXP}$.

So far, we have hinted how subtle restrictions or enhancements to quantum proofs and verifiers – such as limiting relative phase, disallowing entanglement, or introducing non-collapsing measurements, can significantly impact computational power. These investigations reveal that quantum complexity classes are remarkably sensitive to structural assumptions about the quantum witness or the verifier's capabilities.

For instance, we show that removing relative phase from quantum proofs yields the class QMA^+, which surprisingly equals NEXP Bassirian et al. [2023b], showing that destructive interference is not necessary for exponential verification power. Similarly, we showed that granting a verifier access to a single non-collapsing measurement, a feature inspired by hidden-variable theories, also suffices to achieve NEXP power. Central to both results is the ability to detect superposition, a task that either requires new measurement capabilities or is trivially solved when witnesses are restricted to non-negative amplitudes.

These findings naturally bring us back to one of the most persistent open questions in quantum complexity theory: What is the power of $\mathsf{QMA}(2)$? That is, how much additional computational power does a verifier gain by receiving two unentangled quantum proofs, instead of one unrestricted proof?

In Chapter 5, we advance this program by proposing a new approach to separate QMA

from $\mathsf{QMA}(2)$. Instead of weakening the structure of the quantum proofs (as in QMA^+), we impose additional entanglement structure. We show that in certain computational models, a single quantum proof is provably weaker than two unentangled quantum proofs, unless $\mathsf{EXP} = \mathsf{NEXP}$. This gives new evidence that $\mathsf{QMA}(2)$ is strictly more powerful than QMA, and opens up a new path to proving that $\mathsf{QMA}(2) = \mathsf{NEXP}$.

Crucially, our approach matches the strengths of earlier techniques while avoiding their core limitations. By embedding the notion of multipartite entanglement into the computational model, we define new complexity classes that generalize $\mathsf{QMA}(2)$ and reveal fresh structure in the landscape of quantum verification. These classes do not just serve our immediate goals – they also offer a broader framework to understand how entanglement interacts with computational power, and we believe they will be of independent interest to the quantum complexity community.

CHAPTER 2

CLASSICAL PROOFS

This chapter is adapted (with minor modifications) from the work of Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. On the power of nonstandard quantum oracles. In *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*, 2023a. doi:10.4230/LIPIcs.TQC.2023.11. URL `https://arxiv.org/abs/2212.00098`.

2.1 Techniques and Contributions

We use a well-known fact of Petersen to encode the edges of any even-degree regular graph in an invertible function f. We then consider natural ways to install f within an oracle; we say that f is *presented* as a particular kind of oracle. For example, a standard oracle presents f through the map $|c, x\rangle \mapsto |c \oplus f(x), x\rangle$, while an in-place oracle presents f through the map $|x\rangle \mapsto |f(x)\rangle$. In general, we consider oracles that give access both to f and f^{-1}. An oracle may also have internal randomness: on every query to a *randomized* oracle, f is chosen uniformly at random from a fixed set of functions F.

Consider the Laplacian L_f of a graph encoded in f. We first provide a test such that for any input state $|\psi\rangle$, the test succeeds with probability expressible in terms of $\langle\psi| L_f |\psi\rangle$, independently of how an oracle presents f. We use this test to construct a QMA protocol verifying that the graph is not an *expander* graph. This problem is primarily motivated by the preimage-testing problem of Fefferman and Kimmel Fefferman and Kimmel [2018], which separates QMA and QCMA relative to a nonstandard oracle. They encode an invertible function π in an oracle *without efficient access to* π^{-1}, and test a property of π^{-1}; by design, this property can be verified but not easily computed. Crucially, we view a permutation and its inverse as the edges of an *undirected graph*; properties of undirected graphs are not sensitive to the ordering of $(x, \pi(x))$. We use multiple permutations to study graphs

of higher degree, and notice that detecting if a graph has a non-expanding region is hard without traversing most of the graph. Some of these ideas are related to the *component mixer* concept of Lutomirski Lutomirski [2011], and are simultaneously and independently explored by Natarajan and Nirkhe Natarajan and Nirkhe [2022].

A randomized oracle presenting a set of functions F can be seen as a quantum channel, so small changes to F cause statistically indistinguishable changes to the oracle. We use this flexibility to modify non-expansion testing to a simple permutation problem: do the functions $f \in F$ stabilize a small set $V \subseteq [N]$, or is F the set of all permutations on $[N]$? Notice that F is a group in both cases. When an oracle presenting F preserves the group structure of F, we can use representation theory. For this problem, this is satisfied by an *in-place* oracle; the oracle is then an orthogonal projector onto one of two symmetric subspaces of matrices. After finding an orthogonal basis for each subspace, we construct a hybrid argument to prove that only witnesses with knowledge of V can help a quantum verifier efficiently decide this problem. We also use representation theory to give a QCMA protocol for an analogous permutation problem in randomized standard oracles.

We finally study the permutation problem in a randomized phase oracle. We directly analyze the effect of the oracle on any input density matrix; with high probability, the oracle decreases the magnitude of every off-diagonal term by a $\frac{1}{2^{\mathrm{poly}(n)}}$ factor. We then construct a hybrid argument, bounding our measure of progress using an inequality relating the sizes of Schatten p-norms. When the state space is not too large, we prove that an exponential number of queries are required to distinguish most YES instances from the NO instance, *regardless of input state*. As a result, no witness can help a verifier distinguish YES from NO.

Note that our quantum verifiers are not fully general. Our lower bound techniques restrict the number of extra workspace qubits in the verifier; however, our upper bounds also work in this setting. In Section 2.4.2, we explain these restrictions in more detail and discuss prospects for generalizing our results.

2.2 Related work

Quantum oracle models A fundamental constraint of quantum oracle models is that they must be unitary. We describe several nonstandard oracle models used in quantum computing:

- A *quantum* oracle is any unitary operation U in the full Hilbert space. Although the operation is unitary, the verifier doesn't necessarily have access to U^{-1}. Oracles like these are not typically classical because the unitary's action is not efficiently and classically representable.

- An *in-place* oracle maps $|x\rangle \to |\pi(x)\rangle$ for some classical invertible function π. Again, this computation is not efficiently reversible since the verifier may not have access to π^{-1}. When a standard oracle gives access to π^{-1}, an in-place oracle query can be simulated in two queries; otherwise, an exponential number of queries are required to construct one from the other Kashefi et al. [2002].

- A *phase* oracle puts the output of a classical function f in the phase of a basis state. We consider the map $|x\rangle \to e^{f(x)\cdot 2\pi i/N} |x\rangle$. To contrast, note that the map $|c, x\rangle \to e^{cf(x)\cdot 2\pi i/N} |c, x\rangle$ is unitarily equivalent to the standard oracle.

All of these oracles can optionally have *internal randomness*, as considered by Harrow and Rosenbaum Harrow and Rosenbaum [2011]; we call these *randomized* oracles. On every query to a randomized oracle, a unitary is chosen at random from a fixed set. This can be very powerful; for example, Harrow and Rosenbaum [2011] gives examples of randomized oracles where problems *impossible* to decide with classical queries can be decided with a single quantum query.

QMA and QCMA The *Merlin-Arthur* style of complexity classes considers a decision problem and two players. The magician (Merlin) has claimed the answer to the decision problem is YES, and gives the verifier a token (the *proof* or *witness*) to convince them. The verifier

(Arthur) must then ensure the answer is actually YES. Given a problem with size n, the verifier must accept a correct witness (i.e. when the answer is YES) with probability $1/q$ higher than a "lying" witness (i.e. when the answer is NO) for some $q = \text{poly}(n)$. The set of problems that can be decided this way in a classical setting is known as Merlin Arthur (MA). If the verifier is a quantum computer, this is QCMA; if the witness can be any quantum state, this is QMA.

	verifier is classical	**verifier is quantum**
witness is classical	MA	QCMA
witness is quantum	-	QMA

Table 2.1: Complexity classes in the style of *Merlin-Arthur*. QCMA is a subset of QMA where the witness can be efficiently written as a classical bitstring.

Since any classical bitstring can be efficiently written as a quantum state, QCMA $\subseteq$ QMA. But is the reverse true? Even the *oracle* version of this problem is open: at the top of a recent list of open problems, Aaronson asks for a standard oracle that separates the two classes Aaronson [2021]. All previous progress Aaronson and Kuperberg [2007], Fefferman and Kimmel [2018], Natarajan and Nirkhe [2022] relies on specifically chosen *nonstandardness* in the oracle.

Natarajan and Nirkhe Natarajan and Nirkhe [2022] make progress on a standard oracle separation of QMA and QCMA by constructing an oracle with randomness. They simultaneously and independently provide a QMA protocol for testing non-expansion of a graph in an oracle. To prove their lower bound, they combine the adversary method, the polynomial method, and a reduction to a problem of Ambainis, Childs, and Liu Ambainis et al. [2011a]. However, their notion of randomness is different from ours and other works Harrow and Rosenbaum [2011], Fefferman and Kimmel [2018], Arora et al. [2022], and acts as follows: when an oracle is first queried, it chooses a function f from a distribution, but on subsequent queries, it uses the same function f. By contrast, our notion of randomness is memoryless: an oracle chooses f from a uniform distribution on F for *every* query. This allows one to

make small changes to F without affecting the success of the QMA protocol; we use this flexibility to study a simpler permutation problem.

2.3 Our setup

Consider a d-regular graph on $N := 2^n$ vertices for any n and even d. We show that an invertible function can list the edges adjacent to each vertex in G.

Definition 2 (Graph-coded function)**.** *Consider a d-regular graph G (for even d) on N vertices. A G-coded function is a function $f : [N] \times [d/2] \to [N]$, such that $f_i(x) := f(x, i)$ is a bijection for each $i \in [d/2]$, and each edge is uniquely represented by a tuple $(x, f_i(x))$.*

Remark 3 (Even-degree regular graphs have graph-coded functions)**.** *Every regular graph G of even degree has a G-coded function.*

Proof. A d-regular graph G of even degree always has a 2-factorization Petersen [1900]. This means that the edges of G can be partitioned into $d/2$ edge-disjoint subgraphs $[E_1, \ldots, E_{d/2}]$ where in each E_i, all vertices have degree two (i.e. a collection of cycles). Thus, we can represent each E_i with a permutation π_i, where the edge $(x, y) \in E_i$ if and only if $\pi_i(x) = y$ or $\pi_i(y) = x$. Then $f(x, i) := \pi_i(x)$ is a G-coded function. □

Graph-coded functions f are bijective, and therefore invertible. We now *present* f in various oracle models. Note that we define all oracles with access both to f and f^{-1}.

Definition 4 (An oracle model *presents* a function f)**.** *For each oracle model below (e.g. standard oracle), we say that this oracle model* presents *the function f.*

Remark 5. *For notational convenience, we refer to a qubit z that controls the inversion of a function f as taking on values in $\{\pm 1\}$, so that f^z is either $f^1 = f$ or f^{-1}.*

Definition 6 (Standard oracle)**.** *For any $f : [N] \to [N]$, define $U_f : \mathbb{C}^{2N^2} \to \mathbb{C}^{2N^2}$ as*

$$U_f \sum_{c,x \in [N], z \in \{\pm 1\}} \alpha_{c,x,z} \left|c, x, z\right\rangle := \sum_{c,x \in [N], z \in \{\pm 1\}} \alpha_{c,x,z} \left|c \oplus f^z(x), x, z\right\rangle . \tag{2.1}$$

Definition 7 (In-place oracle Kashefi et al. [2002]). *For any permutation $\pi : [N] \to [N]$, define $\widetilde{U}_\pi : \mathbb{C}^{2N} \to \mathbb{C}^{2N}$ as*

$$\widetilde{U}_\pi \sum_{x\in[N],z\in\{\pm1\}} \beta_{x,z} |x,z\rangle := \sum_{x\in[N],z\in\{\pm1\}} \beta_{x,z} |\pi^z(x),z\rangle \ . \tag{2.2}$$

Remark 8 (Kashefi et al. [2002]). *A standard oracle U_f (with access to f^{-1}) can simulate an in-place oracle $\widetilde{U}_f$ in two queries:*

$$(\mathbb{I}\otimes X)\circ U_f \circ (SWAP_{n,n}\otimes X)\circ U_f |0\rangle^{\otimes n} |x,z\rangle = |0\rangle^{\otimes n} |f^z(x),z\rangle \ . \tag{2.3}$$

Definition 9 (N^{th} root of unity). *Define the N^{th} root of unity as $\omega_N := e^{2\pi i/N}$.*

Definition 10 (Phase oracle). *For any function $f : [N] \to [N]$, define $\overline{U}_f : \mathbb{C}^{2N} \to \mathbb{C}^{2N}$ as*

$$\overline{U}_f \sum_{x\in[N],z\in\{\pm1\}} \alpha_{x,z} |x,z\rangle := \sum_{x\in[N],z\in\{\pm1\}} \alpha_{x,z}\omega_N^{f^z(x)} |x,z\rangle \ . \tag{2.4}$$

We describe how an oracle in our setup exhibits internal randomness. On each query, a *randomized* oracle chooses a function uniformly from a set F. We say that a randomized oracle *presents* F.

Remark 11. *Given a unitary U, we use the notation $\mathcal{U}$ to denote an operator on density matrices; that is,*

$$\mathcal{U}[\rho] := U\rho U^\dagger \, . \tag{2.5}$$

Definition 12 (Randomized oracle (e.g. Harrow and Rosenbaum [2011], Fefferman and Kimmel [2018])). *For any set F of functions $f : [N] \to [N]$ corresponding to oracles $\{U_f \mid f \in$*

$F\}$, *define the linear operator* $\mathcal{O}_F$ *as*

$$\mathcal{O}_F := \frac{1}{|F|} \sum_{f \in F} \mathcal{U}_f \,. \tag{2.6}$$

We match the notation of randomized oracle $\mathcal{O}_F$ *with oracle* U_f*; e.g.* $\widetilde{\mathcal{O}}_F$ *is a randomized* in-place *oracle.*

2.3.1 Problem statements

The problems below are not fully specified without the choice of oracle model. We prepend the names below with the choice of oracle model; for example, we denote Problem 13 in a standard oracle as STANDARD NON-EXPANSION(d, α, ϵ).

Problem 13 (NON-EXPANSION(d, α, ϵ))**.** *Consider an oracle* U_f *presenting a* G*-coded function* f*.*

1. *In a YES instance, we are promised that* G *is a union of two disconnected* d*-regular graphs, and that the smaller graph has* N^α *vertices.*

2. *In a NO instance, we are promised that* G *is* d*-regular and has spectral gap at least* ϵ *(for example, an* expander graph*).*

The problem is to decide whether U_f *is a YES instance or NO instance.*

We also consider a version of this problem with *randomized* oracles, where each randomized YES instance is specified by the set of vertices V of the smaller graph. On each query, an oracle chooses an graph-coded function f uniformly at random that corresponds to a graph where V and $[N]/V$ are disconnected.

Problem 14 (RANDOMIZED NON-EXPANSION(d, α, ϵ))**.** *Consider a randomized oracle* $\mathcal{O}_F$ *presenting a set of graph-coded functions* F*.*

1. *Each subset $V \subseteq [N]$ of size $|V| = N^\alpha$ specifies a YES instance $\mathcal{O}_{F_V}$. Let F_V be the set of all G-coded functions of d-regular graphs G with no edges between V and $[N]/V$.*

2. *There is a single NO instance $\mathcal{O}_{F_\varnothing}$. Let $F_\varnothing$ be the set of all G-coded functions of d-regular graphs G with spectral gap at least ϵ.*

The problem is to decide whether $\mathcal{O}$ is a YES instance or a NO instance.

In the configuration model of a random graph, F_V contains all functions $f(x,i)$ such that $f_i(x) := f(x,i)$ is the union of a permutation on $[N]/V$ and a permutation on V. In fact, we can use the oracle's internal randomness to adjust the underlying set F, and even consider graphs that are not typically expander graphs.

Definition 15 (Subset indicator)**.** *For a set $V \subseteq [N]$, define the function $i_V : [N] \to \{V, [N]/V\}$ as*

$$i_V(x) = \begin{cases} V & x \in V \\ [N]/V & x \notin V\,. \end{cases} \tag{2.7}$$

Definition 16 (Permutations that stabilize a subset)**.** *For a set $V \subseteq [N]$, define the set of permutations*

$$T_V := \{\pi : [N] \to [N] \,:\, i_V(x) = i_V(\pi(x))\, \forall x \in [N]\}\,. \tag{2.8}$$

We say that T_V stabilizes *the subset V.*

Problem 17 (RANDOMIZED HIDDEN SUBSET(α))**.** *Consider a randomized oracle $\mathcal{O}_F$ presenting a set of functions F.*

1. *Each subset $V \subseteq [N]$ of size $|V| = N^\alpha$ specifies a YES instance $\mathcal{O}_{T_V}$.*

2. *There is a single NO instance $\mathcal{O}_{T_\varnothing}$, where $T_\varnothing$ is the set of all permutations of $[N]$.*

The problem is to decide whether $\mathcal{O}$ is a YES instance or a NO instance.

Notice that T_V is a group under function composition. One can generalize this algebraic structure to a problem distinguishing oracles presenting subgroups of $T_\varnothing$ from an oracle presenting $T_\varnothing$:

Problem 18 (RANDOMIZED HIDDEN SUBGROUP$(\times, \{H_i\})$)**.** *Consider the set $T_\varnothing$ of all permutations on $[N]$ as a group with operation $\times$, such that each $H_i \subsetneq T_\varnothing$ is also a group. Suppose a randomized oracle $\mathcal{O}$ presents either $T_\varnothing$ or any H_i.*

1. *Each subgroup H_i specifies a YES instance $\mathcal{O}_{H_i}$.*
2. *There is a single NO instance $\mathcal{O}_{T_\varnothing}$, where $T_\varnothing$ is the set of all permutations of $[N]$.*

The problem is to decide whether $\mathcal{O}$ is a YES instance or a NO instance.

For example, RANDOMIZED HIDDEN SUBSET(α) is a special case of RANDOMIZED HIDDEN SUBGROUP$(\times, \{H_i\})$ using the group operation of function composition.

2.4 Our results

2.4.1 Non-expansion and quantum witness

Our first result shows that there is a one-query QMA protocol for NON-EXPANSION(d, α, ϵ) in many oracle models presenting a graph-coded function.

Theorem 19. *There is a* QMA *protocol for* STANDARD NON-EXPANSION(d, α, ϵ) *and* IN-PLACE NON-EXPANSION(d, α, ϵ) *at every even $d \geq 4$, all $0 < \alpha < \frac{1}{2}$, and all constant $\epsilon > 0$.*

Graphs with good *expansion* are well-connected despite their sparsity. For any graph G, let A_G be the adjacency matrix of G, and $L_G = d\mathbb{I} - A_G$ be the *graph Laplacian* of G. The smallest eigenvalue of L_G is $\lambda_1(L_G) = 0$, and the next-smallest eigenvalue $\lambda_2(L_G)$ measures

the expansion of G. In this framework, NON-EXPANSION(d, α, ϵ) asks if an oracle presenting a G-coded function has $\lambda_2(L_G) = 0$ (YES), or if $\lambda_2(L_G) \geq \epsilon$ (NO).

At the heart of our protocol is the *spectral test*, which takes an input state $|\psi\rangle$ and fails with probability proportional to $\langle\psi| L_G |\psi\rangle$. We describe the spectral test for both standard oracles and in-place oracles in Section 2.5.1. A state that passes the spectral test is essentially supported on a subspace according to $\lambda(L_G) = o(\frac{1}{\text{poly}(n)})$; in a NO instance, this is one-dimensional, and in a YES instance, this is at least two-dimensional. In fact, the uniform superposition over all inputs, $|+\rangle^{\otimes n}$, is always in this subspace. As a result, our protocol (Theorem 19) either runs the spectral test, or checks if the input state is close to $|+\rangle^{\otimes n}$.

Consider the randomized variant of NON-EXPANSION(d, α, ϵ). The graph of any graph-coded function presented in a YES instance is guaranteed to have a small set V (i.e. $|V| = N^\alpha$) disconnected from the rest of the graph. As a result, there is a state, defined only by the vertices of V, that is all-but-negligibly supported in the $\lambda(L_G) = 0$ subspace. This state is the *subset state* $|V\rangle$:

Definition 20 (Subset state)**.** *For any non-empty subset $S \subseteq [N]$, define the subset state of S as*

$$|S\rangle := \frac{1}{\sqrt{|S|}} \sum_{x \in S} |x\rangle \ . \tag{2.9}$$

Since $|V\rangle$ is a good witness for *every* graph encoded in a YES instance, the QMA protocol works just as well in the randomized setting (Theorem 26).

Randomized oracles that present a set F of graph-coded functions are stable to small changes in the set F. In fact, an oracle presenting F encoding all d-regular expander graphs is indistinguishable from an oracle presenting F encoding all d-regular graphs. The latter oracle can be simulated with $d/2$ queries to the NO instance of RANDOMIZED HIDDEN SUBSET(α); we show in Theorem 27 that the same QMA protocol can also decide this problem.

2.4.2 Randomized in-place oracles and classical witness

Our second result shows that a general class of verifiers cannot decide IN-PLACE RANDOMIZED HIDDEN SUBSET(α).

Theorem 21 (informal)**.** *No quantum verifier with $O(\log(n))$ additional workspace qubits can efficiently decide* IN PLACE RANDOMIZED HIDDEN SUBSET(α) *given a polynomial sized classical witness.*

Recall that in this problem, a verifier has access to a quantum channel and a polynomial-sized classical witness, and must distinguish whether the oracle presents a uniformly random permutation or a permutation that stabilizes a hidden subset V. Let $\mathcal{Y}$ be the set of all YES instances; note that each instance is uniquely defined by a subset V.

Suppose there exists a QCMA algorithm for this problem. Since there are at most $O(2^{\text{poly}(n)})$ different classical witnesses, there exists a set of YES instances $\mathcal{Y}'$ that share the same witness, such that $\left|\mathcal{Y}'\right| / |\mathcal{Y}| = \Omega(2^{-\text{poly}(n)})$. We can refute the existence of such an algorithm by proving that the same verification "strategy" cannot distinguish all instances of $\mathcal{Y}'$ from the NO case with non-negligible probability. A "strategy" is exactly a quantum algorithm: a series of unitaries and oracle queries, followed by a POVM. Without loss of generality, a T-query algorithm alternates between unitaries and oracle queries on $\mathcal{H}_O \otimes \mathcal{H}_W$ followed by a measurement[1], where $\mathcal{H}_O$ is the Hilbert space of the "oracle" qubits and $\mathcal{H}_W$ is the extra workspace:

$$\mathcal{E}_O[\rho_0] = (\mathcal{O} \otimes \mathbb{I}) \circ \mathcal{U}_T \circ \ldots \circ \mathcal{U}_2 \circ (\mathcal{O} \otimes \mathbb{I}) \circ \mathcal{U}_1[\rho_0]\,. \tag{2.10}$$

One may try to use the hybrid argument of Bennett, Bernstein, Brassard, and Vazirani Bennett et al. [1997] and Ambainis Ambainis [2000] to prove that diamond norm $\left|\mathcal{E}_{\widetilde{\mathcal{O}}_{T_V}} - \mathcal{E}_{\widetilde{\mathcal{O}}_{T_\varnothing}}\right|_\diamond$ is small in expectation over the choice of $\widetilde{\mathcal{O}}_{T_V} \in \mathcal{Y}'$. This would imply that the verifier can-

1. Note that the last operation does not have to be a unitary – one can simply replace a unitary followed by a POVM with another equivalent POVM.

not distinguish all instances of $\mathcal{Y}'$ with the same strategy. We can consider the optimal distinguishing probability in terms of $\left|\mathcal{E}_{\widetilde{\mathcal{O}}_{T_V}}[\rho_0] - \mathcal{E}_{\widetilde{\mathcal{O}}_{T_\varnothing}}[\rho_0]\right|_1$ for some fixed $\rho_0 \in \mathcal{H}_O \otimes \mathcal{H}_W$.

However, this statistical argument does not hold for some choices of $\mathcal{Y}'$. Consider the following simple example: $\mathcal{Y}'$ contains all V such that $1 \in V$. First, $\mathcal{Y}'$ satisfies the size implied by the pigeonhole principle. Second, for $\rho_0 = |1\rangle\langle 1| \otimes \mathbb{I}$, $\left|\widetilde{\mathcal{O}}_{T_V}[\rho_0] - \widetilde{\mathcal{O}}_{T_\varnothing}[\rho_0]\right|_1$ is large for *all* instances in $\mathcal{Y}'$, since $|1\rangle\langle 1|$ mixes only within a small subset. Note that this only implies the existence of an *instance-specific* POVM distinguishing each YES instance in $\mathcal{Y}'$ from the NO instance. By contrast, a verification strategy has a *fixed* POVM $\{E, \mathbb{I} - E\}$. This allows us to prove that the following value is small on average over the choice of V:

$$\left|\mathrm{Tr}\left[E\mathcal{E}_{\widetilde{\mathcal{O}}_{T_V}}[\rho_0]\right] - \mathrm{Tr}\left[E\mathcal{E}_{\widetilde{\mathcal{O}}_{T_\varnothing}}[\rho_0]\right]\right| \tag{2.11}$$

We must bound this value for arbitrary choices of E, ρ_0 and $\mathcal{U}_i$ fixed in the algorithm. In order to do this, we leverage tools from representation theory; this allows us to see randomized oracles in our problem as *orthogonal projectors* into a subspace of matrices with low dimension, and prove that the density matrix of good distinguishers is characterized by the hidden subset. One caveat of our technique is that the verifier is only allowed to have $O(\log(n))$ extra workspace qubits. This restriction is necessary to reduce the subspace dimension to regimes we can handle.

Representation theory has been previously used to study symmetric operators on variables (in probability) or qubits (in quantum computing) using the language of de Finetti theorems (e.g. Harrow [2013b]); these operators project into subspaces of permutation-invariant sequences or quantum states. By contrast, we notice that some randomized oracles are symmetric operators on *density matrices*. This allows us to explicitly find an orthogonal basis for the associated symmetric subspaces. We match oracle models with problems with the same group structure: RANDOMIZED HIDDEN SUBSET(α) for in-place oracles in Section 2.6.1, and an analogous special case of RANDOMIZED HIDDEN SUBGROUP($\times, \{H_i\}$)

for standard oracles in Section 2.6.2.

2.4.3 Randomized phase oracles: no witness can help

Our third result shows that deciding RANDOMIZED HIDDEN SUBSET(α) in a phase oracle is much harder than other oracle models we consider. A random phase has *zero* expectation. We use this fact to show that queries to most YES instances and the NO instance reduce the magnitude of each off-diagonal of the density matrix by an exponential factor, regardless of the input state. We bound the Frobenius norm of the difference of query outputs to show that these instances are statistically indistinguishable when the state space is not too large. As a result, no untrustworthy witness can help decide this problem.

Theorem 22 (informal)**.** *No quantum verifier with $o(n)$ additional workspace qubits can efficiently decide* PHASE RANDOMIZED HIDDEN SUBSET(α) *given* any *unbounded witness. Moreover, these verifiers require an exponential number of queries to statistically distinguish a YES instance from the NO instance, for* each *of asymptotically all YES instances.*

We defer the formal proof to Section 2.7. Note that the query lower bound here is *statistical*. In the NO instance, a witness is designed to fool the verifier; in order to overcome this, the verifier must use the witness in tandem with the oracle. But this cannot be done efficiently: distinguishing the NO instance from nearly any YES instance requires an exponential number of queries, *regardless of input state*.

In fact, Theorem 22 holds for any oracle that sends $|c, x, z\rangle \to \omega_N^{c \cdot f^z(x)} |c, x, z\rangle$, where the c register has $k = o(n)$ qubits, while at $k = n$ qubits, the oracle is unitarily equivalent to a standard oracle, and thus has a QMA protocol for RANDOMIZED HIDDEN SUBSET(α) by Theorem 19.

2.5 Verifying non-expansion with a quantum witness

2.5.1 The spectral test

We give a test that takes an input state $|\psi\rangle = \sum_{x\in[N]} a_x |x\rangle$ on n qubits, and fails with probability proportional to $\langle\psi| L_G |\psi\rangle$. This relies on a curious fact:

Lemma 23. *Consider a d-regular graph G (for even d) on 2^n vertices and a G-coded function f. Suppose we have a normalized quantum state $|\psi\rangle = \sum_{x\in[N]} a_x |x\rangle$ on n qubits. Then*

$$\sum_{i\in[d/2]} \sum_{x\in[N]} \|a_x \pm a_{f(x,i)}\|^2 = d \pm \langle\psi| A_G |\psi\rangle \ . \tag{2.12}$$

Proof.

$$\sum_{i\in[d/2]} \sum_{x\in[N]} \|a_x \pm a_{f(x,i)}\|^2 = \sum_{i\in[d/2]} \sum_{x\in[N]} \|a_x\|^2 + \|a_{f(x,i)}\|^2 \pm (a_x a^*_{f(x,i)} + a^*_x a_{f(x,i)}) \tag{2.13}$$

$$= d \pm \sum_{i\in[d/2]} \sum_{x\in[N]} (a_x a^*_{f(x,i)} + a^*_x a_{f(x,i)}) \tag{2.14}$$

$$= d \pm \langle\psi| A_G |\psi\rangle \, . \tag{2.15}$$

□

We construct the spectral test with one query either to a standard oracle or in-place oracle presenting a graph-coded function f. The former (Procedure 24) is a SWAP test but with an oracle query in the middle. The latter (Procedure 25) relies on controlled access to the in-place oracle.

Procedure 24 (Spectral test with a standard oracle)**.** *Consider a d-regular graph G on $N = 2^n$ vertices where d is even, and normalized state $|\psi\rangle = \sum_{x\in[N]} a_x |x\rangle \in \mathbb{C}^N$. We assume access to a standard oracle $U_f : \mathbb{C}^{k\times k}$ for $k = N^2 2^{\lceil \log_2 d \rceil}$, which acts on a basis*

vector as

$$U_f \left|c, x, i, z\right\rangle = \left|c \oplus f^z(x, i), x, i, z\right\rangle \ , \tag{2.16}$$

for $c, x \in [N]$, $i \in 2^{\lceil \log_2 d \rceil - 1}$, *and* $z \in \{\pm 1\}$.

1. *Pick* $i \in [d/2]$ *uniformly at random, and prepare the state* $\left|i\right\rangle \in \mathbb{C}^{2^{\lceil \log_2 d \rceil - 1}}$.

2. *Prepare a qubit in the state* $\left|+\right\rangle = \frac{\left|1\right\rangle + \left|-1\right\rangle}{\sqrt{2}} \in \mathbb{C}^2$. *(Recall that we label the values of this register in* $\{\pm 1\}$*.)*

3. *Combine* n *registers* $\left|0\right\rangle^{\otimes n}$, *the input state* $\left|\psi\right\rangle$, $\left|i\right\rangle$ *and* $\left|+\right\rangle$ *to create* $\left|0\right\rangle^{\otimes n} \left|\psi\right\rangle \left|i\right\rangle \left|+\right\rangle$.

4. *Apply the oracle* U_f, *which creates the state*

$$\frac{1}{\sqrt{2}} \sum_{x \in [N]} a_x \left(\left|f(x,i)\right\rangle \left|x\right\rangle \left|i\right\rangle \left|1\right\rangle + \left|f^{-1}(x,i)\right\rangle \left|x\right\rangle \left|i\right\rangle \left|-1\right\rangle \right) . \tag{2.17}$$

5. *Swap the first two sets of* n *qubits, controlled by the last qubit. This creates the state*

$$\frac{1}{\sqrt{2}} \sum_{x \in [N]} a_x \left(\left|f(x,i)\right\rangle \left|x\right\rangle \left|i\right\rangle \left|1\right\rangle + \left|x\right\rangle \left|f^{-1}(x,i)\right\rangle \left|i\right\rangle \left|-1\right\rangle \right) \tag{2.18}$$

$$= \frac{1}{\sqrt{2}} \sum_{x \in [N]} a_x \left|f(x,i)\right\rangle \left|x\right\rangle \left|i\right\rangle \left|1\right\rangle + \frac{1}{\sqrt{2}} \sum_{x \in [N]} a_{f(x,i)} \left|f(x,i)\right\rangle \left|x\right\rangle \left|i\right\rangle \left|-1\right\rangle \ . \tag{2.19}$$

6. *Apply a Hadamard on the last qubit, which creates the state*

$$\frac{1}{2} \sum_{x \in [N]} \left(a_x + a_{f(x,i)}\right) \left|f(x,i)\right\rangle \left|x\right\rangle \left|i\right\rangle \left|1\right\rangle + \frac{1}{2} \sum_{x \in [N]} \left(a_x - a_{f(x,i)}\right) \left|f(x,i)\right\rangle \left|x\right\rangle \left|i\right\rangle \left|-1\right\rangle \ . \tag{2.20}$$

7. *Measure the last qubit and accept if it is* 1.

Moreover, by Lemma 23, this procedure fails with probability

$$\frac{1}{d/2}\sum_{i\in[d/2]}\sum_{x\in[N]}\frac{\|a_x - a_{f(x,i)}\|^2}{4} = \frac{\langle\psi| L_G |\psi\rangle}{2d}\,. \tag{2.21}$$

Procedure 25 (Spectral test with an in-place oracle)**.** *Consider a d-regular graph G on $N = 2^n$ vertices where d is even, and normalized state $|\psi\rangle = \sum_{x\in[N]} a_x |x\rangle \in \mathbb{C}^N$. We assume controlled access to an in-place oracle $\widetilde{U}_f : \mathbb{C}^{k\times k}$ for $k = N2^{\lceil\log_2 d\rceil+1}$, which acts on a basis vector as*

$$\widetilde{U}_f |a, x, i, z\rangle = |a, f^{a\cdot z}(x, i), i, z\rangle\ . \tag{2.22}$$

for control qubit $a \in \{0, 1\}$, $x \in [N]$, $i \in 2^{\lceil\log_2 d\rceil-1}$, and $z \in \{\pm 1\}$.[2]

1. *Pick $i \in [d/2]$ uniformly at random, and prepare the state $|i\rangle \in \mathbb{C}^{2^{\lceil\log_2 d\rceil-1}}$.*

2. *Prepare a qubit in the state $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}} \in \mathbb{C}^2$.*

3. *Combine $|+\rangle$, the input state $|\psi\rangle$, $|i\rangle$, and a register $|1\rangle$ to create $|+\rangle |\psi\rangle |i\rangle |1\rangle$.*

4. *Apply the oracle $\widetilde{U}_f$, which creates the state*

$$\frac{1}{\sqrt{2}}\sum_{x\in[N]} a_x \left(|0\rangle |x\rangle + |1\rangle |f(x, i)\rangle\right) |i\rangle |1\rangle \tag{2.23}$$

$$=\frac{1}{\sqrt{2}}\sum_{x\in[N]} \left(a_x |0\rangle + a_{f^{-1}(x,i)} |1\rangle\right) |x\rangle |i\rangle |1\rangle\ . \tag{2.24}$$

5. *Apply a Hadamard on the first qubit, which creates the state*

$$\frac{1}{2}\sum_{x\in[N]} \Big(\left(a_x + a_{f^{-1}(x,i)}\right) |0\rangle + \left(a_x - a_{f^{-1}(x,i)}\right) |1\rangle\Big) |x\rangle |i\rangle |1\rangle\ . \tag{2.25}$$

2. Note that this procedure does not actually need access to the inverse of f to conduct the spectral test.

6. Measure the first qubit and accept if it is 0.

Moreover, by Lemma 23, this procedure fails with probability

$$\frac{1}{d/2}\sum_{i\in[d/2]}\sum_{x\in[N]}\frac{\|a_x - a_{f^{-1}(x,i)}\|^2}{4} = \frac{\langle\psi| L_G |\psi\rangle}{2d}\,. \tag{2.26}$$

2.5.2 A one-query protocol

Theorem 19 (restated)**.** *For every even $d \geq 4$, all $0 < \alpha < \frac{1}{2}$, and all constant $\epsilon > 0$, there is a* QMA *protocol for* STANDARD NON-EXPANSION(d, α, ϵ) *and* IN-PLACE NON-EXPANSION(d, α, ϵ)*, .*

Proof. Suppose the oracle presents a G-coded function. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ be the eigenvalues of the graph Laplacian L_G. Note that the smallest eigenvalue of a regular graph G is $\lambda_1 = 0$. We always choose the eigenvector associated with λ_1 as a uniform superposition over vertices of the graph (i.e. $|\lambda_1\rangle := |[N]\rangle = |+\rangle^{\otimes n}$).

Suppose Arthur receives a state $|\psi\rangle = \sum_{i\in[N]} \alpha_i |\lambda_i\rangle$ from Merlin. Consider the following strategy:

- With probability $\frac{1}{2}$, measure $|\psi\rangle$ in the Hadamard basis. Fail if it is in the basis state according to $|+\rangle^{\otimes n}$, and pass otherwise.
- With probability $\frac{1}{2}$, use the spectral test (Procedure 24 or Procedure 25, respectively).

The probability of failure FAIL is

$$\text{FAIL} = \frac{1}{2}\| \langle\psi| |+\rangle^{\otimes n} \|^2 + \frac{1}{2}\frac{\langle\psi| L_G |\psi\rangle}{2d} \tag{2.27}$$

$$= \frac{1}{2}\Big(\|\alpha_1\|^2 + \frac{1}{2d}\sum_{i=1}^{N}\lambda_i\|\alpha_i\|^2\Big)\,. \tag{2.28}$$

In a NO instance, $\lambda_k = \Omega(1)$ for all $k > 1$. So the probability of failure is always a positive

constant:

$$\text{FAIL}_{\text{NO}} = \frac{1}{2}\|\alpha_1\|^2 + \frac{1}{2d}\sum_{i=2}^{N}\Omega(\|\alpha_i\|^2) = \Omega(\sum_{i=1}^{N}\|\alpha_i\|^2) = \Omega(1)\,. \tag{2.29}$$

In a YES instance, the spectrum of L_G is the combined spectrum of the two disconnected graphs. This means $\lambda_1 = \lambda_2 = 0$, and the associated eigenvectors are linear combinations of $|V\rangle$ and $|[N]/V\rangle$. Recall that $|\lambda_1\rangle := |+\rangle^{\otimes n}$. We find the orthogonal eigenvector $|\lambda_2\rangle$ in this subspace by inspection:

$$|\lambda_2\rangle = \sqrt{\frac{N-|V|}{N}}\,|V\rangle + \sqrt{\frac{|V|}{N}}\,|[N]/V\rangle\ . \tag{2.30}$$

Note that any vector with $\|\alpha_2\|^2 = 1 - o(\frac{1}{\text{poly}(n)})$ has negligible probability of failure:

$$\text{FAIL}_{\text{YES}} = \frac{1}{2}\|\alpha_1\|^2 + \frac{1}{2d}\sum_{i=1}^{N}\lambda_i\|\alpha_i\|^2 = O(\|\alpha_1\|^2 + \sum_{i=3}^{N}\|\alpha_i\|^2) = O(1-\|\alpha_2\|^2)\,. \tag{2.31}$$

Suppose Merlin sends the subset state $|V\rangle$. Since $\|\,\langle V|\lambda_2\rangle\,\|^2 = 1 - \frac{|V|}{N} = 1 - O(\frac{1}{2^{\text{poly}(n)}})$, the strategy has probability of failure $O(\frac{1}{2^{\text{poly}(n)}})$. □

In general, the spectral test can be used in a QMA protocol to test the magnitude of the second-smallest or largest eigenvalue of a graph Laplacian to inverse polynomial precision. The former is a measure of the quality of a graph's expansion, and the latter is related to a measure of a graph's bipartiteness named the *bipartiteness ratio* Trevisan [2008].

Because this QMA protocol requires only one query of either a standard oracle or an in-place oracle, it works even when these oracles are randomized.

Theorem 26. *For all even $d \geq 4$, all $0 < \alpha < \frac{1}{2}$, and all constant $\epsilon > 0$, there is a* QMA *protocol for* STANDARD RANDOMIZED NON-EXPANSION(d, α, ϵ) *and* IN-PLACE RANDOMIZED NON-EXPANSION(d, α, ϵ) .

Proof. The strategy in Theorem 19 also works here. Consider any G-coded function pre-

sented in a YES instance; the same vertices V are exactly the vertices of the smaller component of G. So the witness $|V\rangle$ is close to the second eigenvector $|\lambda_2\rangle$, and the failure probability is negligible. Now consider any G-coded function presented in a NO instance. By definition, G is an expander graph, so the failure probability is always a positive constant. □

Because a randomized oracle chooses a function uniformly from a set F, it is statistically indistinguishable from an oracle with exponentially small changes to F. We use this fact to simplify the NO instance in RANDOMIZED NON-EXPANSION(d, α, ϵ). Suppose the NO instance instead presents graph-coded functions of *all* d-regular graphs. Since $1 - O(\frac{1}{\text{poly}(N)}) = 1 - O(\frac{1}{2^{\text{poly}(n)}})$ graphs have a constant spectral gap Friedman [2004] when $d > 2$, the failure probability in the QMA protocol changes by at most $O(\frac{1}{2^{\text{poly}(n)}})$.

Notice that with this modification, the oracles are exactly $d/2$ copies of the oracles in RANDOMIZED HIDDEN SUBSET(α). One way to interpret this is that the randomization offers a substitute for expander graphs. An expander graph is sparse but well-mixing; a randomized oracle query instantaneously mixes across a graph's connected component. As a result, we can distinguish degree-2 graphs with this QMA protocol, even though they are not typically expander graphs:

Theorem 27. *For all* $0 < \alpha < \frac{1}{2}$*, there is a* QMA *protocol for* STANDARD RANDOMIZED HIDDEN SUBSET(α) *and* IN-PLACE RANDOMIZED HIDDEN SUBSET(α).

Proof. Perhaps surprisingly, the strategy in Theorem 19 also works here:

- Consider the graph G of any G-coded function presented in a YES instance. By definition, the vertices V are disconnected from all vertices in $[N]/V$. So the witness $|V\rangle$ is close to the second eigenvector $|\lambda_2\rangle$, and the failure probability is negligible.
- Consider the NO instance. Then f is chosen uniformly from the set $T_\varnothing$ of all permu-

tations of $[N]$. Then the spectral test fails with probability

$$\mathop{\mathbf{E}}_{\pi\in T_\varnothing}\left[\frac{d-\langle\psi|A_\pi|\psi\rangle}{2d}\right]\Bigg|_{d=2} = \frac{1}{2}-\frac{1}{4}\mathop{\mathbf{E}}_{\pi\in T_\varnothing}\left[\langle\psi|A_\pi|\psi\rangle\right] \tag{2.32}$$

$$= \frac{1}{2}-\frac{1}{4}\left(\frac{1}{N!}\sum_{\pi\in T_\varnothing}\langle\psi|A_\pi|\psi\rangle\right) \tag{2.33}$$

$$= \frac{1}{2}-\frac{1}{8}\left(\frac{1}{(N!)^2}\sum_{\pi_1,\pi_2\in T_\varnothing}\langle\psi|\left(A_{\pi_1}+A_{\pi_2}\right)|\psi\rangle\right). \tag{2.34}$$

The matrix $A_{\pi_1}+A_{\pi_2}$ determines the adjacency matrix of a random 4-regular graph in the configuration model; as a result,

$$\mathop{\mathbf{E}}_{\pi\in T_\varnothing}\left[\frac{d-\langle\psi|A_\pi|\psi\rangle}{2d}\right]\Bigg|_{d=2} = \mathop{\mathbf{E}}_{\pi_1,\pi_2\in T_\varnothing}\left[\frac{d-\langle\psi|A_{\pi_1,\pi_2}|\psi\rangle}{2d}\right]\Bigg|_{d=4}. \tag{2.35}$$

Since a random 4-regular graph has a constant spectral gap with probability $1-O(\frac{1}{\mathrm{poly}(N)})$ Friedman [2004], the failure probability is at least $\mathrm{FAIL}_{\mathrm{YES}}$ from Theorem 19, less $O(\frac{1}{2^{\mathrm{poly}(n)}})$. So the failure probability is $\Omega(1)$, just as before. □

2.6 Randomized oracles and symmetric subspaces

We first formalize how randomized oracles are orthogonal projectors. We defer the proofs to Section 2.10.

Definition 28 (Representation of a group)**.** *Consider a group G and a vector space V. A* representation *of G is a map R that sends each $g\in G$ to a linear operator $R(g):\mathsf{V}\to\mathsf{V}$ such that $R(g_1g_2)=R(g_1)\circ R(g_2)$ for all $g_1,g_2\in G$.*

Theorem 29 (Projecting onto the symmetric subspace [Harrow, 2013b, Proposition 2])**.** *Consider a finite group G, a vector space V, and a representation $R:G\to L(\mathsf{V})$. Then the*

operator

$$\Pi_R := \frac{1}{|G|} \sum_{g \in G} R(g) \tag{2.36}$$

is an orthogonal projector onto $\mathsf{V}^G \subseteq \mathsf{V}$*, where*

$$\mathsf{V}^G := \{v \in \mathsf{V} \,:\, R(g)[v] = v \,\forall g \in G\}\,. \tag{2.37}$$

Theorem 30 (Oracles on density matrices form a representation)**.** *Consider a group* G *of functions* $f : [N] \to [N]$ *with bitwise* $\oplus$ *as the group operation. Then the map* $f \mapsto \mathcal{U}_f$ *is a representation over the vector space of* $2N^2 \times 2N^2$ *complex matrices.*

Similarly, consider a group $\widetilde{G}$ *of permutations* $\pi : [N] \to [N]$ *with composition as the group operation. Then the map* $\pi \mapsto \widetilde{\mathcal{U}}_\pi$ *is a representation over the vector space of* $2N \times 2N$ *complex matrices.*

Theorem 31 (Some randomized oracles are orthogonal projectors)**.** *Consider a group* G *of functions* $f : [N] \to [N]$ *with bitwise* $\oplus$ *as the group operation. Then* $\mathcal{O}_G$ *is an orthogonal projector, under the Frobenius inner product* $(x|y) = \mathrm{Tr}\left[x^\dagger y\right]$ *for* $x, y \in \mathbb{C}^{2N^2 \times 2N^2}$*, onto*

$$\mathsf{V}_G := \{\rho \in \mathbb{C}^{2N^2 \times 2N^2} \,:\mathcal{U}_f[\rho] = \rho \,\forall f \in G\}\,. \tag{2.38}$$

Similarly, consider a group $\widetilde{G}$ *of permutations* $\pi : [N] \to [N]$ *with composition as the group operation. Then* $\widetilde{\mathcal{O}}_{\widetilde{G}}$ *is an orthogonal projector, under the Frobenius inner product* $(x|y) = \mathrm{Tr}\left[x^\dagger y\right]$ *for* $x, y \in \mathbb{C}^{2N \times 2N}$*, onto*

$$\widetilde{\mathsf{V}}_{\widetilde{G}} := \{\rho \in \mathbb{C}^{2N \times 2N} \,:\widetilde{\mathcal{U}}_\pi[\rho] = \rho \,\forall \pi \in \widetilde{G}\}\,. \tag{2.39}$$

In IN-PLACE RANDOMIZED HIDDEN SUBSET(α), a quantum verifier is either given $\widetilde{\mathcal{O}}_{T_\varnothing}$ (NO) or $\widetilde{\mathcal{O}}_{T_V}$ for some $V \subseteq [N]$ where $|V| = N^\alpha$ (YES). Since $T_V \subseteq T_\varnothing$, the symmetric

subspace according to $T_\varnothing$ is a subspace of that according to T_V, i.e. $\widetilde{\mathsf{V}}_{T_\varnothing} \subseteq \widetilde{\mathsf{V}}_{T_V}$. So we can exactly find the basis of the symmetric subspaces $\widetilde{\mathsf{V}}_{T_\varnothing}$ and $\widetilde{\mathsf{V}}_{T_V}$ (see Section 2.8 for details). This key property is used throughout Section 2.6.1.

2.6.1 In-place oracles: when classical witnesses are not enough

We interpret IN-PLACE RANDOMIZED HIDDEN SUBSET(α) as distinguishing the set of all permutations from a subgroup that stabilizes a small subset $V \subseteq [N]$. In Theorem 21, we prove that classical witnesses designed for the verifier to choose YES cannot help a quantum verifier efficiently decide this problem. This requires three main lemmas. First, we show in Lemma 33 that input states distinguishing a YES instance or NO instance must have knowledge of the hidden subset V (either as a subset state $|V\rangle$ or a mixed state $\mathbb{I}_V$). However, no density matrix can be close to too many subset states $|V\rangle$ (Lemma 34), and no POVM can choose the right answer for too many mixed states $\mathbb{I}_V$ (Lemma 35). We combine these facts in a hybrid argument; note that we must fix an algorithm by its unitaries *and* its POVM. We formally state the lemmas (deferring the proofs to Section 2.8 and Section 2.10), and then prove Theorem 21.

We use the following measure of "progress" for the hybrid argument:

Definition 32 (Difference of oracle queries)**.** *For any ρ, let $d_{V,\rho}$ be the difference of the two oracle queries*

$$d_{V,\rho} := \widetilde{\mathcal{O}}_{T_V}[\rho] - \widetilde{\mathcal{O}}_{T_\varnothing}[\rho]\,. \tag{2.40}$$

If the nuclear norm of $d_{V,\rho}$ is non-negligible, we say that ρ is a good distinguisher of $\widetilde{\mathcal{O}}_{T_V}$ and $\widetilde{\mathcal{O}}_{T_\varnothing}$. We show that every good distinguisher ρ has a certain form; the proof is deferred to Section 2.8.

Lemma 33 (Good distinguishers have a certain form)**.** *Consider a density matrix ρ and up to $O(\log(n))$ extra workspace qubits. Suppose $\|d_{V,\rho}\|_1 = \Omega(\frac{1}{\mathrm{poly}(n)})$. Then among the*

quantities

$$\langle V, z| \rho |V, z\rangle \ , \tag{2.41}$$

$$\mathrm{Tr}\left[\rho \left(\mathbb{I}_{V,z} - \frac{|V|}{N}\mathbb{I}_{[N],z}\right)\right], \tag{2.42}$$

for any $z \in \{\pm 1\}$*, at least one has magnitude* $\Omega(\frac{1}{\mathrm{poly}(n)})$.

We now state two lemmas about subsets and subset states. These help us prove that no quantum state can be a good distinguisher of too many YES instances. We defer the proofs to Section 2.10.

Lemma 34 (Can't approximate too many subset states)**.** *Consider a Hermitian* $N \times N$ *matrix* ρ *that is positive semidefinite and has trace at most* 1*. Consider the set of all subsets* $V \subseteq [N]$*, where* $|V| = N^\alpha$ *for a fixed* $0 < \alpha < \frac{1}{2}$*. Then the fraction of subsets* V *such that* $\langle V| \rho |V\rangle = \Omega(\frac{1}{\mathrm{poly}(n)})$ *decreases faster than any exponential in* $\mathrm{poly}(n)$.

Lemma 35 (Not too many subsets can have elevated mean)**.** *Consider any* $N \times N$ *POVM* $\{E, \mathbb{I} - E\}$*, and the set of all subsets* $V \subseteq [N]$*, where* $|V| = N^\alpha$ *for a fixed* $0 < \alpha < \frac{1}{2}$*. Then the fraction of subsets* V *where*

$$|f(V)| := \left|\frac{1}{|V|}\,\mathrm{Tr}[\mathbb{I}_V E] - \frac{1}{N}\,\mathrm{Tr}[E]\right| = \Omega(\frac{1}{\mathrm{poly}(n)})\,, \tag{2.43}$$

decreases faster than any exponential in $\mathrm{poly}(n)$.

Intuitively, Lemma 34 and Lemma 35 hold because subset states can approximate *any* quantum state well. Grilo, Kerenidis, and Sikora Grilo et al. [2015] show that for any n-qubit quantum state $|\psi\rangle$, there exists a subset state $|S\rangle$ such that $|\langle S|\psi\rangle| \geq \frac{1}{8\sqrt{n+3}}$.

We now prove the main statement:

Theorem 21. *No quantum verifier that entangles oracle queries with at most* $O(\log(n))$ *additional qubits can efficiently decide* IN PLACE RANDOMIZED HIDDEN SUBSET(α) *for*

any $0 < \alpha < \frac{1}{2}$*, even with a polynomial-length classical witness designed for the verifier to choose YES.*

Proof. Let the set of YES instances be $\mathcal{Y}$; note that each YES instance corresponds to a set $V \subseteq [N]$ where $|V| = N^\alpha$, for some fixed $0 < \alpha < \frac{1}{2}$.

Suppose for contradiction that there is a protocol for this problem at some $\alpha < \frac{1}{2}$. Then the verifier can distinguish $\widetilde{\mathcal{O}}_{T_\varnothing}$ from any $\widetilde{\mathcal{O}}_{T_V}$ in a polynomial number of queries using a classical witness of size $O(\text{poly}(n))$. By the pigeonhole principle, there must exist a set of YES instances $\mathcal{Y}'$ such that $|\mathcal{Y}'|/|\mathcal{Y}| = \Omega(\frac{1}{2^{\text{poly}(n)}})$, where the verifier can use the *same algorithm* to distinguish $\widetilde{\mathcal{O}}_{T_\varnothing}$ from *every* YES instance in $\mathcal{Y}'$.

We then construct a *hybrid argument* in the style of Bennett, Bernstein, Brassard, and Vazirani Bennett et al. [1997] and Ambainis Ambainis [2000], which interpolates from queries of one oracle to queries of another oracle. For simplicity we write the proof without extra workspace qubits; however, we can have up to $O(\log(n))$ extra workspace qubits to satisfy Lemma 33. Any polynomial query algorithm can be written as a set of unitaries $A = \{U^{(1)}, \ldots, U^{(k)}\}$ for some $k = O(\text{poly}(n))$ (alternating between unitary evolutions and oracle queries), and a POVM $\{E, \mathbb{I} - E\}$. Consider the following "hybrid" algorithms:

Definition 36. *Given any set of* k *unitaries* $A = \{U^{(1)}, \ldots, U^{(k)}\}$*, define the hybrid algorithm*

$$A_{V,\ell}\left[\rho_0\right] = \widetilde{\mathcal{O}}_{T_V}^{(k)} \circ \mathcal{U}^{(k)} \circ \cdots \circ \widetilde{\mathcal{O}}_{T_V}^{(\ell+1)} \circ \mathcal{U}^{(\ell)} \circ \widetilde{\mathcal{O}}_{T_\varnothing}^{(\ell)} \circ \mathcal{U}^{(\ell)} \circ \ldots \widetilde{\mathcal{O}}_{T_\varnothing}^{(1)} \circ \mathcal{U}^{(1)}\left[\rho_0\right], \tag{2.44}$$

which evolves ρ_0 *under the oracle* $\mathcal{O}_{T_\varnothing}$ *for* ℓ *steps and under* $\mathcal{O}_{T_V}$ *for the other* $k - \ell$ *steps.*

Then the following is true for each $\widetilde{\mathcal{O}}_{T_V} \in \mathcal{Y}'$:

$$\Omega(\frac{1}{\text{poly}(n)}) = \left|\text{Tr}\left[EA_{V,k}[\rho_0]\right] - \text{Tr}\left[EA_{V,0}[\rho_0]\right]\right| \tag{2.45}$$

$$\leq \sum_{i=0}^{k-1} \left|\text{Tr}\left[EA_{V,i+1}[\rho_0]\right] - \text{Tr}\left[EA_{V,i}[\rho_0]\right]\right|, \tag{2.46}$$

which implies

$$\Omega(\frac{1}{\text{poly}(n)}) = \sum_{i=0}^{k-1} \left|\text{Tr}\Big[E\left(\widetilde{\mathcal{O}}_{T_V}^{(k)} \circ \ldots \mathcal{U}^{(i)} \circ \left(\widetilde{\mathcal{O}}_{T_V}^{(i)} - \widetilde{\mathcal{O}}_{T_\varnothing}^{(i)}\right)[\rho^{(i)}]\right)\Big]\right| \tag{2.47}$$

$$= \sum_{i=0}^{k-1} \left|\text{Tr}\Big[E^{V,(i)} d_{V,\rho^{(i)}}\Big]\right| , \tag{2.48}$$

for the operator $E^{V,(i)}$ constructed by

$$E^{V,(i)} = \mathcal{U}^{\dagger(i)} \circ \widetilde{\mathcal{O}}_{T_V}^{(i)} \circ \cdots \circ \mathcal{U}^{\dagger(k)} \circ \widetilde{\mathcal{O}}_{T_V}^{(k)}[E] . \tag{2.49}$$

By Fact 53, the operators $E^{V,(i)}$ and $\mathbb{I} - E^{V,(i)}$ are also Hermitian and positive semidefinite, so $\{E^{V,(i)}, \mathbb{I} - E^{V,(i)}\}$ is a POVM.

Using the pigeonhole principle, there must be a step ℓ in the summation with magnitude $\Omega(\frac{1}{\text{poly}(n)})$. Each $\widetilde{\mathcal{O}}_{T_V} \in \mathcal{Y}'$ has such a step. Again by the pigeonhole principle, there is a ℓ^* and set $\mathcal{Y}^* \subseteq \mathcal{Y}'$ where

$$\left|\text{Tr}\Big[E^{V,(\ell^*)} d_{V,\rho^{(\ell^*)}}\Big]\right| = \Omega(\frac{1}{\text{poly}(n)}) , \tag{2.50}$$

and $|\mathcal{Y}^*|/|\mathcal{Y}'| \geq \frac{1}{k} = \Omega(\frac{1}{\text{poly}(n)})$. Notice that this implies $|\mathcal{Y}^*|/|\mathcal{Y}| = \Omega(\frac{1}{2^{\text{poly}(n)}})$.

Since the trace of M with a POVM operator is at most $\|M\|_1$ (Fact 50), we have for all $\widetilde{\mathcal{O}}_{T_V} \in \mathcal{Y}^*$,

$$\Omega(\frac{1}{\text{poly}(n)}) = \left|\text{Tr}\Big[E^{V,(\ell^*)} d_{V,\rho^{(\ell^*)}}\Big]\right| \leq \left\|d_{V,\rho^{(\ell^*)}}\right\|_1 . \tag{2.51}$$

When queries are entangled with at most $O(\log(n))$ qubits, the premise of Lemma 33 holds; then one of the quantities in the theorem statement must be large. However, Lemma 34 says that a given ρ can only satisfy either of the first two quantities for a smaller-than-exponential

fraction of $\mathcal{Y}$. So for most choices of $\widetilde{\mathcal{O}}_{T_V} \in \mathcal{Y}^*$,

$$\mathrm{Tr}\left[\rho^{(\ell^*)}\left(\mathbb{I}_{V,z} - \frac{|V|}{N}\mathbb{I}_{[N],z}\right)\right] = \Omega(\frac{1}{\mathrm{poly}(n)})\,. \tag{2.52}$$

for at least one of $z \in \{\pm 1\}$.

Inspecting the proof of Lemma 33, this implies $d_{V,\rho^{(\ell^*)}}$ can only have $\Omega(\frac{1}{\mathrm{poly}(n)})$ weight on $C_{4,z}$ for some $z \in \{\pm 1\}$ across all matrices in $\mathcal{C}$. In fact, for most choices of $\widetilde{\mathcal{O}}_{T_V} \in \mathcal{Y}^*$, we show that this is also true for

$$d_{V,\ell^*,j} := \widetilde{\mathcal{O}}_{T_V}^{(\ell^*+j)} \circ \mathcal{U}^{(\ell^*+j)} \circ \cdots \circ \widetilde{\mathcal{O}}_{T_V}^{(\ell^*+1)} \circ \mathcal{U}^{(\ell^*+1)} \circ d_{V,\rho^{(\ell^*)}}\,, \tag{2.53}$$

for all $0 \leq j \leq k - \ell^*$. We show this by induction. Note that by Fact 50 and the fact that $d_{V,\ell^*,k-\ell^*}$ is the difference of two objects with nuclear norm 1, $\|d_{V,\ell^*,k-\ell^*}\|_1 = \Omega(\frac{1}{\mathrm{poly}(n)}) = O(1)$.

Consider $d_{V,\ell^*,i}$ for some $1 \leq i \leq k - \ell^*$, which can be represented with the basis $\mathcal{C}$. By Fact 54, it has Frobenius norm at most $\|d_{V,\rho^{(\ell^*)}}\|_{Fr} = O(\frac{1}{\sqrt{|V|}})$. So it must have $o(\frac{1}{\mathrm{poly}(n)})$ weight on pure states. Inspecting the basis $\mathcal{C}$, this means $d_{V,\rho^{(\ell^*)}}$ can only have $\Omega(\frac{1}{\mathrm{poly}(n)})$ weight on $C_{4,z}$ or $\frac{1}{N}\mathbb{I}_{[N],z}$ for $z \in \{\pm 1\}$. By Fact 54, $d_{V,\rho^{(\ell^*)}}$ has nuclear norm at least $\|d_{V,\ell^*,k-\ell^*}\|_1 = \Omega(\frac{1}{\mathrm{poly}(n)})$, so it must have $\Omega(\frac{1}{\mathrm{poly}(n)})$ weight on at least one such matrix. Suppose for contradiction that the matrix is $\frac{1}{N}\mathbb{I}_{[N],z}$ for $z \in \{\pm 1\}$. Then

$$\Omega(\frac{1}{\mathrm{poly}(n)}) = \mathrm{Tr}\left[\mathbb{I}_{[N],z}\widetilde{\mathcal{O}}_{T_V}\left[\mathcal{U}^{(\ell^*+i)}\left[d_{V,\ell^*,i-1}\right]\right]\right] = \mathrm{Tr}\left[\left(\mathcal{U}^{(\ell^*+i)\dagger} \circ \widetilde{\mathcal{O}}_{T_V}^{\dagger}[\mathbb{I}_{[N],z}]\right) d_{V,\ell^*,i-1}\right]. \tag{2.54}$$

By the inductive hypothesis, $d_{V,\ell^*,i-1}$ only has $\Omega(\frac{1}{\mathrm{poly}(n)})$ weight on some $C_{4,z}$ for $z \in \{\pm 1\}$. Then for some $z' \in \{\pm 1\}$,

$$\Omega(\frac{1}{\mathrm{poly}(n)}) = \mathrm{Tr}\left[\left(\mathcal{U}^{(\ell^*+i)\dagger} \circ \widetilde{\mathcal{O}}_{T_V}^{\dagger}[\mathbb{I}_{[N],z}]\right)\left(\frac{1}{|V|}\mathbb{I}_{V,z'} - \frac{1}{N}\mathbb{I}_{[N],z'}\right)\right]. \tag{2.55}$$

Notice that for any unitary U, the object

$$\{\mathcal{U}^{(\ell^*+i)\dagger} \circ \widetilde{\mathcal{O}}^\dagger_{T_V}[\mathbb{I}_{[N],z=+1}], \mathcal{U}^{(\ell^*+i)\dagger} \circ \widetilde{\mathcal{O}}^\dagger_{T_V}[\mathbb{I}_{[N],z=-1}]\}$$

forms a POVM. By Lemma 35, this can only be satisfied at either $z \in \{\pm 1\}$ for a smaller-than-exponential fraction of choices of V. So for most choices of $\widetilde{\mathcal{O}}_{T_V} \in \mathcal{Y}^*$ (i.e. a $\Omega(\frac{1}{2^{\text{poly}(n)}})$ fraction of choices of V), $d_{V,\ell^*,i}$ has $\Omega(\frac{1}{\text{poly}(n)})$ weight on $C_{4,z}$ for at least one of $z \in \{\pm 1\}$, and for no other matrices in $\mathcal{C}$.

Since $\Omega(\frac{1}{\text{poly}(n)}) = \left|\text{Tr}\left[E d_{V,\ell^*,k-\ell^*}\right]\right|$, our supposition then implies that for one of $z \in \{\pm 1\}$,

$$\Omega(\frac{1}{\text{poly}(n)}) = \left|\text{Tr}\left[E C_{4,z}\right]\right| = \left|\text{Tr}\left[E\left(\frac{1}{|V|}\mathbb{I}_{V,z} - \frac{1}{N}\mathbb{I}_{[N],z}\right)\right] \pm O(\frac{1}{2^{\text{poly}(n)}})\right| . \tag{2.56}$$

But by Lemma 35, this can only be satisfied at either $z \in \{\pm 1\}$ for a smaller-than-exponential fraction of $\mathcal{Y}$. This is a contradiction. So there can be no efficient protocol for this problem.

□

2.6.2 Standard oracles: when classical witnesses are enough

As shown in Theorem 30, randomized standard oracles can also form a representation. But the preserved group structure is much different than for randomized in-place oracles. Consider the set $T_\varnothing$ of permutations on $[N]$. For any $f_1, f_2 \in T_\varnothing$, the element $f_1 f_2$ in this group structure acts for all $x \in [N]$ and $z \in \{\pm 1\}$ as

$$(f_1 f_2)^z(x) = f_1^z(x) \oplus f_2^z(x) . \tag{2.57}$$

Note that this operation is abelian; that is, $(f_1 f_2) = (f_2 f_1)$. Any finite abelian group can always be represented as the direct sum of cyclic groups. In fact, under this group operation,

$T_\varnothing$ can be decomposed by the input $x \in [N]$ and function inverter $z \in \{\pm 1\}$:

$$T_\varnothing = \bigoplus_{x \in [2^n], z \in \{\pm 1\}} \mathbb{Z}_{2^n} . \tag{2.58}$$

With this group operation, the only possible subgroups of $T_\varnothing$ have the form

$$\bigoplus_{x \in [2^n], z \in \{\pm 1\}} \mathbb{Z}_{2^{k_{x,z}}} , \tag{2.59}$$

for $0 \leq k_{x,z} \leq n$. As a result, there is a QCMA protocol to distinguish any strict subgroup of $T_\varnothing$ from $T_\varnothing$.

Theorem 37. *There is a one-query* QCMA *protocol for* STANDARD RANDOMIZED HIDDEN SUBGROUP$(\times, \{H_i\})$ *when the group operation* $\times$ *is bitwise XOR, for any valid* $\{H_i\}$.

Proof. Suppose the classical witness is a bitstring of length at least $n + 1$. The verifier can then:

1. Use the first n bits to construct x and the next bit to construct z.
2. Prepare the state $|0\rangle^{\otimes n} |x, z\rangle$.
3. Apply $\mathcal{O}_H$, creating the state $|f^z(x)\rangle |x, z\rangle$ for some $f \in H$.
4. Measure the first n qubits, and accept if the result is even.[3]

Consider a YES instance associated with a subgroup $H \subsetneq T_\varnothing$. Then H will have some $x \in [N], z \in \{\pm 1\}$ such that $k_{x,z} < n$. A witness can store x and z; since $k_{x,z} < n$, $f^z(x)$ will be even with probability 1.

In the NO instance, $H = T_\varnothing$. Then $f^z(x)$ is even with probability 0.5 for every $x \in [N], z \in \{\pm 1\}$. □

3. Depending on the encoding, one can simply measure the n^{th} qubit, and accept if the result is 0.

Note that Theorem 37 holds even if the randomized standard oracle $\mathcal{O}_F$ does not have access to the function inverse.

2.7 No witness is enough for phase oracles

Theorem 22. *No quantum verifier that entangles oracle queries with at most $o(n)$ additional qubits can efficiently decide* PHASE RANDOMIZED HIDDEN SUBSET(α) *for any $0 < \alpha < \frac{1}{2}$, even with* any *witness designed for the verifier to choose YES. Moreover, these verifiers require an exponential number of queries to statistically distinguish a YES instance from the NO instance, for* each *of asymptotically all YES instances.*

Proof. We now prove the query lower bound. Let k be the number of queries required to distinguish $\overline{\mathcal{O}}_{T_V}$ from $\overline{\mathcal{O}}_{T_\varnothing}$. Consider any algorithm that distinguishes the two instances, defined by a starting state ρ_0, k unitaries, k oracle queries, and a POVM $\{E, \mathbb{I} - E\}$. In the framework of hybrid algorithms (Definition 36),

$$\begin{aligned}
\Omega(\frac{1}{\text{poly}(n)}) &= \left|\text{Tr}\left[EA_{V,k}[\rho_0]\right] - \text{Tr}\left[EA_{V,0}[\rho_0]\right]\right| && (2.60)\\
&\leq \left\|A_{V,k}[\rho_0] - A_{V,0}[\rho_0]\right\|_1 && (2.61)\\
&\leq k \max_{i\in\{0,\ldots,k-1\}} \left\|A_{V,i+1}[\rho_0] - A_{V,i}[\rho_0]\right\|_1 && (2.62)\\
&\leq k \max_{i\in\{0,\ldots,k-1\}} \left\|\overline{\mathcal{O}}_{T_V}[\rho^{(i)}] - \overline{\mathcal{O}}_{T_\varnothing}[\rho^{(i)}]\right\|_1 , && (2.63)
\end{aligned}$$

where the last line follows because randomized oracles do not increase the nuclear norm (Fact 54).

We now bound $\left\|\overline{\mathcal{O}}_{T_V}[\rho] - \overline{\mathcal{O}}_{T_\varnothing}[\rho]\right\|_1$ for *any* ρ. Recall that a phase oracle $\overline{O}_F$ acts as

$$\overline{O}_F\left[|x_1, z_1\rangle\langle x_2, z_2|\right] = \frac{1}{|F|}\sum_{f\in F} \omega_N^{f^{z_1}(x_1) - f^{z_2}(x_2)} |x_1, z_1\rangle\langle x_2, z_2| , \tag{2.64}$$

for any $x_1, x_2 \in [N]$ and $z_1, z_2 \in \{\pm 1\}$. So every basis vector $|x_1, z_1\rangle\langle x_2, z_2|$ acquires a

coefficient c_{x_1,z_1,x_2,z_2}.

We start with $\overline{O}_{T_\varnothing}$ (the NO instance). When $(x_1, z_1) = (x_2, z_2)$, the coefficient is 1. When $x_1 \neq x_2$, $f^z(x_1)$ and $f^{-z}(x_2)$ are uniformly likely to be any value, so the coefficient is

$$\frac{1}{N^2} \sum_{a\in[N],b\in[N]} \omega_N^{a-b} = \frac{1}{N^2} \left\| \sum_{a\in[N]} \omega_N^{a} \right\|^2 = 0\,. \tag{2.65}$$

Similarly, when $x_1 \neq x_2$, $f^z(x_1)$ and $f^z(x_2)$ are uniformly likely to be any unequal values; the coefficient is

$$\frac{1}{N(N-1)} \sum_{a\in[N],b\in[N],a\neq b} \omega_N^{a-b} = \frac{1}{N(N-1)} \Big[\sum_{a\in[N],b\in[N]} \omega_N^{a-b} - \sum_{a\in[N]} \omega_N^{a-a} \Big] = -\frac{1}{N-1}\,. \tag{2.66}$$

We now consider $\overline{O}_{T_V}$ (a YES instance). When $(x_1, z_1) = (x_2, z_2)$, the coefficient is again 1. When $x_1 \neq x_2$, the values of $f^z(x_1)$ and $f^{-z}(x_2)$ are uniformly likely to be any value in $i_V(x_1)$ and $i_V(x_2)$, respectively, so the coefficient is

$$\frac{1}{|i_V(x_1)| \times |i_V(x_2)|} \sum_{a\in i_V(x_1),b\in i_V(x_2)} \omega_N^{a-b}\,. \tag{2.67}$$

Similarly, when $x_1 \neq x_2$, $f^z(x_1)$ and $f^z(x_2)$ are uniformly likely to be any unequal values in $i_V(x_1)$ and $i_V(x_2)$, respectively, so the coefficient is

$$\frac{1}{|i_V(x_1)| \times |i_V(x_2)|} \sum_{a\in i_V(x_1),b\in i_V(x_2),a\neq b} \omega_N^{a-b} = \frac{\left(\sum_{a\in i_V(x_1),b\in i_V(x_2)} \omega_N^{a-b}\right) - \delta|i_V(x_1)|}{|i_V(x_1)| \times |i_V(x_2) - \delta|}\,, \tag{2.68}$$

where δ is 1 if $i_V(x_1) = i_V(x_2)$ and 0 otherwise.

Consider the object $\overline{\mathcal{O}}_{T_V}[\rho] - \overline{\mathcal{O}}_{T_\varnothing}[\rho]$ as the sum of two matrices $A_V + B_V$. Let A_V contain the $(V, z) \times (V, z)$ submatrix for both $z \in \{\pm 1\}$, and B_V contain the rest of the entries. When

the oracle query is entangled with $o(n)$ additional qubits, A_V has rank $O(|V|)$, and B_V has rank $O(N)$.

Since the roots of unity sum to zero, $\sum_{a\in V}\omega_N^a = -\sum_{a\in[N]/V}\omega_N^a$ for any $V\subseteq[N]$. Because of this,

$$\left\|\sum_{a,b\in i_V(x)}\omega_N^{a-b}\right\| \leq \left\|\sum_{a\in i_V(x)}\omega_N^a\right\|^2 \leq \left\|\sum_{a\in V}\omega_N^a\right\|^2 = O(|V|^2)\,. \tag{2.69}$$

As a result, all coefficients in B_V are $O(\frac{1}{N^{1-\alpha}})$.

In Lemma 38, we show that for asymptotically all choices of V, all coefficients in A_V are $O(\frac{1}{N^{3\alpha/4}})$. This argument uses a Chernoff bound and a central limit argument on samples without replacement.

We bound the nuclear norm of $\overline{\mathcal{O}}_{T_V}[\rho] - \overline{\mathcal{O}}_{T_\varnothing}[\rho]$ with the rank and Frobenius norm of A_V and B_V (Fact 48):

$$\begin{aligned}
\left\|\overline{\mathcal{O}}_{T_V}[\rho] - \overline{\mathcal{O}}_{T_\varnothing}[\rho]\right\|_1 &\leq \|A_V\|_1 + \|B_V\|_1 && (2.70)\\
&= O(\sqrt{V})\|A_V\|_{Fr} + O(\sqrt{N})\|B_V\|_{Fr} && (2.71)\\
&\leq \left(O(\sqrt{V})O(N^{-3\alpha/4}) + O(\sqrt{N})O(N^{\alpha-1})\right)\|\rho\|_{Fr} && (2.72)\\
&= O(N^{-\alpha/4} + N^{\alpha-1/2})\,. && (2.73)
\end{aligned}$$

Thus, for most choices of V, distinguishing $\overline{\mathcal{O}}_{T_V}$ and $\overline{\mathcal{O}}_{T_\varnothing}$ requires

$$k = \Omega(\min(N^{\alpha/4}, N^{1/2-\alpha}))$$

queries. □

We now prove the Chernoff bound:

Lemma 38. *Fix any $0 < \alpha < \frac{1}{2}$, and consider all subsets $V\subseteq[N]$ such that $|V| = N^\alpha$.*

Then for all but a doubly exponentially small fraction of choices of V,

$$\left\| \frac{1}{N^{2\alpha}} \sum_{a,b \in V} \omega_N^{a-b} \right\| = O(\frac{1}{N^{3\alpha/4}}) \,. \tag{2.74}$$

Proof. Consider the distribution $X = \{\omega_N^k\}$ where k is chosen uniformly from N. Both $Re(X)$ and $Im(X)$ have mean zero and variance at most 1.

Take a size-N^α sample from the distribution X, *without replacement*. Denote Y as the distribution of the sample mean. Both $Re(Y)$ and $Im(Y)$ have expectation $Re(X) = Im(X) = 0$, and variance

$$\frac{\sigma_X^2}{N^\alpha}(1 - \frac{N^\alpha - 1}{N-1}) \leq \frac{1}{N^\alpha} \,. \tag{2.75}$$

Even when sampling without replacement, Y is asymptotically normally distributed Erdos [1959]. So its moment generating function is

$$\mathrm{MGF}_Y[t] = e^{t\mu_Y + \sigma_Y^2 t^2/2} \leq e^{t^2/N^\alpha} \,. \tag{2.76}$$

We use a Chernoff bound to estimate when Y has magnitude at least $N^{-3\alpha/8}$. Notice that

$$\mathbf{Pr}[Y \geq a] = \mathbf{Pr}[e^{tY} \geq e^{ta}] \leq e^{-at}\mathrm{MGF}_Y[t] \leq e^{t^2/N^\alpha} e^{-at} \,, \tag{2.77}$$

so

$$\mathbf{Pr}[Y \geq \frac{0.5}{N^{3\alpha/8}}] \leq \inf_{t \geq 0} \exp\left(\frac{t^2}{N^\alpha} - \frac{0.5t}{N^{3\alpha/8}}\right) \underset{t=2N^{\alpha/2}}{\leq} \exp\left(4 - N^{\alpha/8}\right) = O(\frac{1}{\exp(\exp(n))}) \,. \tag{2.78}$$

This implies that Y has magnitude at most $N^{-3\alpha/8}$ (and Y^2 at most $N^{-3\alpha/4}$) except in a doubly exponentially small fraction of choices of V. □

Remark 39. *Consider an oracle that sends* $|c, x, z\rangle \to \omega_N^{c \cdot f^z(x)} |c, x, z\rangle$*, where the* c *register has* k *qubits. Note that Theorem 22 applies whenever* $k = o(n)$*. However, there must be a phase transition, since at* $k = n$*, this oracle is unitarily equivalent to a standard oracle, and thus has a* QMA *protocol for* RANDOMIZED HIDDEN SUBSET(α) *in Section 2.5.*

2.8 Basis elements of symmetric subspaces

In this section we find an orthogonal basis for $\widetilde{\mathsf{V}}_{T_\varnothing}$ and $\widetilde{\mathsf{V}}_{T_V}$, and prove Lemma 33.

Theorem 40 (Basis elements of $\widetilde{\mathsf{V}}_{T_\varnothing}$)**.** *Consider* $T_\varnothing$*, the set of all permutations of* $[N]$*. Then* $\widetilde{\mathsf{V}}_{T_\varnothing}$ *is spanned by the matrices*

$$A_{z_1,z_2} = |+\rangle^{\otimes n} \langle +|^{\otimes n} \otimes |z_1\rangle \langle z_2| \,, \tag{2.79}$$

$$B'_{z_1} = \mathbb{I}_{[N]} \otimes |z_1\rangle \langle z_1| \,, \tag{2.80}$$

for $z_1, z_2 \in \{\pm 1\}$*.*

Proof. By definition, $\widetilde{\mathsf{V}}_{T_\varnothing}$ contains the set of matrices M such that $U_\pi M U_\pi^\dagger = M$ for all permutations $\pi : [N] \to [N]$. We can always write M in the following form:

$$M := \sum_{x_1,x_2 \in [N], z_1,z_2 \in \{\pm 1\}} \alpha_{x_1,z_1,x_2,z_2} |x_1, z_1\rangle \langle x_2, z_2| \,. \tag{2.81}$$

Then $U_\pi M U_\pi^\dagger$ has the form

$$U_\pi M U_\pi^\dagger = \sum_{x_1,x_2 \in [N], z_1,z_2 \in \{\pm 1\}} \alpha_{x_1,z_1,x_2,z_2} |\pi^{z_1}(x_1), z_1\rangle \langle \pi^{z_2}(x_2), z_2| \tag{2.82}$$

$$= \sum_{x_1,x_2 \in [N], z_1,z_2 \in \{\pm 1\}} \alpha_{\pi^{-z_1}(x_1),z_1,\pi^{-z_2}(x_2),z_2} |x_1, z_1\rangle \langle x_2, z_2| \,. \tag{2.83}$$

Thus, for all permutations π, and for all $x_1, x_2 \in [N]$ and $z_1, z_2 \in \{\pm 1\}$,

$$\alpha_{x_1,z_1,x_2,z_2} = \alpha_{\pi^{-z_1}(x_1),z_1,\pi^{-z_2}(x_2),z_2} \,. \tag{2.84}$$

Given (x_1, z_1, x_2, z_2), this restriction depends on the possible values of $(\pi^{-z_1}(x_1), \pi^{-z_2}(x_2))$:

- When $z_1 \neq z_2$, this can take on any value $(j, k) \in [N] \times [N]$.
- When $z_1 = z_2$ and $x_1 \neq x_2$, this can take on any value $(j, k) \in [N] \times [N]$ such that $j \neq k$.
- When $z_1 = z_2$ and $x_1 = x_2$, this can take on any value $(j, j) \in [N] \times [N]$.

As a result, for a given choice of $z_1, z_2 \in \{\pm 1\}$, all coefficients are equal except possibly on the diagonal when $z_1 = z_2$. There are then six basis elements, one proportional to $\mathbb{1}\!\!\!\!-_N \otimes |z_1\rangle\langle z_2| \propto |+\rangle^{\otimes n}\langle +|^{\otimes n} \otimes |z_1\rangle\langle z_2|$ for each $z_1, z_2 \in \{\pm 1\}$, and one proportional to $\mathbb{I}_{[N]} \otimes |z\rangle\langle z|$ for each $z \in \{\pm 1\}$. □

Theorem 41 (Basis elements of $\widetilde{\mathsf{V}}_{T_V}$)**.** *Consider T_V for some non-empty $V \subsetneq [N]$. Then $\widetilde{\mathsf{V}}_{T_V}$ is spanned by the matrices*

$$C'_{S_1,S_2,z_1,z_2} = |S_1\rangle\langle S_2| \otimes |z_1\rangle\langle z_2| \,, \tag{2.85}$$

$$D_{S_1,z_1} = \mathbb{I}_{S_1} \otimes |z_1\rangle\langle z_1| \,, \tag{2.86}$$

for $S_1, S_2 \in \{V, [N]/V\}$ and $z_1, z_2 \in \{\pm 1\}$.

Proof. The proof is similar to Theorem 40. By definition, $\widetilde{\mathsf{V}}_{T_V}$ contains the set of matrices M such that $U_\pi M U_\pi^\dagger = M$ for all permutations $\pi \in T_V$. We can always write M in the following form:

$$M := \sum_{x_1,x_2 \in [N], z_1,z_2 \in \{\pm 1\}} \alpha_{x_1,z_1,x_2,z_2} |x_1, z_1\rangle\langle x_2, z_2| \,. \tag{2.87}$$

Then for all $\pi \in T_V$, $x_1, x_2 \in [N]$ and $z_1, z_2 \in \{\pm 1\}$,

$$\alpha_{x_1,z_1,x_2,z_2} = \alpha_{\pi^{-z_1}(x_1),z_1,\pi^{-z_2}(x_2),z_2} \,. \tag{2.88}$$

Given (x_1, z_1, x_2, z_2), this restriction depends on the possible values of $(\pi^{-z_1}(x_1), \pi^{-z_2}(x_2))$:

- When $z_1 \neq z_2$ or $i_V(x_1) \neq i_V(x_2)$, $\alpha_{x_1,z_1,x_2,z_2} = \alpha_{j,z_1,k,z_2}$ whenever $j \in i_V(x_1)$ and $k \in i_V(x_2)$.
- When $z = z_1 = z_2$ and $i_V(x_1) = i_V(x_2)$, $\alpha_{x_1,z,x_2,z} = \alpha_{j,z,k,z}$ whenever $j, k \in i_V(x_1)$ and $j \neq k$.
- When $z = z_1 = z_2$ and $x = x_1 = x_2$, $\alpha_{x,z,x,z} = \alpha_{j,z,j,z}$ whenever $j \in i_V(z)$.

As a result, for a given choice of $z_1, z_2 \in \{\pm 1\}$ and $i_V(x_1), i_V(x_2) \in \{V, [N]/V\}$, all coefficients are equal except possibly on the diagonal when $z_1 = z_2$ and $i_V(x_1) = i_V(x_2)$. There are then twenty basis elements; one proportional to $|i_V(x_1)\rangle \langle i_V(x_2)| \otimes |z_1\rangle \langle z_2|$ for each $i_V(x_1), i_V(x_2) \in \{V, [N]/V\}$ and $z_1, z_2 \in \{\pm 1\}$, and one proportional to $\mathbb{I}_{i_V(x)} \otimes |z\rangle \langle z|$ for each $i_V(x) \in \{V, [N]/V\}$ and $z \in \{\pm 1\}$. □

We can now find an orthogonal basis for $\widetilde{\mathsf{V}}_{T_\varnothing}$ and $\widetilde{\mathsf{V}}_{T_V}$:

Theorem 42 (Orthogonal basis elements of $\widetilde{\mathsf{V}}_{T_\varnothing}$ and $\widetilde{\mathsf{V}}_{T_V}$)**.** *The symmetric subspace according to $T_\varnothing$ is spanned by the orthogonal matrices*

$$A_{z_1,z_2} = |+\rangle^{\otimes n} \langle +|^{\otimes n} \otimes |z_1\rangle \langle z_2| \,, \tag{2.89}$$

$$B_{z_1} = \frac{1}{N} \left(\mathbb{I}_{[N]} - |+\rangle^{\otimes n} \langle +|^{\otimes n} \right) \otimes |z_1\rangle \langle z_1| \,, \tag{2.90}$$

for $z_1, z_2 \in \{\pm 1\}$.

Let $\delta_{i,j}$ be the Kronecker delta function. The symmetric subspace according to T_V for any nonempty $V \subsetneq [N]$ is spanned by the matrices above, and the orthogonal matrices in $\mathcal{C}$,

which are

$$C_{1,z_1,z_2} = (|V\rangle\langle[N]/V| - |[N]/V\rangle\langle V|) \otimes |z_1\rangle\langle z_2| \,, \tag{2.91}$$

$$C_{2,z_1,z_2} = \Bigg(|V\rangle\langle[N]/V| + |[N]/V\rangle\langle V| \tag{2.92}$$

$$- \frac{2}{N\sqrt{|V|(N-|V|)}} \left(|+\rangle^{\otimes n}\langle+|^{\otimes n} - \delta_{z_1,z_2}\frac{1}{N}\mathbb{I}_{[N]} \right) \Bigg) \otimes |z_1\rangle\langle z_2| \,, \tag{2.93}$$

$$C_{3,z_1,z_2} = \Bigg(|V\rangle\langle V| - \frac{|V|}{N-|V|}|[N]/V\rangle\langle[N]/V| \tag{2.94}$$

$$- \delta_{z_1,z_2}\frac{1-\frac{|V|}{N-|V|}}{N}\mathbb{I}_{[N]} \Bigg) \otimes |z_1\rangle\langle z_2| \,, \tag{2.95}$$

$$\tag{2.96}$$

$$C_{4,z_1} = \left(\frac{1}{|V|}(\mathbb{I}_V - |V\rangle\langle V|) - \frac{1}{N-|V|}(\mathbb{I}_{[N]/V} - |[N]/V\rangle\langle[N]/V|) \right) \otimes |z_1\rangle\langle z_1| \,, \tag{2.97}$$

for $z_1, z_2 \in \{\pm 1\}$*. Moreover, these matrices have Frobenius norm at most* $O(1)$ *and nuclear norm* $\Theta(1)$*.*

Proof. All matrices are orthogonal when the last qubit $|z_1\rangle\langle z_2|$ doesn't match. We first consider the basis elements of $\widetilde{\mathsf{V}}_{T_\varnothing}$, as listed in Theorem 40. When $z_1 \neq z_2$, A_{z_1,z_2} is the only basis vector; when $z = z_1 = z_2$, observe that B_z is a linear combination of $A_{z,z}$ and B'_z, and

$$\mathrm{Tr}\left[A^\dagger_{z,z}B_z\right] = \frac{1}{N}\left(\langle+|^{\otimes n}\,\mathbb{I}_{[N]}\,|+\rangle^{\otimes n} - \langle+|^{\otimes n}\,|+\rangle^{\otimes n}\,\langle+|^{\otimes n}\,|+\rangle^{\otimes n}\right) = 0\,. \tag{2.98}$$

We now construct the basis of $\widetilde{\mathsf{V}}_{T_V}/\widetilde{\mathsf{V}}_{T_\varnothing}$. When $z_1 \neq z_2$, the only vectors are linear combinations of C'_{S_1,S_2,z_1,z_2} for $S_1, S_2 \in \{V, [N]/V\}$ and $z_1, z_2 \in \{\pm 1\}$. Note that $\langle+|^{\otimes n}\,|V\rangle = \frac{1}{\sqrt{|V|N}}$.

- First of all, C_{1,z_1,z_2} is antisymmetric; for every state $|\psi\rangle$, $\langle\psi|\,C_{1,z_1,z_2}\,|\psi\rangle = 0$.
- C_{2,z_1,z_2} is constructed as the *symmetric* version of C_{1,z_1,z_2}, and then offset by $|+\rangle\langle+|^{\otimes n}$

to account for overlap with A_{z_1,z_2}.

- C_{3,z_1,z_2} is immediately orthogonal C_{1,z_1,z_2} and C_{2,z_1,z_2}, and inspection shows orthogonality with A_{z_1,z_2}.

When $z = z_1 = z_2$, we modify $C_{1,z,z}, C_{2,z,z}, C_{3,z,z}$ to be traceless (i.e. orthogonal to $\mathbb{I}_N \otimes |z\rangle\langle z|$). The last matrix $C_{4,z}$ must include $\mathbb{I}_V$; it is immediately orthogonal to $C_{1,z,z}, C_{2,z,z}$, and inspection shows that it is traceless and orthogonal to any $|S\rangle\langle S| \otimes |z\rangle\langle z|$ for $S \in \{V, [N]/V\}$.

We now calculate the norms of each vector. The Frobenius norm and nuclear norm of an outer product $|\psi\rangle\langle\psi|$ is exactly 1. The nuclear norm of $\frac{1}{N}\mathbb{I}_{[N]}$ is 1, and the Frobenius norm is $\frac{1}{\sqrt{N}}$. Using Cauchy-Schwarz, the Frobenius norm and nuclear norm of A_{z_1,z_2}, C_{1,z_1,z_2}, C_{2,z_1,z_2}, C_{3,z_1,z_2} are all $\Theta(1)$. The nuclear norm of B_z and $C_{4,z}$ are both $\Theta(1)$, and their Frobenius norms are $\Theta(\frac{1}{\sqrt{N}})$ and $\Theta(\frac{1}{\sqrt{|V|}})$, respectively. □

We use the orthogonal basis in Theorem 42 to describe the form of good distinguishers of $\widetilde{\mathcal{O}}_{T_V}$ and $\widetilde{\mathcal{O}}_{T_\varnothing}$.

Fact 43 (Representing the difference of randomized oracles)**.** *Since $\widetilde{\mathcal{O}}_{T_V}$ and $\widetilde{\mathcal{O}}_{T_\varnothing}$ are both orthogonal projectors under the Frobenius inner product, the difference of their outputs can be represented with the basis elements $\mathcal{C}$ of $\widetilde{\mathsf{V}}_{T_V}/\widetilde{\mathsf{V}}_{T_\varnothing}$ (computed in Section 2.8). That is,*

$$d_{V,\rho} := \widetilde{\mathcal{O}}_{T_V}[\rho] - \widetilde{\mathcal{O}}_{T_\varnothing}[\rho] = \sum_{M \in \mathcal{C}} c_M M\,, \tag{2.99}$$

where

$$c_M = \frac{\mathrm{Tr}\Big[M^\dagger \rho\Big]}{\|M\|_{Fr}^2}\,. \tag{2.100}$$

We refer to c_M as the weight *(of ρ) on the matrix M.*

Lemma 44 (Lemma 33, restated)**.** *Consider a density matrix ρ and up to $O(\log(n))$ extra*

workspace qubits. Suppose $\|d_{V,\rho}\|_1 = \Omega(\frac{1}{\text{poly}(n)})$. *Then among the quantities*

$$\langle V, z| \rho |V, z\rangle \,, \tag{2.101}$$

$$\text{Tr}\left[\rho \left(\mathbb{I}_{V,z} - \frac{|V|}{N}\mathbb{I}_{[N],z}\right)\right], \tag{2.102}$$

for any $z \in \{\pm 1\}$, *at least one has magnitude* $\Omega(\frac{1}{\text{poly}(n)})$.

Proof. By Fact 43, $d_{V,\rho}$ can be written as a linear combination of basis elements in $\mathcal{C}$, i.e.

$$d_{V,\rho} = \sum_{M\in\mathcal{C}} c_M M = \sum_{M\in\mathcal{C}} \frac{\text{Tr}\left[M^\dagger \rho\right] M}{\|M\|_{Fr}^2} \,. \tag{2.103}$$

By the pigeonhole principle and the triangle inequality, there is an $M \in \mathcal{C}$ where

$$\|c_M M\|_1 \geq \frac{1}{|\mathcal{C}|}\Omega(\frac{1}{\text{poly}(n)}) = \Omega(\frac{1}{\text{poly}(n)}) \,, \tag{2.104}$$

Note that the last equality holds when $|\mathcal{C}| = O(\text{poly}(n))$; this is true allowing up to $O(\log(n))$ extra workspace qubits. Inspecting each condition:

- Suppose the matrix M is C_{1,z_1,z_2} or C_{2,z_1,z_2} for some $z_1, z_2 \in \{\pm 1\}$. By the proof of Theorem 42, the Frobenius norm and nuclear norm are both $\Theta(1)$, so

$$\Omega(\frac{1}{\text{poly}(n)}) = \text{Tr}\left[C^\dagger_{1.5\pm 0.5,z_1,z_2}\rho\right] \tag{2.105}$$

$$= \langle [N]/V, z_2| \rho |V, z_1\rangle \mp \langle V, z_2| \rho |[N]/V, z_1\rangle \pm O(\frac{1}{2^{\text{poly}(n)}}) \,. \tag{2.106}$$

  By pigeonhole, this implies that there is some $z_1, z_2 \in \{\pm 1\}$ such that

$$|\langle [N]/V, z_2| \rho |V, z_1\rangle| = \Omega(\frac{1}{\text{poly}(n)}),$$

  and by Fact 55, $\langle V, z_1| \rho |V, z_1\rangle = \Omega(\frac{1}{\text{poly}(n)})$.

- Suppose the matrix M is C_{3,z_1,z_2} for some $z_1, z_2 \in \{\pm 1\}$. By the proof of Theorem 42, the Frobenius norm and nuclear norm are both $\Theta(1)$, so

$$\Omega(\frac{1}{\text{poly}(n)}) = \text{Tr}\left[C^\dagger_{3,z_1,z_2}\rho\right] = \langle V, z_2|\,\rho\,|V, z_1\rangle \pm O(\frac{1}{2^{\text{poly}(n)}})\,, \tag{2.107}$$

  and by Fact 55, $\langle V, z_1|\,\rho\,|V, z_1\rangle = \Omega(\frac{1}{\text{poly}(n)})$.

- Suppose the matrix M is $C_{4,z}$ for some $z \in \{\pm 1\}$. By the proof of Theorem 42, the Frobenius norm is $\Theta(\frac{1}{\sqrt{|V|}})$ and the nuclear norm is $\Theta(1)$, so

$$\Omega(\frac{1}{\text{poly}(n)}) = |V|\,\text{Tr}\left[C^\dagger_{4,z}\rho\right] \tag{2.108}$$

$$= \text{Tr}\left[\rho\left(\mathbb{I}_{V,z} - \frac{|V|}{N}\mathbb{I}_{[N]/V,z}\right)\right] - \langle V, z|\,\rho\,|V, z\rangle \pm O(\frac{1}{2^{\text{poly}(n)}})\,. \tag{2.109}$$

  By pigeonhole, this implies that at least one of the two terms has magnitude $\Omega(\frac{1}{\text{poly}(n)})$.

This proof holds even if we allow controlled access to the oracle. Consider the addition of a control qubit $|a\rangle$, such that if $a = 0$ the oracle has no effect, and if $a = 1$ the oracle acts as usual. We can separate $d_{V,\rho}$ into four matrices $d^{(a_1,a_2)}_{V,\rho}$ based on the values (a_1, a_2).

If $\|d_{V,\rho}\|_1 = \Omega(\frac{1}{\text{poly}(n)})$, one of the four matrices $d^{(a_1,a_2)}_{V,\rho}$ must have $\Omega(\frac{1}{\text{poly}(n)})$ nuclear norm. We have already seen what happens if that matrix is $d^{(1,1)}_{V,\rho}$. Recall that the oracle has no effect when $a_1 = a_2 = 0$. Every matrix M with control register $|0\rangle\langle 0|$ satisfies $U_\pi M U^\dagger_\pi = M$ for all permutations π, and so $M \in \widetilde{\mathsf{V}}_{T_\varnothing}$. As a result, the matrix $d^{(0,0)}_{V,\rho} \in \widetilde{\mathsf{V}}_{T_V}/\widetilde{\mathsf{V}}_{T_\varnothing}$ is identically zero.

Suppose the matrix $d^{(1,0)}_{V,\rho}$ has $\Omega(\frac{1}{\text{poly}(n)})$ nuclear norm. (The proof for $d^{(0,1)}_{V,\rho}$ is nearly identical.) Despite an exponentially large basis of matrices of this form, we show in Lemma 45 that this matrix has small rank. In particular, up to $O(\frac{1}{2^{\text{poly}(n)}})$ multiplicative error, $d^{(1,0)}_{V,\rho}$ is a weighted sum of outer products $P_{z=+1} = |1, V, +1\rangle\langle\psi_{z=+1}|$ and $P_{z=-1} = |1, V, -1\rangle\langle\psi_{z=-1}|$ for some normalized states $|\psi_z\rangle$. These outer products have nuclear norm 1, so by the pigeonhole principle, there is a $z \in \{\pm 1\}$ such that the weight c_{P_z} on P_z has magnitude at

least $O(\frac{1}{\text{poly}(n)})$. Since the Frobenius norm of each outer product is also 1, the weight is $c_{P_z} = \text{Tr}\left[P_z^\dagger \rho\right] = \langle 1, V, z| \rho |\psi_z\rangle$. Thus, by Fact 55, $\langle 1, V, z| \rho |1, V, z\rangle = \Omega(\frac{1}{\text{poly}(n)})$ for some $z \in \{\pm 1\}$. □

We now prove the lemma:

Lemma 45. *Consider any matrix $A \in \widetilde{\mathsf{V}}_{T_V}/\widetilde{\mathsf{V}}_{T_\varnothing}$ where the control register is*

$$|a_1 = 1\rangle \langle a_2 = 0| \text{ or } |a_1 = 0\rangle \langle a_2 = 1| .$$

Up to $O(\frac{1}{2^{\text{poly}(n)}})$ multiplicative error, A is a linear combination of outer products

$$|1, V, +1\rangle \langle \psi_{z=+1}| \text{ and } |1, V, -1\rangle \langle \psi_{z=-1}|$$

or their conjugate transposes for some states $|\psi_{z=+1}\rangle$ and $|\psi_{z=-1}\rangle$.

Proof. We set $a_1 = 1$ and $a_2 = 0$; the proof in the reverse case is nearly identical. Consider any matrix with the form

$$M := \sum_{x_1, x_2 \in [N], z_1, z_2 \in \{\pm 1\}} \alpha_{x_1, z_1, x_2, z_2} |1, x_1, z_1\rangle \langle 0, x_2, z_2| . \tag{2.110}$$

For any permutation π, the matrix $U_\pi M U_\pi^\dagger$ is

$$U_\pi M U_\pi^\dagger = \sum_{x_1, x_2 \in [N], z_1, z_2 \in \{\pm 1\}} \alpha_{x_1, z_1, x_2, z_2} |1, \pi^{z_1}(x_1), z_1\rangle \langle 0, x_2, z_2| \tag{2.111}$$

$$= \sum_{x_1, x_2 \in [N], z_1, z_2 \in \{\pm 1\}} \alpha_{\pi^{-z_1}(x_1), z_1, x_2, z_2} |1, x_1, z_1\rangle \langle 0, x_2, z_2| . \tag{2.112}$$

For every matrix $M \in \widetilde{\mathsf{V}}_{T_\varnothing}$ and permutation π, $M = U_\pi M U_\pi^\dagger$. Thus, the coefficients do not

depend on x_1: every $M \in \widetilde{\mathsf{V}}_{T_\varnothing}$ has the form

$$M = \sum_{x_2 \in [N], z_1, z_2 \in \{\pm 1\}} \alpha_{z_1, x_2, z_2} \left|1, +^{\otimes n}, z_1\right\rangle \left\langle 0, x_2, z_2\right| = \sum_{z \in \{\pm 1\}} \alpha_z \left|1, +^{\otimes n}, z\right\rangle \left\langle \chi_z\right| . \tag{2.113}$$

Similarly, any matrix in $\widetilde{\mathsf{V}}_{T_V}$ satisfies $M = U_\pi M U_\pi^\dagger$ when π stabilizes the subset V. So if $M \in \widetilde{\mathsf{V}}_{T_V}$,

$$M = \sum_{S \in \{V, [N]/V\}, x_2 \in [N], z_1, z_2 \in \{\pm 1\}} \alpha_{S, z_1, x_2, z_2} \left|1, S, z_1\right\rangle \left\langle 0, x_2, z_2\right| \tag{2.114}$$

$$= \sum_{S \in \{V, [N]/V\}, z \in \{\pm 1\}} \alpha_{S,z} \left|1, S, z\right\rangle \left\langle \chi_z\right| . \tag{2.115}$$

Any matrix in $\widetilde{\mathsf{V}}_{T_V}/\widetilde{\mathsf{V}}_{T_\varnothing}$ must be orthogonal to all matrices in $\widetilde{\mathsf{V}}_{T_\varnothing}$. Thus, if $M \in \widetilde{\mathsf{V}}_{T_V}/\widetilde{\mathsf{V}}_{T_\varnothing}$,

$$M = \sum_{z \in \{\pm 1\}} \alpha_z \left|1\right\rangle \left|V'\right\rangle \left|z\right\rangle \left\langle \psi_z\right| , \tag{2.116}$$

where

$$\left|V'\right\rangle = \sqrt{\frac{N - |V|}{N}} \left|V\right\rangle + \sqrt{\frac{|V|}{N}} \left|[N]/V\right\rangle . \tag{2.117}$$

Note that $\left|V'\right\rangle$ is equal to $|V\rangle$ up to $O(\sqrt{\frac{|V|}{N}}) = O(\frac{1}{2^{\mathrm{poly}(n)}})$ multiplicative error. □

2.9 Norms and inner products

Note that we work with arbitrary matrices, not just positive semidefinite ones.

Definition 46 (Nuclear norm of a matrix)**.** *The nuclear norm of a matrix M is the sum of*

its singular values; that is,

$$\|M\|_1 = \sum_i \sigma_i(M) = \mathrm{Tr}\left[\sqrt{M^\dagger M}\right]. \tag{2.118}$$

Definition 47 (Frobenius norm and inner product of a matrix)**.** *The Frobenius inner product of $N \times N$ matrices A, B is*

$$(A|B) = \mathrm{Tr}\left[A^\dagger B\right] \tag{2.119}$$

This induces a norm, which is the square root of the sum of squares of the singular values:

$$\|A\|_{Fr} = \sqrt{\sum_i \sigma_i(A)^2} = \sqrt{\sum_{ij\in[N]} |A_{ij}|^2}. \tag{2.120}$$

Fact 48. *The nuclear norm of a matrix is at most the product of its Frobenius norm and the square root of its rank.*

Proof. See Rennie Rennie [2005] for a proof with explanation. □

Fact 49 (Nuclear norm of a positive semidefinite matrix)**.** *The nuclear norm of a positive semidefinite Hermitian matrix is simply its trace; that is, if ρ is Hermitian and positive semidefinite, then*

$$\|\rho\|_1 = \mathrm{Tr}[\rho]. \tag{2.121}$$

Proof. For a Hermitian and positive semidefinite matrix, the eigenvalues are all real and nonnegative, so the singular values are exactly the eigenvalues. Alternatively, notice that $\rho = \sqrt{\rho^\dagger \rho}$ and use Definition 46. □

Fact 50 (POVM trace is at most the nuclear norm)**.** *Consider any Hermitian matrix M*

and a POVM $\{E, \mathbb{I} - E\}$*. Then*

$$\mathrm{Tr}[EM] \leq \|M\|_1 \,. \tag{2.122}$$

Proof. Consider the singular value decomposition of a Hermitian $M = UDU^\dagger$. Then

$$\mathrm{Tr}[EM] = \mathrm{Tr}\Big[(U^\dagger EU)D\Big] = \mathrm{Tr}\big[E'D\big] \tag{2.123}$$

for some matrix E'. Note that $\{E', \mathbb{I} - E'\}$ make a POVM; they have the same eigenvalues of $\{E, \mathbb{I} - E\}$, respectively, and so are both positive semidefinite. Recall that the diagonal elements of a POVM are all nonnegative and at most 1. Then

$$\mathrm{Tr}[EM] = \mathrm{Tr}\big[E'D\big] = \sum_i E'_{ii} D_i \leq \sum_i |D_i| = \|D\|_1 = \|M\|_1 \,. \tag{2.124}$$

□

Fact 51 (Trace of outer product is inner product)**.** *Consider vectors* $|x\rangle, |y\rangle \in \mathbb{C}^m$ *and a matrix* $A \in \mathbb{C}^{m \times m}$*. Then the inner product of* $|y\rangle \langle x|$ *and* A *is*

$$\mathrm{Tr}\Big[(|y\rangle \langle x|)^\dagger A\Big] = \mathrm{Tr}[A |x\rangle \langle y|] = \langle y| A |x\rangle \,. \tag{2.125}$$

Proof.

$$\mathrm{Tr}[A |x\rangle \langle y|] = \mathrm{Tr}\left[\sum_{i,k \in [m]} \left(\sum_{j \in [m]} A_{ij} x_j\right)_i y_k^\dagger\right] = \sum_{k,j \in [m]} A_{kj} x_j y_k^\dagger = \langle y| A |x\rangle \,. \tag{2.126}$$

□

Remark 52 (Orthogonal basis for an input density matrix)**.** *We can decompose* ρ *into a*

basis $\mathcal{M}$ that is orthogonal under the Frobenius inner product $(a|b) = \mathrm{Tr}\Big[a^\dagger b\Big]$:

$$\rho = \sum_{M \in \mathcal{M}} c_M M \,. \tag{2.127}$$

Because the basis is orthogonal, for any $M \in \mathcal{M}$,

$$\mathrm{Tr}\Big[M^\dagger \rho\Big] = \sum_{M' \in \mathcal{M}} c_{M'} \,\mathrm{Tr}\Big[M^\dagger M'\Big] = c_M \|M\|_{Fr}^2 \,. \tag{2.128}$$

Moreover, by Cauchy-Schwarz, the inner product of M and ρ is at most the product of the norm of each, so

$$\|c_M M\|_{Fr} = \frac{\Big|\mathrm{Tr}\Big[M^\dagger \rho\Big]\Big|}{\|M\|_{Fr}} \leq \|\rho\|_{Fr} \,. \tag{2.129}$$

We also state two properties that hold for *any* randomized oracle:

Fact 53 (Randomized oracles preserve trace, Hermiticity, and positive semidefiniteness)**.** *Consider any randomized oracle $\mathcal{O}_F$ corresponding to a set of functions $f \in F$. Then $\mathcal{O}_F$ preserves the trace of its input. Moreover, if the input M is Hermitian, so is $\mathcal{O}_F[M]$; if M is also positive semidefinite, so is $\mathcal{O}_F[M]$.*

Proof. Consider any input matrix M. Then

$$\mathrm{Tr}[\mathcal{O}_F[M]] = \mathrm{Tr}\left[\frac{1}{|F|}\sum_{f \in F} \mathcal{U}_f[M]\right] = \frac{1}{|F|}\sum_{f \in F} \mathrm{Tr}\Big[U_f M U_f^\dagger\Big] = \frac{1}{|F|}\sum_{f \in F} \mathrm{Tr}[M] = \mathrm{Tr}[M] \,. \tag{2.130}$$

Now suppose M is Hermitian; that is, $M^\dagger = M$. Then

$$\mathcal{O}_F[M]^\dagger = \big(\frac{1}{|F|}\sum_{f \in F} \mathcal{U}_f[M]\big)^\dagger \tag{2.131}$$

$$= \frac{1}{|F|} \sum_{f \in F} \left(U_f M U_f^\dagger\right)^\dagger = \frac{1}{|F|} \sum_{f \in F} U_f M^\dagger U_f = \mathcal{O}_F[M] \,. \tag{2.132}$$

Furthermore, suppose M is positive semidefinite; that is, there is a matrix B such that $M = B^\dagger B$. Then

$$\mathcal{O}_F[M] = \frac{1}{|F|} \sum_{f \in F} \mathcal{U}_f[M] = \frac{1}{|F|} \sum_{f \in F} U_f B^\dagger B U_f^\dagger = \sum_{f \in F} \left(BU_f\right)^\dagger \left(BU_f\right) \,, \tag{2.133}$$

which is a sum of positive semidefinite matrices. Thus, $\mathcal{O}_F[M]$ is positive semidefinite. □

Fact 54 (Randomized oracles do not increase nuclear norm or Frobenius norm)**.** *Consider any randomized oracle $\mathcal{O}_F$ corresponding to a set of functions $f \in F$. Then $\mathcal{O}_F$ does not increase the nuclear norm nor the Frobenius norm of its input.*

Proof. Recall that both the nuclear norm and Frobenius norm are unitarily invariant. Now consider any input matrix M. Then the nuclear norm of $\mathcal{O}_F[M]$ is

$$\|\mathcal{O}_F[M]\|_1 = \left\| \frac{1}{|F|} \sum_{f \in F} U_f M U_f^\dagger \right\|_1 \leq \frac{1}{|F|} \sum_{f \in F} \left\| U_f M U_f^\dagger \right\|_1 \tag{2.134}$$

$$= \frac{1}{|F|} \sum_{f \in F} \|M\|_1 = \|M\|_1 \,. \tag{2.135}$$

The Frobenius norm of $\mathcal{O}_F[M]$ follows in exactly the same way. □

We use one additional property of density matrices in the proof of Lemma 33:

Fact 55. *Consider any $N \times N$ density matrix ρ and normalized states $|v\rangle, |w\rangle$. If*

$$|\langle v| \rho |w\rangle| = \Omega(\frac{1}{\text{poly}(n)}),$$

then both $\langle v| \rho |v\rangle$ and $\langle w| \rho |w\rangle$ are $\Omega(\frac{1}{\text{poly}(n)})$.

Proof. Recall that a density matrix is Hermitian and positive semidefinite, so it is diagonalizable and has real and nonnegative eigenvalues. As a result, it has a decomposition

$$\rho = S^\dagger \Lambda S = S^\dagger \sqrt{\Lambda}\sqrt{\Lambda} S \tag{2.136}$$

$$= (\sqrt{\Lambda} S)^\dagger (\sqrt{\Lambda} S) = A^\dagger A\,, \tag{2.137}$$

for some diagonal Λ and $A := \sqrt{\Lambda} S$. Then by Cauchy-Schwarz,

$$|\langle v|\rho|w\rangle|^2 = \left|(A|v\rangle)^\dagger (A|w\rangle)\right|^2 \leq \left|(A|v\rangle)^\dagger (A|v\rangle)\right| \cdot \left|(A|w\rangle)^\dagger (A|w\rangle)\right| = \langle v|\rho|v\rangle \cdot \langle w|\rho|w\rangle\ . \tag{2.138}$$

Since $\mathrm{Tr}[\rho] = 1$, $\langle\psi|\rho|\psi\rangle \leq 1$ for all normalized states $|\psi\rangle$. Thus, both $\langle v|\rho|v\rangle$ and $\langle w|\rho|w\rangle$ are at least $|\langle v|\rho|w\rangle|^2 = \Omega(\frac{1}{\mathrm{poly}(n)})$. □

2.10 Deferred proofs

Theorem 56 (Theorem 29, restated; [Harrow, 2013b, Proposition 2])**.** *Consider a finite group G, a vector space V, and a representation $R : G \to L(\mathsf{V})$. Then the operator*

$$\Pi_R := \frac{1}{|G|} \sum_{g \in G} R(g) \tag{2.139}$$

is an orthogonal projector onto $\mathsf{V}^G \subseteq \mathsf{V}$, where

$$\mathsf{V}^G := \{v \in \mathsf{V} \,:\, R(g)[v] = v \,\forall g \in G\}\,. \tag{2.140}$$

Proof. We include the proof for completeness.

Note that for any $g \in G$,

$$R(g)\Pi_R = R(g)\frac{1}{|G|}\sum_{g' \in G} R(g') = \frac{1}{|G|}\sum_{g' \in G} R(gg') = \frac{1}{|G|}\sum_{g^{-1}g' \in G} R(g') = \Pi_R \,. \tag{2.141}$$

This implies $\Pi_R \Pi_R = \Pi_R$:

$$\Pi_R \Pi_R = \frac{1}{|G|}\sum_{g \in G} R(g)\Pi_R = \frac{1}{|G|}\sum_{g \in G} \Pi_R = \Pi_R \,. \tag{2.142}$$

So Π_R is a projection.

Note that for any $v \in \mathsf{V}$ and $g \in G$,

$$R(g)[\Pi_R[v]] = (R(g) \circ \Pi_R)[v] = \Pi_R[v] \,, \tag{2.143}$$

so $\Pi_R[v] \in \mathsf{V}^G$. And similarly, for all $w \in \mathsf{V}^G$,

$$\Pi_R[w] = \frac{1}{|G|}\sum_{g \in G} R(g)[w] = \frac{1}{|G|}\sum_{g \in G} w = w \in \mathsf{V}^G \,. \tag{2.144}$$

So the image of Π_R is exactly V^G.

In order to consider orthogonality, we must define an inner product. Consider an inner product $(u|v)$ for $u, v \in \mathsf{V}$. If for some $g \in G$, $(R(g)[u]|R(g)[v]) \neq (u|v)$, use the inner product

$$\langle u, v \rangle = \frac{1}{|G|}\sum_{g \in G} (R(g)[u]|R(g)[v]) \,. \tag{2.145}$$

Then under this inner product, $R(g)$ is a unitary operator for all $g \in G$:

$$\langle R(g)[u], R(g)[v] \rangle = \sum_{g' g^{-1} \in G} (R(g')[u]|R(g')[v]) = \langle u, v \rangle \,. \tag{2.146}$$

This implies that Π_R is an orthogonal projection:

$$\begin{aligned}\langle \Pi_R[u], v\rangle &= \frac{1}{|G|}\sum_{g,g'\in G}(R(gg')[u]|R(g)[v]) \\ &= \frac{1}{|G|}\sum_{h,h'\in G}(R(h)[u]|R(h')[v]) \\ &= \frac{1}{|G|}\sum_{j,j'\in G}(R(j)[u]|R(jj')[v]) = \langle u, \Pi_R[v]\rangle\,. \end{aligned} \tag{2.147}$$

□

Theorem 57 (Theorem 30, restated; oracles on density matrices form a representation)**.** *Consider a group G of functions $f : [N] \to [N]$ with bitwise $\oplus$ as the group operation. Then the map $f \mapsto \mathcal{U}_f$ is a representation over the vector space of $2N^2 \times 2N^2$ complex matrices.*

Similarly, consider a group $\widetilde{G}$ of permutations $\pi : [N] \to [N]$ with composition as the group operation. Then the map $\pi \mapsto \widetilde{\mathcal{U}}_\pi$ is a representation over the vector space of $2N \times 2N$ complex matrices.

Proof. See that $f_1 f_2$ acts as $(f_1 f_2)^z(x) = f_1^z(x) \oplus f_2^z(x)$ for any $f_1, f_2 \in G$. The associated unitary acts as

$$U_{f_1 f_2}\sum_{c,x\in[N],z\in\{\pm1\}}\alpha_{c,x,z}\,|c,x,z\rangle = \sum_{c,x\in[N],z\in\{\pm1\}}\alpha_{c,x,z}\,|c\oplus(f_1f_2)^z(x),x,z\rangle \tag{2.148}$$

$$= \sum_{c,x\in[N],z\in\{\pm1\}}\alpha_{c,x,z}\,|c\oplus f_1^z(x)\oplus f_2^z(x),x,z\rangle \tag{2.149}$$

$$= U_{f_1}U_{f_2}\sum_{c,x\in[N],z\in\{\pm1\}}\alpha_{c,x,z}\,|c,x,z\rangle\,. \tag{2.150}$$

Since $U_{f_1f_2} = U_{f_1}U_{f_2}$, $\mathcal{U}_{f_1f_2} = \mathcal{U}_{f_1}\circ\mathcal{U}_{f_2}$. So the map $f \mapsto \mathcal{U}_f$ respects the group operation of G.

Similarly, $\pi_1\pi_2$ acts as $(\pi_1\pi_2)^z(x) = \pi_1^z(\pi_2^z(x))$ for any $\pi_1, \pi_2 \in \widetilde{G}$. The associated

unitary acts as

$$\widetilde{U}_{\pi_1\pi_2} \sum_{x\in[N],z\in\{\pm1\}} \alpha_{x,z} \left|x,z\right\rangle = \sum_{x\in[N],z\in\{\pm1\}} \alpha_{x,z} \left|(\pi_1\pi_2)^z(x),z\right\rangle \tag{2.151}$$

$$= \sum_{x\in[N],z\in\{\pm1\}} \alpha_{x,z} \left|\pi_1^z(\pi_2^z(x)),z\right\rangle \tag{2.152}$$

$$= \widetilde{U}_{\pi_1}\widetilde{U}_{\pi_2} \sum_{x\in[N],z\in\{\pm1\}} \alpha_{x,z} \left|x,z\right\rangle . \tag{2.153}$$

Since $\widetilde{U}_{\pi_1\pi_2} = \widetilde{U}_{\pi_1}\widetilde{U}_{\pi_2}, \widetilde{\mathcal{U}}_{\pi_1\pi_2} = \widetilde{\mathcal{U}}_{\pi_1} \circ \widetilde{\mathcal{U}}_{\pi_2}$. So the map $\pi \mapsto \widetilde{U}_\pi$ respects the group operation of $\widetilde{G}$. □

Theorem 58 (Theorem 31, restated; some randomized oracles are orthogonal projectors)**.** *Consider a group G of functions $f : [N] \to [N]$ with bitwise $\oplus$ as the group operation. Then $\mathcal{O}_G$ is an orthogonal projector, under the Frobenius inner product $(x|y) = \mathrm{Tr}\left[x^\dagger y\right]$ for $x, y \in \mathbb{C}^{2N^2\times 2N^2}$, onto*

$$\mathsf{V}_G := \{\rho \in \mathbb{C}^{2N^2\times 2N^2} : \mathcal{U}_f[\rho] = \rho \,\forall f \in G\} . \tag{2.154}$$

Similarly, consider a group $\widetilde{G}$ of permutations $\pi : [N] \to [N]$ with composition as the group operation. Then $\widetilde{\mathcal{O}}_{\widetilde{G}}$ is an orthogonal projector, under the Frobenius inner product $(x|y) = \mathrm{Tr}\left[x^\dagger y\right]$ for $x, y \in \mathbb{C}^{2N\times 2N}$, onto

$$\widetilde{\mathsf{V}}_{\widetilde{G}} := \{\rho \in \mathbb{C}^{2N\times 2N} : \widetilde{\mathcal{U}}_\pi[\rho] = \rho \,\forall \pi \in \widetilde{G}\} . \tag{2.155}$$

Proof. By Theorem 30, the maps $f \mapsto \mathcal{U}_f$ for $f \in G$ and $\pi \mapsto \widetilde{\mathcal{U}}_\pi$ for $\pi \in \widetilde{G}$ both are representations. Note that for any unitary U and square matrices x, y of the same dimensions,

$$(UxU^\dagger|UyU^\dagger) = \mathrm{Tr}\left[(UxU^\dagger)^\dagger UyU^\dagger\right] = \mathrm{Tr}\left[Ux^\dagger U^\dagger UyU^\dagger\right] = \mathrm{Tr}\left[Ux^\dagger yU^\dagger\right] = \mathrm{Tr}\left[x^\dagger y\right] . \tag{2.156}$$

So the representation of each group element is a unitary operator under the Frobenius inner product. The proof then follows by Theorem 29. □

Lemma 59 (Lemma 34, restated; can't approximate too many subset states)**.** *Consider a Hermitian $N \times N$ matrix ρ that is positive semidefinite and has trace at most 1. Consider the set of all subsets $V \subseteq [N]$, where $|V| = N^\alpha$ for a fixed $0 < \alpha < \frac{1}{2}$. Then the fraction of subsets V such that $\langle V| \rho |V\rangle = \Omega(\frac{1}{\mathrm{poly}(n)})$ decreases faster than any exponential in* $\mathrm{poly}(n)$*.*

Proof. Let $\mathcal{W}$ be any set of subsets V, with size $|\mathcal{W}| = \omega(\mathrm{poly}(n)) = o(\frac{2^{\sqrt{n}}}{\mathrm{poly}(n)})$, such that $|\langle V_1|V_2\rangle| = O(2^{-\sqrt{n}})$ for all $V_1, V_2 \in \mathcal{W}$. Consider an approximate basis for ρ, consisting of

- $|V\rangle\langle V|$, for all $V \in \mathcal{W}$, and
- a set of matrices $\mathcal{M}'$ orthogonal to all other matrices, and with Frobenius norm 1.

Representing ρ into this approximate basis is not unique, but the coefficient on each term must be close to $\langle V| \rho |V\rangle$. Recall that the Frobenius norm of a density matrix ρ is at most 1; by Remark 52, each $|c_M| \leq 1$:

$$\rho := \sum_{V \in \mathcal{W}} c_V |V\rangle\langle V| + \sum_{M \in \mathcal{M}'} c_M M\,, \tag{2.157}$$

$$\langle V| \rho |V\rangle = \sum_{V' \in \mathcal{W}} c_{V'} \left|\langle V'|V\rangle\right|^2 = c_V + \sum_{V' \in \mathcal{W}, V' \neq V} c_{V'} \left|\langle V'|V\rangle\right|^2 = c_V \pm o(\frac{1}{\mathrm{poly}(n)})\,. \tag{2.158}$$

Let's inspect $\mathrm{Tr}\left[\rho^2\right] \leq \mathrm{Tr}[\rho]^2 = 1$:

$$1 \geq \mathrm{Tr}\left[\rho^2\right] = \sum_{V_1 \neq V_2 \in \mathcal{W}} \mathrm{Tr}\left[c_{V_1} c_{V_2}^\dagger |V_1\rangle\langle V_1|V_2\rangle\langle V_2|\right] + \sum_{V \in \mathcal{W}} |c_V|^2 + \sum_{M \in \mathcal{M}'} |c_M|^2\,. \tag{2.159}$$

Notice that the first summation has $o(2^{2\sqrt{n}})$ terms of size $O(2^{-2\sqrt{n}})$, so it has magnitude $o(1)$. Since the last summation is nonnegative, the second summation must be at most a

constant $O(1)$. Thus, for at most a $O(\text{poly}(n))$ distinct choices of $V \in \mathcal{W}$, c_V (and thus $\langle V| \rho |V\rangle$) can have magnitude $\Omega(\frac{1}{\text{poly}(n)})$.

So, after choosing a polynomial number of $|V\rangle \langle V|$ for $V \in \mathcal{W}$, all other subsets V that have good overlap with ρ must have $\omega(2^{-\sqrt{n}})$ overlap with some previously chosen $V \in \mathcal{W}$. The fractional number of subsets V with this property is at most

$$O(\text{poly}(n)) \times \frac{\binom{N}{N^\alpha(1-\omega(2^{-\sqrt{n}}))}}{\binom{N}{N^\alpha}} = \frac{O(\text{poly}(n))}{N^{N^\alpha \omega(2^{-\sqrt{n}})}} = \frac{1}{2^{n 2^{\alpha n(1-o(1))}}} = o(\frac{1}{2^{\text{poly}(n)}}) \,. \tag{2.160}$$

□

Lemma 60 (Lemma 35, restated; not too many subsets can have elevated mean)**.** *Consider any $N \times N$ POVM $\{E, \mathbb{I} - E\}$, and the set of all subsets $V \subseteq [N]$, where $|V| = N^\alpha$ for a fixed $0 < \alpha < \frac{1}{2}$. Then the fraction of subsets V where*

$$|f(V)| := \left| \frac{1}{|V|} \text{Tr}[\mathbb{I}_V E] - \frac{1}{N} \text{Tr}[E] \right| = \Omega(\frac{1}{\text{poly}(n)}) \,, \tag{2.161}$$

decreases faster than any exponential in $\text{poly}(n)$.

Proof. Suppose for contradiction that $|f(V)| = \Omega(\frac{1}{\text{poly}(n)})$ for at least a $\Omega(\frac{1}{2^{\text{poly}(n)}})$ fraction of all subsets $V \subseteq [N]$ where $|V| = N^\alpha$. Because $f(V)$ is real, it can be either positive or negative; by pigeonhole, a $\frac{1}{2}\Omega(\frac{1}{2^{\text{poly}(n)}}) = \Omega(\frac{1}{2^{\text{poly}(n)}})$ fraction of subsets V must have the same sign of $f(V)$ (without loss of generality, assume it is positive). Let $\mathcal{W}$ contain the subsets V with this property.

Let $k = N^{1-\alpha} - N^{1-\alpha-(\alpha/3-\epsilon)}$ for any $\epsilon > 0$. We now build a finite sequence $\mathcal{Q} = \{V_1, V_2, \ldots, V_k\} \subseteq \mathcal{W}$ that nearly cover $[N]$, and where no two elements share more than $N^{\alpha/3}$ elements. First, choose $V_1 \in \mathcal{W}$. On subsequent draws, choose another $V_m \in \mathcal{W}$ such that

$$\frac{|V_m \cap \left(\cup_{1 \leq j < m} V_j\right)|}{|V|} \leq N^{\alpha/3} \,. \tag{2.162}$$

Suppose we could not draw $V_m \in \mathcal{W}$; then all remaining elements of $\mathcal{W}$ share at least $N^{\alpha/3}$ elements with one of $\{V_1, \ldots, V_{m-1}\}$. The fractional size of $\mathcal{W}$ (compared to all subsets) would then be at most

$$(m-1) + \frac{\binom{(m-1)N^\alpha}{N^{\alpha/3}}\binom{N}{N^\alpha - N^{\alpha/3}}}{\binom{N}{N^\alpha}} \leq \frac{\binom{(k-1)N^\alpha}{N^{\alpha/3}}}{\binom{N-(N^\alpha - N^{\alpha/3})}{N^{\alpha/3}}} \tag{2.163}$$

$$\leq \left(\frac{k}{N^{1-\alpha}-1}\right)^{N^{\alpha/3}} \leq \left(1 - \frac{1}{N^{\alpha/3-\epsilon}}\right)^{N^{\alpha/3}} \leq o\left(\frac{1}{2^{\mathrm{poly}(n)}}\right), \tag{2.164}$$

which is a contradiction. So we can always construct $\mathcal{Q}$. Now consider

$$\sum_{V \in \mathcal{Q}} f(V) = \frac{1}{|V|} \sum_{V \in \mathcal{Q}} \mathrm{Tr}[E \mathbb{I}_V] - \frac{|\mathcal{Q}|}{N} \mathrm{Tr}[E]\,. \tag{2.165}$$

By supposition, all $f(V) > 0$ and have magnitude $\Omega(\frac{1}{\mathrm{poly}(n)})$, so this quantity should be $\Omega(|\mathcal{Q}|\frac{1}{\mathrm{poly}(n)})$. We show however that this sum is small because the two sums are very close. The number of "overcounts" of an element $x \in [N]$ is at most $|\mathcal{Q}|N^{\alpha/3}$, and the number of elements $x \in [N]$ not in some $V \in \mathcal{Q}$ is at most

$$N - k(N^\alpha - N^{\alpha/3}) = O(N^{1-(\alpha/3-\epsilon)})\,. \tag{2.166}$$

Recall that the diagonal elements of a POVM are each nonnegative and at most 1. So then

$$\left|\frac{1}{|V|}\sum_{V \in \mathcal{Q}} \mathrm{Tr}[E\mathbb{I}_V] - \frac{|\mathcal{Q}|}{N}\mathrm{Tr}[E]\right| = O(\frac{|\mathcal{Q}|}{|V|}N^{\alpha/3}) + O(\frac{|\mathcal{Q}|}{N}N^{1-(\alpha/3-\epsilon)}) = o(\frac{|\mathcal{Q}|}{\mathrm{poly}(n)})\,. \tag{2.167}$$

This contradicts our supposition. So the fraction of subsets V where $|f(V)| = \Omega(\frac{1}{\mathrm{poly}(n)})$ decreases faster than any exponential in $\mathrm{poly}(n)$. □

2.11 Our setup contrasted with a discrete-time quantum walk

The way one stores a graph in an oracle drastically changes the difficulty of some problems. Consider a *discrete-time quantum walk* Watrous [1998], which allows a vertex access to a superposition of its neighbors.[4] Given a d-regular graph $G(V, E)$, the operator $W : \mathbb{C}^{N^2 \times N^2}$ acts as

$$W = \left(\sum_{(j,k) \in E} |j,k\rangle \langle k,j| \right) C \tag{2.168}$$

$$C = \sum_{j \in V} |j\rangle \langle j| \otimes (2 \left|\partial_j\right\rangle \left\langle \partial_j\right| - \mathbb{I}) \tag{2.169}$$

$$\left|\partial_j\right\rangle = \frac{1}{\sqrt{d}} \sum_{(j,k) \in E} |k\rangle \ . \tag{2.170}$$

Using a discrete-time quantum walk, we can learn about the mixing properties of the associated graph; these are fundamentally related to the graph's spectral gap Godsil and Zhan [2019].

By contrast, we query each neighbor of a vertex $v \in G$ with the value of the registers encoding $i \in [d/2]$ (defined by a G-coded function). For example, Ambainis et al. [2011a] uses a similar oracle to show that deciding whether a graph is a single expander graph or two equal-sized disconnected expander graphs is outside of BQP. Intuitively, a lack of superposition access to neighbors of a vertex makes it harder for a quantum computer to "traverse" the graph.

4. [Childs, 2022, Chapter 17] has a good introduction to this topic.

CHAPTER 3

PROOFS WITHOUT RELATIVE PHASE

This chapter is adapted (with minor modifications) from the work of Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. Quantum merlin-arthur and proofs without relative phase. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference, ITCS 2024, January 30 to February 2, 2024, Berkeley, CA, USA*, volume 287 of *LIPIcs*, pages 9:1–9:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024b. doi:10.4230/LIPICS.ITCS.2024.9. URL `https://doi.org/10.4230/LIPIcs.ITCS.2024.9`.

3.1 Techniques and Contributions

Our primary technical contribution is to show a QMA^+ protocol for a NEXP-complete problem. This directly extends the work of Jeronimo and Wu Jeronimo and Wu [2023b], who show a $\mathsf{QMA}^+(2)$ protocol for a NEXP-complete problem. As in Jeronimo and Wu [2023b], we study constraint satisfaction problems (CSPs) with *constant* gap. In $(1,\delta)$-GapCSP, either *all* constraints can be satisfied, or at most a δ fraction of constraints can be satisfied. These problems are known to be NP-hard or NEXP-hard (depending on the problem size) using the PCP theorem Arora and Safra [1992], Arora et al. [1998a], Harsha [2004].

Before proving QMA^+ with some constant gap equals NEXP, we prove $\mathsf{QMA}^+_{\log}$ (with some other constant gap) equals NP. This choice (also taken by Jeronimo and Wu [2023b]) is pedagogical: it allows us to explain the protocol without worrying about input encoding size, since QMA^+ has a polynomial amount of space and verifier runtime. Here, we consider $(1,\delta)$-GapCSP with polynomially many variables and clauses; the quantum proof must certify that there is a satisfying assignment to all clauses.

The $\mathsf{QMA}^+_{\log}(2)$ protocol of Jeronimo and Wu [2023b] crucially relies on an estimate of *sparsity* (ℓ_1 norm) of a quantum state without relative phase. The overlap of a m-qubit quantum state without relative phase $|\psi\rangle$ with $|+\rangle^{\otimes m}$ is exactly the value $2^{-m/2}$.

$\| |\psi\rangle \|_1$. With multiple quantum proofs $|\psi_1\rangle \otimes \ldots |\psi_k\rangle$, one can *estimate* the sparsity by repeating this "sparsity test" on each $|\psi_i\rangle$, and using a swap test to ensure that all $|\psi_i\rangle$ are approximately equal. Interestingly, no other part of their protocol requires the *no relative phase* assumption.[1]

In $\mathsf{QMA}^{+}_{\log}$, we have a single quantum proof, so we cannot use this test to estimate sparsity. Instead, we design a similar test that directly enforces a *rigidity* property of the proof.[2] The required form is $\frac{1}{\sqrt{R}} \sum_{j \in [R]} |j\rangle \left|\vec{v}_j\right\rangle$, where the second register is constant-sized. Using a "sparsity test" over the second register, Arthur ensures that the second register has one $\vec{v}_j$ per j; but using the *complement* of a "sparsity test" over the whole proof, Arthur ensures the overall state maximizes ℓ_1 norm. States of the required form are optimal for this combination of tests. We make use of the *no relative phase* property in Lemmas 66 and 68.

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

Checking *consistency* is done using a technique called "regularization" from the PCP literature Dinur [2007]; for each constraint j, we verify that each variable participating in j has the same value in exactly d other constraints for some constant d, in a way that the edges form an *expander graph*. The expansion property guarantees that cheating on this test is as damaging as cheating on the satisfiability test. Jeronimo and Wu Jeronimo and Wu [2023b]

1. Formally, Jeronimo and Wu [2023b] studies states with non-negative amplitudes. Recall that the set of these states, up to global phase, are equivalent to states with *no relative phase*.

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

use a swap test to implement these new checks, but this requires multiple quantum proofs. We show how to use a Hadamard test (which requires only one quantum proof) to achieve the same result, building on ideas from previous work Bassirian et al. [2023a]. Since there exists a δ such that $(1, \delta)$-GapCSP is NP-hard, this completes the proof of $\mathsf{NP} \subseteq \mathsf{QMA}^{+}_{\log}$ with some constant gap.

When scaling up to QMA^{+}, one must be careful of how to succinctly encode the input of a NEXP-complete problem. The PCP theorem allows us to choose $(1, \delta)$-GapCSP that is *succinct*, but we need stronger properties. Following the adjustments taken in Jeronimo and Wu [2023b], we choose a PCP system for NEXP that is both *doubly explicit* and *strongly uniform*. *Doubly explicit* means that one can efficiently compute the variables participating in a given constraint *and* the constraints a given variable participates in; using this, we can implement the consistency checks in polynomial time. *Strongly uniform* means that the number of constraints a variable participates in is efficiently computable, and one of a fixed number of possibilities; using this, we only need to build a fixed number of expander graphs during regularization. Recent work also shows how to construct exponentially-sized expander graphs in polynomial time Lubotzky [2009], Alon [2021]. Once we are through these input encoding difficulties, our protocol is identical to that for NP.

Fundamentally, the *no relative phase* property allows Arthur to verify a number of constraints exponential in the number of qubits. Attempts to do this for $\mathsf{QMA}(2)$ gave too small a promise gap Blier and Tapp [2010], Pereszlényi [2012b], Gall et al. [2012], too many provers Aaronson et al. [2008], Chen and Drucker [2010], Chiesa and Forbes [2011], or too much space or time Harrow and Montanaro [2013], Natarajan and Zhang [2023]. Jeronimo and Wu Jeronimo and Wu [2023b] show that $\mathsf{QMA}^{+}(2)$ circumvents this difficulty: using *no relative phase* and *unentanglement*, Arthur enforces the *sparsity* of a quantum proof to solve a NEXP-complete problem. At the center of our work is the insight that *no relative phase* is enough for Arthur to require constant-sized answers to exponentially many questions, solving a NEXP-complete problem with a single polynomial-size quantum proof.

3.2 Related work

The complexity class $\mathsf{QMA}(2)$ The complexity class $\mathsf{QMA}(2)$ is known to have promise gap amplification, and to be equal to $\mathsf{QMA}(k)$ for any k at most polynomial in n Harrow and Montanaro [2013]. It is not obvious how to test for entanglement; even determining whether a polynomially-sized vector is entangled is NP-hard Gharibian [2009]. If there exist efficient approximate "disentanglers" that can create any separable state, then $\mathsf{QMA} = \mathsf{QMA}(2)$; see Aaronson et al. [2008] for some progress. Gharibian et al. [2018] describe quantum variants of the polynomial hierarchy and connect their properties to bounds on $\mathsf{QMA}(2)$. It is not even known whether there is a *quantum* oracle separating QMA and $\mathsf{QMA}(2)$ Aaronson [2021].

PCPs and expander graphs Probabilistically checkable proofs (PCPs) show hardness for CSPs with a constant gap Arora and Safra [1992], Arora et al. [1998a], Harsha [2004]. Dinur Dinur [2007] proves the PCP theorem using a *regularization* step, which adds new constraints associated with the edges of a *regular* expander graph. *Polynomial-time* regularization for NEXP requires an efficient description of exponentially-sized expander graphs. Recent advances in expander graph constructions Lubotzky [2009], Alon [2021] allow for this type of regularization, first used in Jeronimo and Wu [2023b].

Quantum states and relative phase Up to a global phase, states with non-negative amplitudes are equivalent to states with no relative phase. Jeronimo and Wu [2023b] propose the class QMA^+ and $\mathsf{QMA}^+(2)$, and show $\mathsf{QMA}^+(2) = \mathsf{NEXP}$. Note that by contrast, QMA restricted to states with *real* amplitudes is equal to QMA McKague [2013]. Relative phase was recently proposed as a quantum resource Xu [2023]. For both QMA and $\mathsf{QMA}(2)$, restricting Merlin in completeness to send a *subset state* does not change the power of the complexity class (i.e., $\mathsf{QMA} = \mathsf{SQMA}$ and $\mathsf{QMA}(2) = \mathsf{SQMA}(2)$). Grilo et al. [2014] also shows why their proof strategy fails if Merlin is restricted in both completeness and soundness.

Rigidity and games Rigidity was first formally introduced in the context of non-local games Mayers and Yao [2004], and have been used to prove several complexity class equalities. For example, the CHSH game Clauser et al. [1969] tests for a maximally entangled state on two qubits Tsirelson [1993], and was used to prove $\mathsf{QMIP} = \mathsf{MIP}^*$ Reichardt et al. [2013]. The Mermin-Peres magic square game tests for two copies of a maximally entangled quantum state, and was used to prove $\mathsf{MIP}^* = \mathsf{RE}$ Ji et al. [2021b]. Rigidity is known to exist in broader contexts, including some (but not all) linear constraint games Cui et al. [2020] and monogamy-of-entanglement games Broadbent and Culf [2023].

3.3 Our setup

We restate the definition of $\mathsf{QMA}^+(k)$ from Jeronimo and Wu [2023b]. When the proof length is not specified, it is allowed to be at most any polynomial in input size. We follow the conventions $\mathsf{QMA}^+ := \mathsf{QMA}^+(1)$, $\mathsf{QMA}^+(k) := \bigcup_{c-s=\Omega(1)} \mathsf{QMA}^+(k)_{c,s}$, and $\mathsf{QMA}^+_{\log} := \mathsf{QMA}^+$ with proof length at most $O(\log n)$.

Definition 61 ($\mathsf{QMA}^+_\ell(k)_{c,s}$)**.** *Let $k : \mathbb{N} \to \mathbb{N}$ and $s, c, \ell : \mathbb{N} \to \mathbb{R}^+$ be polynomial time computable functions. A promise problem $\mathcal{L}_{yes}, \mathcal{L}_{no} \subseteq \{0,1\}^*$ is in $\mathsf{QMA}^+_\ell(k)_{c,s}$ if there exists a BQP verifier V such that for every $n \in \mathbb{N}$ and every $x \in \{0,1\}^n$:*

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

$$\mathbf{Pr}[V(x, |\psi_1\rangle \otimes \ldots \otimes \left|\psi_{k(n)}\right\rangle) \text{ accepts}] \geq c(n)\,. \tag{3.1}$$

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

$$\mathbf{Pr}[V(x, |\psi_1\rangle \otimes \ldots \otimes \left|\psi_{k(n)}\right\rangle) \text{ accepts}] \leq s(n)\,. \tag{3.2}$$

We make a few remarks on this complexity class, with extended discussion in Section 3.6. First, we stress that the restriction to quantum proofs with non-negative amplitudes is *promise-symmetric*, i.e. both in completeness and in soundness. This is unlike, for example, the class SQMA Grilo et al. [2014]. Although the restriction to *subset states* is stronger than *non-negative amplitudes*,[3] its use *only* in completeness allows for SQMA = QMA. In fact, our work implies that QMA with a promise-symmetric *subset state* restriction also interpolates from QMA to NEXP, depending on the size of the promise gap.

We also explain why the promise gap of QMA^+ cannot obviously be amplified. The first strategy one might try is *parallel repetition*: an honest Merlin sends multiple copies of the original proof and Arthur verifies each copy of the original proof. For QMA, entangling the copies in soundness does not help Merlin, since Arthur's protocol is sound for all quantum states. But perhaps unintuitively, it *can* help for QMA^+. (See Fact 88 for a simple example.) This is because partial measurement can *destroy* the restriction on the quantum proof! For example, Arthur's first measurement may introduce relative phase in the rest of the proof. This fact also obstructs more clever amplification strategies for QMA such as the proof-length preserving variant Marriott and Watrous [2005a].

Furthermore, it is not clear how to upper-bound QMA^+ beyond the trivial NEXP.[4] One technique to upper-bound QMA is to find the optimal proof using a semidefinite program (or a general convex program). This shows that $\mathsf{QMA} \subseteq \mathsf{PSPACE}$ (or $\mathsf{QMA} \subseteq \mathsf{EXP}$ with a convex program). But these arguments do not immediately transfer to QMA^+. Convex optimization over states with non-negative amplitudes is equivalent to optimizing over the *copositive cone* Burer [2011]. Even the *weak membership* problem over the copositive cone (deciding if the optimal vector is close to a non-negative vector) is NP-hard in polynomially-sized vector spaces; recall that quantum states are in *exponentially*-sized vector spaces. These are the same reasons that prevent straightforward upper bounds for QMA(2) Gharibian [2009].

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

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

3.4 $\mathsf{QMA}^{+}_{\log}$ Protocol for NP

We first define the problem we consider:

Definition 62 (CSP system)**.** *A (N, R, q, Σ)-CSP system $\mathcal{C}$ on N variables with values in Σ consists of a set (possibly a multi-set) of R constraints $\{\mathcal{C}_1, \ldots, \mathcal{C}_R\}$ where the arity of each constraint is exactly q.*

Definition 63 (Value of CSP)**.** *The* value *of a (N, R, q, Σ)-CSP system $\mathcal{C}$ is the maximum fraction of satisfiable constraints over all possible assignments $\sigma : [N] \to \Sigma$. The value of $\mathcal{C}$ is denoted* $\mathbf{val}(\mathcal{C})$.

Definition 64 (GapCSP)**.** *The* $(1, \delta)$-GapCSP problem *inputs a CSP system $\mathcal{C}$. The task is to distinguish whether $\mathcal{C}$ is such that (in completeness) $\mathbf{val}(\mathcal{C}) = 1$ or (in soundness) $\mathbf{val}(\mathcal{C}) \leq \delta$.*

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

Theorem 65 (Dinur [2007])**.** *There exist constants $q > 1$ and $|\Sigma|$ such that $(1, 1/2)$-GapCSP is NP-hard.*

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

$$|\psi\rangle := \sum_{j \in [R], x \in \Sigma^q} a_{j,x} |j\rangle |x\rangle \tag{3.3}$$

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

Arthur then applies one of two kinds of tests:

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

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

3.4.1 Rigidity tests

Arthur enforces *rigidity* of a quantum proof using two tests. The first test is the *Density test*, which maximizes ℓ_1 norm. Here, we measure the state in the Hadamard basis and *accept* if the outcome is $|+\rangle$.[5] Given $|\psi\rangle$, the success probability of this test is

$$D(|\psi\rangle) = |\langle+|\psi\rangle|^2 = \frac{1}{\kappa R}\left|\sum_{j\in[R],x\in\Sigma^q} a_{j,x}\right|^2 = \frac{1}{\kappa R}(\|\,|\psi\rangle\,\|_1)^2\,. \tag{3.4}$$

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

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

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

Given $|\psi\rangle$, the success probability of this test is

$$V(|\psi\rangle) = 1 - \langle +| \operatorname{Tr}_R(|\psi\rangle\langle\psi|)|+\rangle = 1 - \frac{1}{\kappa}\sum_{j\in[R]}\left|\sum_{x\in\Sigma^q} a_{j,x}\right|^2, \tag{3.5}$$

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

$$1 - \langle +| \left(\frac{1}{R}\sum_{j\in[R]} |\vec{v}_j\rangle\langle\vec{v}_j|\right)|+\rangle = 1 - \frac{1}{\kappa}. \tag{3.6}$$

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

Lemma 66. *Suppose $|\psi\rangle$ has no relative phase. Then $V(|\psi\rangle) \le 1 - \frac{1}{\kappa}$.*

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

$$1 - \frac{1}{\kappa}\sum_{j\in[R]}(\sum_{x\in\Sigma^q} a_{j,x})^2 = 1 - \frac{1}{\kappa}(\sum_{j\in[R]}\sum_{x,y\in\Sigma^q} a_{j,x}a_{j,y}) \tag{3.7}$$

$$= 1 - \frac{1}{\kappa}(1 + \sum_{j\in[R]}\sum_{x,y\in\Sigma^q;x\neq y} a_{j,x}a_{j,y}) \le 1 - \frac{1}{\kappa}, \tag{3.8}$$

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

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

Lemma 67. $D(|\psi\rangle) + V(|\psi\rangle) \le 1$.

Proof. By Cauchy-Schwarz,

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

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

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

Lemma 68. *Given $|\psi\rangle = \sum_{j,x} a_{j,x}|j\rangle|x\rangle$ with no relative phase (i.e. $a_{j,x} \geq 0$), let*

$$\gamma := \max_{\nu:[R]\to\Sigma^q} \sum_{j\in[R]} a_{j,\nu(j)}^2 \tag{3.9}$$

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

Proof. Note that for all $x \in \Sigma^q$, $a_{j,\sigma(j)} \geq a_{j,x}$; otherwise, σ is not maximizing. Using the proof of Lemma 66,

$$d = \sum_{j\in[R]}\sum_{x,y\in\Sigma^q;x\neq y} a_{j,x}a_{j,y} \geq \sum_{j\in[R]}\sum_{y\in\Sigma^q;y\neq\sigma(j)} a_{j,\sigma(j)}a_{j,y} \geq \sum_{j\in[R]}\sum_{y\in\Sigma^q;y\neq\sigma(j)} a_{j,y}^2\,. \tag{3.10}$$

So then,

$$\gamma = \sum_{j\in[R]} a_{j,\sigma(j)}^2 = (\sum_{j\in[R]}\sum_{x\in\Sigma^q} a_{j,x}^2) - (\sum_{j\in[R]}\sum_{x\in\Sigma^q; x\neq\sigma(j)} a_{j,x}^2) \geq 1-d\,. \tag{3.11}$$

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

Another requirement to get close to the upper bound is near-optimal success probability on *Density test*, up to Lemma 67. Consider any quantum proof that passes *Validity test* with probability close to $1-\frac{1}{\kappa}$ and *Density test* with probability close to $\frac{1}{\kappa}$; we show it must be close to a state of the form $\frac{1}{\sqrt{R}}\sum_{j\in R}|j\rangle\left|\vec{v}_j\right\rangle$. Now we can prove the soundness of the rigidity test by relying on the following fact:

Fact 69. *Let $0 \leq \Pi \leq \mathbb{I}$ be a positive semi-definite matrix, and let $|\psi_1\rangle$ and $|\psi_2\rangle$ be quantum states such that $|\langle\psi_1|\psi_2\rangle|^2 \geq 1-d$. Then $|\langle\psi_1|\Pi|\psi_1\rangle - \langle\psi_2|\Pi|\psi_2\rangle| \leq \sqrt{d}$.*

Proof. The quantity

$$|\langle\psi_1|\Pi|\psi_1\rangle - \langle\psi_2|\Pi|\psi_2\rangle| = |\mathrm{Tr}(\Pi(|\psi_1\rangle\langle\psi_1| - |\psi_2\rangle\langle\psi_2|))| \tag{3.12}$$

is upper-bounded by the trace distance of $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$, which has value

$$\sqrt{1-|\langle\psi_1|\psi_2\rangle|^2} \leq \sqrt{d}.$$ □

Lemma 70 (Rigidity lemma)**.** *Let $d_2 \geq d_1 \geq 0$ be small constants. Suppose $|\psi\rangle$, $|\phi\rangle$, and σ are defined as in Lemma 68, and $|\chi\rangle$ is defined as*

$$|\chi\rangle := \frac{1}{\sqrt{R}}\sum_{j\in[R]}|j\rangle|\sigma(j)\rangle\,. \tag{3.13}$$

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

Proof. By Lemma 68, we know that $|\langle\psi|\phi\rangle|^2 \geq 1-\kappa\cdot d_2$. So by Fact 69, for any quantum

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

$$|\langle\chi|\phi\rangle|^2 = \kappa\,|\langle+|\phi\rangle|^2 \geq 1 - \kappa(d_1 + \sqrt{\kappa \cdot d_2})\,. \tag{3.14}$$

Again by triangle inequality,

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

Intuitively, Lemma 70 allows us to *tune* the probability of each test in the NP protocol. As we explain in the analysis (Section 3.4.3), if the probabilities of running *Validity test* and *Density test* are much higher than that for constraint tests, then if d_1 or d_2 is large, these two tests catch a "deceptive" quantum proof in soundness. This allows constraint tests to focus on the case of small d_1 and d_2; i.e. nearly *rigid* quantum proofs.

3.4.2 Constraint tests

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

(i) (*satisfiability*) For all $j \in [R]$, the assignment $\vec{v}_j$ satisfies $\mathcal{C}_j$.

(ii) (*consistency*) Each variable is assigned the same value when participating in different constraints.

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

given in superposition.[6]

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

But this form of quantum proof gives Merlin a new way to "deceive": for a given variable, send different values depending on the constraint! We prevent this by checking for *consistency*, similarly to the pre-processing step of Dinur [2007] sometimes called *regularization*. As in [Jeronimo and Wu, 2023b, Section 7], we add "consistency constraints" to the CSP system $\mathcal{C}$ as follows:[7]

- For each variable $i\in[N]$, let V_i represent the constraints that i participates in.
- Fix a constant d. For each $i\in[N]$, draw a d-regular graph with vertices V_i that is *expanding*.[8]
- Each edge (j_1,j_2) of each expander V_i represents a "consistency constraint", where we assert that the value of variable i sent with constraint j_1 equals that sent with constraint j_2.

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

Claim 71 ([Dinur, 2007, Lemma 4.1])**.** *Consider a (N,R,q,Σ)-CSP system $\mathcal{C}$, and apply* regularization*. If $\mathbf{val}(\mathcal{C})=1$, then all "consistency constraints" can be simultaneously satisfied. If $\mathbf{val}(\mathcal{C})\leq\delta$, then the total number of unsatisfied constraints (u_s+u_e) is at least $(1-\delta)R$.*

6. There is a way around this limitation for CSP systems consisting of *unique game constraints*, where each (binary) constraint involving variables i_1,i_2 accepts exactly one $a(i_2)$ for each $a(i_1)$. See [Jeronimo and Wu, 2023b, Section 6] for more discussion.

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

8. For technical reasons of Claim 71, we require that the Cheeger constant is at least 2.

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

We now choose d permutations. It is a classical fact that the adjacency matrix of any d-regular graph can always be decomposed to d permutations. Let $\pi_1, \ldots, \pi_d$ be the decomposition of $\widetilde{G}$; recall that these are permutations on $V(\widetilde{G})$ where $|V(\widetilde{G})| = R \cdot q$. For each $k \in [d]$, we identify π_k with a permutation on $[R] \times [N]$, where any $(j, i) \in [R] \times [N]$ that is not a vertex of $\widetilde{G}$ (i.e. variable i does not participate in constraint j) is mapped to itself.[9] Note that this map always preserves the variable $i \in [N]$; without loss of generality, we also identify π_k with its restriction $[R] \times [N] \to [R]$. From here on, we use this last definition of π_k, which maps constraint j_1 that variable i participates in to another constraint j_2 that variable i participates in, and identity otherwise.

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

$$\left|\phi'\right\rangle := \frac{1}{\sqrt{q \cdot R}} \sum_{j \in [R]} \sum_{i \in \mathcal{C}_j} |j\rangle \left|\vec{v}_j\right\rangle |i\rangle \left|v_j(i)\right\rangle . \tag{3.15}$$

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

We now would like to construct a unitary on $\left|\phi'\right\rangle$ that maps $|j\rangle |i\rangle |value\rangle$ to $\left|j'\right\rangle |i\rangle |value\rangle$

9. These permutations (and their inverses) are all efficient because N and R are polynomially-sized.

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

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

$$|\phi\rangle := \frac{1}{\sqrt{q \cdot R}} \sum_{j \in R} \sum_{i \in \mathcal{C}_j} |j\rangle |i\rangle \left|v_j(i)\right\rangle . \tag{3.16}$$

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

$$\Pi_k : |j\rangle |i\rangle \left|v_j(i)\right\rangle \to |\pi_k(j,i)\rangle |i\rangle \left|v_j(i)\right\rangle . \tag{3.17}$$

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

Note that in a satisfiable instance, $\Pi_k |\phi\rangle = |\phi\rangle$. By contrast, if $v_j(i) \neq v_{j'}(i)$, $\Pi_k |\phi\rangle$ is orthogonal to $|\phi\rangle$. Hence, $\langle\phi| \Pi_k |\phi\rangle$ is the fraction of satisfied "consistency constraints" observed by π_k. We use the *Hadamard test* to measure this value, in a similar way to the *Spectral test* in Bassirian et al. [2023a]. Note that unlike the swap test, the Hadamard test only uses one copy of a quantum state.

Definition 72 (Hadamard test)**.** *Let $|\psi\rangle$ be a quantum state and U a unitary operator.*

1. *Prepend a control qubit to $|\psi\rangle$, to create $|0\rangle |\psi\rangle$.*
2. *Apply a Hadamard on the control qubit, to create $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) |\psi\rangle$.*
3. *Apply U, controlled by the control qubit, to create $\frac{1}{\sqrt{2}} (|0\rangle |\psi\rangle + |1\rangle U |\psi\rangle)$.*

4. Apply a Hadamard on the control qubit, to create:

$$\frac{1}{2}|0\rangle\left(|\psi\rangle + U|\psi\rangle\right) + \frac{1}{2}|1\rangle\left(|\psi\rangle - U|\psi\rangle\right).$$

5. Measure the control qubit, and accept if the output is 0*.*

The success probability is then

$$\frac{1}{4}\left\||\psi\rangle + U|\psi\rangle\right\|^2 = \frac{1}{2} + \frac{1}{4}\langle\psi|\, U + U^\dagger\,|\psi\rangle = \frac{1}{2} + \frac{\mathrm{Re}\,\langle\psi|\, U\,|\psi\rangle}{2}. \tag{3.18}$$

We now can describe the constraint tests together:

(i) With probability $\frac{1}{qd\kappa+1}$, check *satisfiability*. This succeeds with probability $1 - \frac{u_s}{R}$.

(ii) W. P. $\frac{qd\kappa}{qd\kappa+1}$, generate $|\phi'\rangle$, and measure the second register in the Hadamard basis. If the output state is not $|+\rangle$, reject. Otherwise, choose a random $k \in [d]$, and perform a Hadamard test with Π_k. This succeeds with probability $\frac{1}{\kappa}(\frac{1}{2} + \frac{1}{2}\mathbb{E}_k[\mathrm{Re}\,\langle\phi|\,\Pi_k\,|\phi\rangle]) = \frac{1}{\kappa}(1 - \frac{u_e}{qdR})$.[10]

The overall success probability of the constraint tests is

$$\frac{1}{qd\kappa+1}(1 - \frac{u_s}{R}) + \frac{qd\kappa}{qd\kappa+1}\left(\frac{1}{\kappa}\cdot(1 - \frac{u_e}{qdR})\right) = \frac{qd+1}{qd\kappa+1} - \frac{u_e + u_s}{R\cdot(qd\kappa+1)}. \tag{3.19}$$

We now show a constant gap between completeness and soundness. In completeness, $u_e = u_s = 0$, so $|\psi\rangle$ passes the constraint tests with probability $C^{\mathrm{YES}} := \frac{qd+1}{qd\kappa+1}$. In soundness, recall that $\mathbf{val}(\mathcal{C}) \le \delta$, so by Claim 71, any *rigid* quantum proof passes the constraint tests with probability at most $C^{\mathrm{YES}} - \frac{1-\delta}{qd\kappa+1}$. We now apply Lemma 70: any quantum proof that passes *Density test* and *Validity test* with probabilities too similar to that in completeness must pass the constraint tests with probability less than C^{YES}.

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

Corollary 73. *In soundness, if* $D(|\psi\rangle) = \frac{1}{\kappa} - d_1$ *and* $V(|\psi\rangle) = 1 - \frac{1}{\kappa} - d_2$*, then*

$$C(|\psi\rangle) \leq C^{\text{YES}} - \frac{1-\delta}{qd\kappa+1} + \left(\kappa d_1 + (\kappa+1)\sqrt{\kappa \cdot d_2}\right)^{1/2} . \tag{3.20}$$

3.4.3 Analysis

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

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

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

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

$$\left(\kappa\varepsilon + (\kappa+1)\sqrt{\kappa \cdot \varepsilon}\right)^{1/2} \leq \frac{\lambda}{2} . \tag{3.21}$$

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

$$p_1 = \frac{1}{2} \cdot \frac{1}{Z} \qquad p_2 = \frac{1}{Z} \qquad p_3 = \frac{\varepsilon}{4(1-C^{\text{YES}})} \cdot \frac{1}{Z} . \tag{3.22}$$

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

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

by *Density test*.

$$P_{\mathrm{NO}} = p_1(\frac{1}{\kappa} - d) + p_2 V(|\psi\rangle) + p_3 C(|\psi\rangle) \tag{3.23}$$
$$\leq P_{\mathrm{YES}} - p_1 d + p_3(1 - C^{\mathrm{YES}}) \tag{3.24}$$
$$\leq P_{\mathrm{YES}} - p_1 \varepsilon + p_3(1 - C^{\mathrm{YES}}) \tag{3.25}$$
$$= P_{\mathrm{YES}} - \frac{\varepsilon}{2Z} + \frac{\varepsilon}{4Z} = P_{\mathrm{YES}} - \frac{\varepsilon}{4Z}\,. \tag{3.26}$$

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

$$P_{\mathrm{NO}} = p_1(\frac{1}{\kappa} + d) + p_2 V(|\psi\rangle) + p_3 C(|\psi\rangle) \tag{3.27}$$
$$\leq p_1(\frac{1}{\kappa} + d) + p_2(1 - \frac{1}{\kappa} - d) + p_3 \tag{3.28}$$
$$= P_{\mathrm{YES}} - (p_2 - p_1)d + p_3(1 - C^{\mathrm{YES}}) \tag{3.29}$$
$$\leq P_{\mathrm{YES}} - (p_2 - p_1)\varepsilon + p_3(1 - C^{\mathrm{YES}}) \tag{3.30}$$
$$= P_{\mathrm{YES}} - \frac{\varepsilon}{2Z} + \frac{\varepsilon}{4Z} = P_{\mathrm{YES}} - \frac{\varepsilon}{4Z}\,, \tag{3.31}$$

where the first inequality follows from Lemma 67.

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

$$P_{\mathrm{NO}} \leq p_1(\frac{1}{\kappa} + |d_1|) + p_2(1 - \frac{1}{\kappa} - d_2) + p_3 \tag{3.32}$$
$$\leq P_{\mathrm{YES}} + p_1|d_1| - p_2 d_2 + p_3(1 - C^{\mathrm{YES}}) \tag{3.33}$$
$$\leq P_{\mathrm{YES}} - (p_2 - p_1)\varepsilon + p_3(1 - C^{\mathrm{YES}}) \tag{3.34}$$
$$= P_{\mathrm{YES}} - \frac{\varepsilon}{4Z}\,. \tag{3.35}$$

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

$$P_{\text{NO}} = p_1(\frac{1}{\kappa} + d_1) + p_2(1 - \frac{1}{\kappa} - d_2) + p_3 C(|\psi\rangle) \tag{3.36}$$

$$\leq P_{\text{YES}} + p_1 d_1 - p_2 d_2 + p_3 \left(-\frac{1-\delta}{qd\kappa + 1} + \left(\kappa d_1 + (\kappa+1)\sqrt{\kappa \cdot d_2} \right)^{1/2} \right) \tag{3.37}$$

$$\leq P_{\text{YES}} + p_1 d_1 - p_2 d_2 - p_3 \frac{\lambda}{2}. \tag{3.38}$$

The first inequality follows from Corollary 73, and the second inequality holds by our choice of ε. Note that $d_2 \geq 0$ by Lemma 66. By Lemma 67, if $d_1 \geq 0$, then $d_1 \leq d_2$; otherwise $d_1 \leq d_2$ trivially. So then

$$P_{\text{NO}} \leq P_{\text{YES}} - (p_2 - p_1) d_2 - p_3 \frac{\lambda}{2} \leq P_{\text{YES}} - \frac{\varepsilon \lambda}{8(1 - C^{\text{YES}})Z}. \tag{3.39}$$

Combining these cases proves the following result:

Theorem 74. *Given an instance of* $(1, \delta)$-GapCSP, *the* $\mathsf{QMA}^+_{\log}$ *protocol succeeds with probability* P^{YES} *in completeness and at most* $P^{\text{YES}} - \Delta$ *in soundness for some constants* $1 > P^{\text{YES}} > \Delta > 0$.

Corollary 75. *There exist constants* $1 > P^{\text{YES}} > \Delta > 0$ *such that* $\mathsf{NP} \subseteq \mathsf{QMA}^+_{\log}$ *with completeness* P^{YES} *and soundness* $P^{\text{YES}} - \Delta$.

3.5 QMA^+ protocol for NEXP

Our goal in this section is to modify the previous protocol to solve an NEXP-complete problem. Again by the PCP theorem, the *succinct* $(1, \delta)$-GapCSP problem with exponentially many variables and clauses is NEXP-complete. The *succinctness* allows us to efficiently describe the problem input. What remains is to ensure that the verifier's protocol is efficient. Previously, the unitary transformations were efficient because the verifier handled poly(n)-

size graphs. Furthermore, the expanders used to check the equality constraints for each variable may have different sizes. Now that there can be exponentially many possibilities for the size of each cluster, naively applying the previous technique is not efficient. These challenges were addressed in Jeronimo and Wu [2023b] by considering a PCP construction for NEXP with strong properties.

Theorem 76 (Jeronimo and Wu [2023b])**.** *There is a* NEXP*-hard* $(1, \delta)$-GAPCSP *instance for some* $(N = 2^{\mathrm{poly}(n)}, R = 2^{\mathrm{poly}(n)}, q = O(1), \Sigma = \{0,1\})$*-CSP system* $\mathcal{C}$ *that is both* τ-strongly uniform *for some constant* τ *and* polylog(NR)-doubly explicit*.*

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

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

Definition 77 (Doubly explicit CSP)**.** *A* (N, R, q, Σ)*-CSP system* $\mathcal{C}$ *is* $Z(N, R)$-doubly explicit *if for all* $i \in [N]$ *and* $j \in [R]$*, the following are computable in time* $Z(N, R)$*:*

(i) Cardinality of $Adj_V(i)$ *and* $Adj_{\mathcal{C}}(j)$ *for all* $i \in [N]$ *and* $j \in [R]$*.*

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

(iii) $Adj^{id}_{\mathcal{C}} : [R] \times [q] \to [N]$*;* $Adj^{id}_{\mathcal{C}}(j, \imath) = i$ *is the* $\imath$*-th variable of* $Adj_{\mathcal{C}}(j)$*.*

(iv) $Adj^{ind}_{V} : [N] \times [R] \to [R]$*; if* i *participates in* $\mathcal{C}_j$*, then* $Adj^{ind}_{V}(i, j) = \jmath$ *is the index of* j *in* $Adj_V(i)$*.*

(v) $Adj^{id}_{V} : [N] \times [R] \to [R]$*;* $Adj^{id}_{V}(i, \jmath) = j$ *is the* $\jmath$*-th variable* $Adj_V(i)$*.*

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

Definition 78 (Strongly uniform CSP)**.** *Let $\tau \in \mathbb{N}$. A (N, R, q, Σ)-CSP system $\mathcal{C}$ is τ-strongly uniform if the variable set $[N]$ can be partitioned into at most τ different subsets $\bigcup_y V_y$ such that $|\,Adj_V(i)| = |\,Adj_V(j)| = n_k$ if i and j belong to the same part V_k. Furthermore, the part $k \in [\tau]$ can be determined in time* $\mathrm{polylog}(NR)$*.*

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

Theorem 79 (Doubly explicit expander graphs Lubotzky [2009], Alon [2021])**.** *There is a constant d such that the following explicit constructions of expander graphs exist:*

1. *For every n, there is a d-regular graph on n vertices.*
2. *For every prime $p > 17$, there is a d-regular graph on $n = p(p^2 - 1)$ vertices, and the graph can be decomposed into d permutations $\pi_1, \ldots, \pi_d$ that can each be evaluated in time* $\mathrm{polylog}(n)$*.*[11]

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

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

Theorem 80 (Primes in short intervals Cheng [2013])**.** *There is an absolute constant k_0 such that for any integer $k > k_0$, there is a prime in the interval $[k - 4k^{2/3}, k]$.*

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

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

3.5.1 Efficient constraint tests

We show how to efficiently implement *consistency* checks that imply a version of Claim 71. Fix any vertex i. Let n be the number of constraints that depend on i and p be the corresponding prime. Let n_0 be a large enough constant. If $n \leq n_0$, then use Construction 1 d-regular expander to wire new copies of the vertex together, just as for NP. Otherwise, use Construction 2 to generate a d-regular expander graph of size $p(p^2-1) \in [n - O(n^{8/9}), n]$ that wires nearly all copies of the vertex together. Then add d self-loops for the remaining vertices. The number of vertices with self loops is at most some ηn (for some small constant η) since $p(p^2-1) \geq n - O(n^{8/9})$; we can make η arbitrarily small by choosing a large enough n_0.

Claim 81 (Jeronimo and Wu [2023b])**.** *Consider a (N, R, q, Σ)-CSP system $\mathcal{C}$, and apply* efficient regularization*. If $\mathbf{val}(\mathcal{C}) = 1$, then all "consistency constraints" can be simultaneously satisfied. If $\mathbf{val}(\mathcal{C}) \leq \delta$, then the total number of unsatisfied constraints is at least $(1 - \delta - q\eta)R$.*

The analysis of Claim 81 is similar to Claim 71. The additional factor of $q\eta$ comes from the self-loop constraints; these can be satisfied without violating any "consistency constraint".

After measuring the primes, let the verifier act on the space $\mathbb{C}^R \otimes \mathbb{C}^\kappa \otimes \mathbb{C}^N \otimes \mathbb{C}^{|\Sigma|}$. We explicitly define the unitary operators that are used in the NP protocol. These definitions exactly match Jeronimo and Wu [2023b]. First, operator $\mathcal{A}$ expands the values from the list

of values of variables involved in each constraint:

$$\mathcal{A} : |j\rangle |v\rangle |0\rangle |0\rangle \to \frac{1}{q} \sum_{r=1}^{q} |j\rangle |\vec{v}\rangle |i_r\rangle |v_r\rangle \ , \tag{3.40}$$

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

$$\mathcal{M}_k : |j\rangle |v\rangle |i\rangle \left|v'\right\rangle \to |\Pi_k(j,i)\rangle |v\rangle |i\rangle \left|v'\right\rangle \tag{3.41}$$

The last operation computes the constraints in superposition:

$$\mathcal{B} |j\rangle |v\rangle |0\rangle \to |j\rangle |v\rangle \left|\mathcal{C}_j(v)\right\rangle \tag{3.42}$$

Theorem 82 (Jeronimo and Wu [2023b])**.** *$\mathcal{A}, \mathcal{B}, \mathcal{M}_k$ can be implemented by* BQP *circuits.*

3.5.2 Analysis

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

(i) For *satisfiability*, the prover computes $\mathcal{B} |\psi\rangle |0\rangle$ and measures the second qubit in the standard basis.

(ii) For *consistency*, the prover computes $\mathcal{A} |\psi\rangle$, selects random $d \in [k]$, and uses $\mathcal{M}_k$ in Hadamard test.

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

$$C(|\psi\rangle) = \frac{qd+1}{qd\kappa+1} - \frac{u_e + u_s}{R \cdot (qd\kappa+1)} \tag{3.43}$$

To be able to analyze soundness, all that is left is to reprove Corollary 73 to handle the subtle difference between Claim 81 and Claim 71. Let $D(|\psi\rangle) = \frac{1}{\kappa} \pm d_1$ and $V(|\psi\rangle) = 1 - \frac{1}{\kappa} - d_2$, then we can write:

$$\begin{aligned} C(|\psi\rangle) &= C^{\text{YES}} - \frac{u_e + u_s}{R \cdot (qd\kappa+1)} && (3.44) \\ &\leq C^{\text{YES}} - \frac{R \cdot (1-\delta-q\eta)}{R \cdot (qd\kappa+1)} + \left(\kappa d_1 + (\kappa+1)\sqrt{\kappa \cdot d_2}\right)^{1/2} . && (3.45) \end{aligned}$$

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

Corollary 83. *If $D(|\psi\rangle) = \frac{1}{\kappa} \pm d_1$ and $V(\psi) = 1 - \frac{1}{\kappa} - d_2$, then:*

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

For soundness, the same analysis of Section 3.4.3 goes through by reducing λ by a factor of 2; this change comes from Corollary 83, which handles the extra $q\eta$ self-loop constraints. All together:

Theorem 84. *Consider $(1,\delta)$-*GapCSP *with a $(N = 2^{\text{poly}(n)}, R = 2^{\text{poly}(n)}, q = O(1), \Sigma = \{0,1\})$-CSP system that is* polylog(NR)*-doubly explicit and $O(1)$-strongly uniform. The* QMA^+ *protocol solves this problem with completeness c and soundness s for some constants $1 > c > s > 0$.*

Corollary 85. *There exist constants $1 > c > s > 0$ such that $\mathsf{NEXP} \subseteq \mathsf{QMA}^+_{c,s}$.*

3.6 Subtle features of QMA^+

3.6.1 Promise symmetry matters

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

Fact 86. *Consider the class* $\mathsf{QMA}^{+'}$*, where the proof must have non-negative amplitudes* only in completeness*. Since subset states have non-negative amplitudes,* $\mathsf{SQMA} \subseteq \mathsf{QMA}^{+'} \subseteq \mathsf{QMA}$*. By Grilo et al. [2014],* $\mathsf{SQMA} = \mathsf{QMA}^{+'} = \mathsf{QMA}$*.*

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

Corollary 87. *Notice that* $\mathsf{QMA}^{+'}_{c,s} \subseteq \mathsf{QMA}^{+}_{c,s}$ *for any choice of* $0 \leq c, s \leq 1$*, since any* $\mathsf{QMA}^{+'}_{c,s}$ *protocol is also a* $\mathsf{QMA}^{+}_{c,s}$ *protocol. Since* $\mathsf{QMA} = \mathsf{QMA}^{+'}$*,* $\mathsf{QMA} \subseteq \mathsf{QMA}^{+}_{c,s}$ *whenever* $c < 1$ *and* $c - s \geq \frac{1}{p(n)}$ *for any polynomial* $p(n)$*.*

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

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

Fact 88. *Consider a Hermitian and positive semidefinite matrix* M*. Let*

$$\|M\|_+ := \max_{|v\rangle \geq 0} \frac{\|M|v\rangle\|_2}{\||v\rangle\|_2}$$

be the maximum value among real non-negative vectors. Then it is possible for $\|M \otimes M\|_+ > \|M\|_+^2$*.*

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

3.6.2 QMA^+ *at some constant gap equals* QMA

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

Proposition 89. $\mathsf{QMA}^+_{c,s} \subseteq \mathsf{QMA}_{c,4s}$.

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

- *(completeness)* The same completeness proof is a valid proof for QMA, accepting with completeness c.

- *(soundness)* Recall that $\langle\chi|\Pi_1|\chi\rangle \leq s$ for any $|\chi\rangle$ with non-negative amplitudes.

  Consider any state $|\psi\rangle$ with *real* (but possibly negative) amplitudes. Separate and normalize its positive entries and negative entries, i.e. $|\psi\rangle = \sqrt{p}\,|\psi_+\rangle - \sqrt{1-p}\,|\psi_-\rangle$. Notice that $\langle\psi_+|\psi_-\rangle = 0$, so $|\psi\rangle$ is of unit norm. Then

$$\begin{aligned}
\langle\psi|\Pi_1|\psi\rangle &= p\langle\psi_+|\Pi_1|\psi_+\rangle + (1-p)\langle\psi_-|\Pi_1|\psi_-\rangle && (3.46)\\
&\quad - \sqrt{p(1-p)}(\langle\psi_-|\Pi_1|\psi_+\rangle + \langle\psi_+|\Pi_1|\psi_-\rangle) && (3.47)\\
&\leq p\langle\psi_+|\Pi_1|\psi_+\rangle + (1-p)\langle\psi_-|\Pi_1|\psi_-\rangle && (3.48)\\
&\quad + 2\sqrt{p(1-p)}\sqrt{\langle\psi_-|\Pi_1|\psi_-\rangle\langle\psi_+|\Pi_1|\psi_+\rangle} && (3.49)\\
&\leq s + 2s\sqrt{p(1-p)} \leq 2s\,, && (3.50)
\end{aligned}$$

  where the first inequality holds by Cauchy-Schwarz because Π_1 is positive semidefinite.

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

□

Corollary 90. *For any* $0 < \varepsilon < 0.2$, $\mathsf{QMA}^{+}_{0.8+\varepsilon,0.2} = \mathsf{QMA}$.

Proof. $\mathsf{QMA}^{+}_{0.8+\varepsilon,0.2} \subseteq \mathsf{QMA}$ follows from Proposition 89. The other direction follows from Corollary 87. □

This is a strange phenomenon: depending on the choice of constants $c > s$, $\mathsf{QMA}^{+}_{c,s}$ could be as small as QMA and as large as NEXP![12] See Figure 1.2 for a pictorial description. An implication of our work is that assuming $\mathsf{EXP} \neq \mathsf{NEXP}$, QMA^{+} simply *cannot* be amplified.

3.7 Open questions

1. **What is the relationship of QMA and $\mathsf{QMA}(2)$?** Our result does not immediately say anything about QMA and $\mathsf{QMA}(2)$. It only suggests that for QMA, the restriction of *relative phase* is maximally strong. For example, it is possible that $\mathsf{QMA}(2) = \mathsf{NEXP}$; i.e. the restriction of *entanglement* across a fixed barrier may be just as powerful. In fact, showing $\mathsf{QMA}(2) = \mathsf{QMA}^{+}(2)$ is still an open route to proving $\mathsf{QMA}(2) = \mathsf{NEXP}$, but in light of this work, amplification for $\mathsf{QMA}^{+}(2)$ must crucially rely on the unentanglement promise.

2. **Are other complexity classes sensitive to different constant-sized promise gaps?** We show that for $\mathsf{QMA}^{+}_{c,s}$, the parameter $\frac{c}{s}$ can be "tuned" to change the power of the class from QMA to NEXP (see also Figure 1.2). Do other complexity classes drastically change power with different promise gaps? One similar class is

12. Note that the same phenomenon holds for $\mathsf{QMA}^{+}(2)$ and $\mathsf{QMA}(2)$ with nearly the same proof. This is why Jeronimo and Wu [2023b] was perceived as "just a constant gap away" from proving $\mathsf{QMA}(2) = \mathsf{NEXP}$.

SBQP Kuperberg [2015], which equals $\mathsf{BQP}_{c,s}$ when $\frac{c}{s} = 2$ (but unlike our work, c and s are exponentially small). However, when $\frac{c}{s}$ is allowed to be any number above 1, $\mathsf{BQP}_{c,s}$ is equal to PP Deshpande et al. [2022]. Note that relative to oracles, SBQP is not closed under intersection, which was used to separate it from QMA Aaronson et al. [2020].

3. **State complexity vs. decision complexity?** Although we prove that there are constants c_1, s_1, c_2, s_2 such that $\mathsf{QMA}^+_{c_1,s_1} = \mathsf{QMA}^+(2)_{c_2,s_2}$, we do not prove the existence of any *product test* (as in Harrow and Montanaro [2013]). In fact, it is possible that no product test exists! This could lead to a separation between the complexity of *decision problems* and *state synthesis problems*; i.e. $\mathsf{QMA}^+ = \mathsf{QMA}^+(2)$ but $\mathsf{stateQMA}^+ \neq \mathsf{stateQMA}^+(2)$. In fact, it is even possible that $\mathsf{QMA} = \mathsf{QMA}(2)$ but $\mathsf{stateQMA} \neq \mathsf{stateQMA}(2)$. This inquiry can help us understand whether (or how) *the power of unentanglement* is useful when solving *decision problems*.

CHAPTER 4

NON-COLLAPSING MEASUREMENTS

This chapter is adapted (with minor modifications) from the work of Roozbeh Bassirian and Kunal Marwaha. Superposition detection and QMA with non-collapsing measurements. *CoRR*, abs/2403.02532, 2024. doi:10.48550/ARXIV.2403.02532. URL `https://doi.org/10.48550/arXiv.2403.02532`.

4.1 Techniques and Contributions

Our techniques primarily follow that of Jeronimo and Wu [2023b], Bassirian et al. [2023b], which are inspired by the work of Blier and Tapp Blier and Tapp [2010]. We describe the general approach and then note the differences in this work.

The central challenge is to place an NEXP-complete problem in a variant of QMA. Following Jeronimo and Wu [2023b], Bassirian et al. [2023b], we use a *succinct* constraint satisfaction problem with *constant* gap. In $(1, \delta)$-GapCSP, either all constraints can be satisfied, or at most a δ fraction of constraints can be satisfied. This problem is NEXP-hard due to the PCP theorem Arora and Safra [1992], Arora et al. [1998a], Harsha [2004].[1]

Consider a witness over two registers: the *constraint index* register with a polynomial number of qubits, and a constant-sized *value* register. The protocol chooses one of two tests to run with some probability. The first test checks the *rigidity* of a quantum state, i.e. accepting if the witness is close to the form $\frac{1}{\sqrt{R}} \sum_{j \in [R]} |j\rangle |\sigma(j)\rangle$. The second test verifies the *constraints* of the CSP by measuring the state in the computational basis and verifying the constraint indexed by the outcome of the first register. Using several tricks from the PCP literature Dinur [2007], Jeronimo and Wu [2023b], the second test is able to decide $(1, \delta)$-GapCSP whenever the witness is *rigid*.

1. As we do not consider NP-completeness in this work, we often omit the word *succinct* when describing $(1, \delta)$-GapCSP.

Surprisingly, the *non-negative amplitudes* guarantee is only used in the *first* test, i.e. checking the *rigidity* of the witness. With some probability, the verifier encourages a large ℓ_1-norm by measuring the overlap with the uniform superposition state $|+\rangle$. It must now enforce that the witness sends at most *one* basis element in the *value* register per basis element in the *constraint index* register. For example, after measuring the *constraint index* register in the computational basis, the verifier must reject any superposition in the remaining quantum state. Jeronimo and Wu [2023b], Bassirian et al. [2023b] can enforce this property assuming the witness has non-negative amplitudes, thus deciding the NEXP-complete problem.

In this work, we consider any algorithm that detects *superposition* over constant-sized computational basis states:

Definition 91 (Superposition detector)**.** *A* superposition detector $\mathsf{SupDetect}_{k,\epsilon,\Delta}$ *is an algorithm that inputs a quantum state* $|\psi\rangle$ *on* k *qubits and outputs* 1 *or* 0*, such that:*

- *If* $|\psi\rangle$ *is a computational basis state on* k *qubits, then* $\mathsf{SupDetect}_{k,\epsilon,\Delta}(|\psi\rangle) = 1$ *with probability* 1*.*
- *If* $|\langle e|\psi\rangle|^2 \leq 1 - \epsilon$ *for every computational basis state* $|e\rangle$ *on* k *qubits, then*

$$\mathsf{SupDetect}_{k,\epsilon,\Delta}(|\psi\rangle) = 1$$

  with probability at most $1 - \Delta$*.*

We show that if a quantum computer can implement a superposition detector over the *value* register of the quantum witness, then it can check *rigidity* of the witness, and thus decide the NEXP-complete problem. Since the *value* register is constant-sized, we only need a superposition detector where k is some constant.

We use this generalization in two places. First, we show that a *single* non-collapsing measurement allows a quantum computer to detect superposition on *any* quantum state, and thus decide any problem in NEXP. This resolves Question 1. Curiously, in both the

verification and advice setting Aaronson [2018], the power of non-collapsing measurements is magnified when the verifier is given a state it cannot prepare.

Second, we show that if the witness has non-negative amplitudes, a quantum computer can efficiently detect superposition; this reproves the statement $\mathsf{QMA}^+ = \mathsf{NEXP}$ Bassirian et al. [2023b]. The *superposition detection* viewpoint gives new intuition for this result: QMA^+ and $\mathsf{QMA}^+(2)$ *pre-select* quantum witnesses where the verifier can detect superposition. However, the verifier cannot detect superposition in *all* quantum witnesses unless $\mathsf{QMA} \neq \mathsf{NEXP}$. As a consequence, any proof of $\mathsf{NEXP} \subseteq \mathsf{QMA}(2)$ must handle *some* witnesses in a fundamentally new way.

We first prove the following statement:

Theorem 92 (Main theorem)**.** *Consider a variant of* QMA *where for any constant* k *and* $\epsilon > 0$*, there exists some* $\Delta = \Omega(\frac{1}{\mathrm{poly}(n)})$ *such that* $\mathsf{SupDetect}_{k,\epsilon,\Delta}$ *can be efficiently applied to the quantum witness after a partial measurement in the computational basis. Then this variant contains* NEXP*.*

This has two direct consequences. First, since a superposition detector can be efficiently implemented on all quantum states if the verifier can make non-collapsing measurements, we resolve Question 1:

Corollary 93. *Consider the variant of* QMA *where the verifier may make* one *non-collapsing measurement of an arbitrary quantum state. Then this variant equals* NEXP*.*

Second, since a superposition detector can be implemented for all quantum witnesses with *non-negative amplitudes*, we recover the main result of Bassirian et al. [2023b]:

Corollary 94 (Bassirian et al. [2023b])**.** $\mathsf{QMA}^+ = \mathsf{NEXP}$.

4.2 Related work

QMA and hidden-variable theories Aaronson initiated the study of quantum complexity subject to hidden-variable theories in Aaronson [2005b]. They introduced the class DQP

(Dynamical Quantum Polynomial-time), and showed evidence that this class was more powerful than BQP. However, DQP has subtleties in its definition which make it particularly challenging to study. To address this, Aaronson et al. [2014] introduced the concept of "non-collapsing" measurement through the classes CQP and naCQP (non-adaptive CQP)[2]. They showed evidence that these classes were only "slightly more powerful" than BQP. By contrast, Aaronson [2018] showed that this power dramatically increased in the presence of quantum advice: naCQP/qpoly = ALL $\neq$ BQP/qpoly. Aharonov and Regev Aharonov and Regev [2003a] define a variant of QMA[3], which can be interpreted as allowing non-collapsing measurements on the accept/reject qubit at the end of the computation. This variant is equal to QMA; see further discussion in Aaronson and Drucker [2013], Aaronson [2023]. Non-collapsing measurements are also connected to "rewindable" quantum computation; see Hiromasa et al. [2023].

The complexity class QMA(2) The power of QMA(2) is not sensitive to the *number* of separable partitions: Harrow and Montanaro [2013] showed that QMA(2) = QMA(k) for k at most polynomial in n using a primitive called the *product test*. One reason to believe that QMA(2) is powerful is that QMA(2) can decide NP using only logarithmic-sized proofs Blier and Tapp [2010]; by contrast, QMA with logarithmic-sized proofs is equal to BQP. However, since the QMA(2) protocol has a subconstant promise gap, it cannot (so far) be scaled to NEXP Chen and Drucker [2010], Chiesa and Forbes [2011], Aaronson et al. [2008], Pereszlényi [2012b], Gall et al. [2012]. If there exist efficient algorithms that "disentangle" quantum states, then QMA = QMA(2); Aaronson et al. [2008], Akibue et al. [2024] show some evidence that no "disentanglers" exist. Although $\mathsf{QMA}^+ = \mathsf{QMA}^+(2) = \mathsf{NEXP}$ Jeronimo and Wu [2023b], Bassirian et al. [2023b], there exist constants c, s where $\mathsf{QMA}^+_{c,s} \subseteq \mathsf{QMA}$ and $\mathsf{QMA}^+(2)_{c,s} \subseteq \mathsf{QMA}(2)$. In fact, there exists a promise gap for $\mathsf{QMA}^+(2)$ where a sharp

2. naCQP was called PDQP (Product Dynamical Quantum Polynomial-time) in a preprint of Aaronson et al. [2014], and again in subsequent work Aaronson [2018], Aaronson et al. [2024b]. We use the name naCQP.

3. Confusingly, this variant is called QMA+, not to be confused with QMA^+ of Jeronimo and Wu [2023b].

transition in complexity occurs (assuming there is a transition) Jeronimo and Wu [2024]; it is unknown if the same is true for QMA^+.

Concurrent work Aaronson et al. [2024b] simultaneously and independently study the power of QMA with non-collapsing measurements. They show that this computational model is equal to NEXP, also resolving Question 1. However, their proof uses $O(n \log n)$ non-collapsing measurements, as opposed to just *one* in this work. Notably, they use different techniques, building on the proofs of $\mathsf{MIP} = \mathsf{NEXP}$ Babai et al. [1990] and $\mathsf{naCQP/qpoly} = \mathsf{ALL}$ Aaronson [2018]. They also explore the power of DQP Aaronson [2005b] in the contexts of quantum verification and quantum advice. Note that one must make minor adjustments to the definitions of CQP and naCQP Aaronson et al. [2014] to allow non-collapsing measurement on a quantum *witness*. Aaronson et al. [2024b] makes a slightly different choice than this work; we discuss this technicality further in Section 4.5.1.

4.3 Our Setup

We first define $\mathsf{QMA}_{c,s}$. When c, s are not specified, $c - s$ is allowed to be at least any inverse polynomial in input size. In general, the promise gap can be amplified via parallel repetition or with more clever methods Marriott and Watrous [2005a], Nagaj et al. [2009].

Definition 95 ($\mathsf{QMA}_{c,s}$)**.** *Let $c, s : \mathbb{N} \to \mathbb{R}^+$ be polynomial-time computable functions. A promise problem $\mathcal{L}_{yes}, \mathcal{L}_{no} \subseteq \{0,1\}^*$ is in $\mathsf{QMA}_{c,s}$ if there exists a BQP verifier V such that for every $n \in \mathbb{N}$ and every $x \in \{0,1\}^n$:*

- ***Completeness:*** *if $x \in \mathcal{L}_{yes}$, then there exists a state $|\psi\rangle$ on $\mathrm{poly}(n)$ qubits, s.t.*

$$\mathbf{Pr}[V(x, |\psi\rangle) \text{ accepts}] \geq c(n)\,. \tag{4.1}$$

- ***Soundness:*** *If $x \in \mathcal{L}_{no}$, then for every state $|\psi\rangle$ on* $\mathrm{poly}(n)$ *qubits, we have*

$$\mathbf{Pr}[V(x, |\psi\rangle) \text{ accepts}] \leq s(n)\,. \tag{4.2}$$

We prove Theorem 92 using the NEXP-complete problem proposed by Jeronimo and Wu [2023b].

Definition 96 (CSP system)**.** *A (N, R, q, Σ)-*CSP system $\mathcal{C}$ *on N variables with values in Σ consists of a set (possibly a multi-set) of R constraints where the arity of each constraint is exactly q.*

Definition 97 (Value of CSP)**.** *The* value *of a (N, R, q, Σ)-CSP system $\mathcal{C}$ is the maximum fraction of satisfiable constraints over all possible assignments $\sigma : [N] \to \Sigma$. The value of $\mathcal{C}$ is denoted* $\mathbf{val}(\mathcal{C})$.

Definition 98 (GapCSP)**.** *The* $(1, \delta)$-GapCSP problem *takes as input a (N, R, q, Σ)-CSP system $\mathcal{C}$. The task is to distinguish whether $\mathcal{C}$ is such that (in completeness) $\mathbf{val}(\mathcal{C}) = 1$ or (in soundness) $\mathbf{val}(\mathcal{C}) \leq \delta$.*

Theorem 99 (Jeronimo and Wu [2023b])**.** *There exist constants q and $\delta < 1$ and a family of succinct $(N_i, R_i, q, \Sigma = \{0, 1\})$-CSP systems $\{\mathcal{C}_i\}_{i \geq 1}$ for which the* $(1, \delta)$-GapCSP *problem is* NEXP*-complete.*

In our proof, we use the same notion of rigidity Definition 143. These states map each constraint to a single assignment in Σ^q. Bassirian et al. [2023b] showed a QMA protocol for $(1, \delta)$-GapCSP, *assuming* the quantum witness is a *rigid* state. In other words, the protocol is sound *only* for rigid quantum witnesses.

Lemma 100 (Bassirian et al. [2023b])**.** *There exist constants $0 < \xi \leq \mathrm{C_{YES}} < 1$ and a* $\mathsf{QMA}_{\mathrm{C_{YES}}, \mathrm{C_{YES}} - \xi}$ *protocol* ConstraintCheck *such that for every* $(1, \delta)$-GapCSP *instance $\mathcal{I}$ guaranteed by Theorem 99,* ConstraintCheck *decides $\mathcal{I}$* assuming *the quantum witness is a* rigid *state.*

Jeronimo and Wu [2023b], Bassirian et al. [2023b] considered QMA^+, a variant of QMA such that the quantum witness always has *non-negative amplitudes*. Given this condition, Bassirian et al. [2023b] designed a protocol to ensure the quantum witness is a *rigid* state, and so $\mathsf{NEXP} \subseteq \mathsf{QMA}^+$. In this work, we design a similar protocol assuming the existence of efficient superposition detectors. We repeatedly make use of the following fact:

Fact 101 (Fuchs and van de Graaf [1998])**.** *Let $0 \leq \Pi \leq \mathbb{I}$ be a positive semi-definite matrix, and let $|\psi_1\rangle$ and $|\psi_2\rangle$ be quantum states such that $|\langle\psi_1|\psi_2\rangle|^2 \geq 1 - d$. Then $|\langle\psi_1|\Pi|\psi_1\rangle - \langle\psi_2|\Pi|\psi_2\rangle| \leq \sqrt{d}$.*

Proof. $|\langle\psi_1|\Pi|\psi_1\rangle - \langle\psi_2|\Pi|\psi_2\rangle| = |\operatorname{Tr}(\Pi(|\psi_1\rangle\langle\psi_1| - |\psi_2\rangle\langle\psi_2|))|$ is at most the trace distance of $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$, which equals $\sqrt{1 - |\langle\psi_1|\psi_2\rangle|^2} \leq \sqrt{d}$. □

4.4 Proof of main theorem

Our protocol depends on R, q, Σ that determine the (N, R, q, Σ)-CSP system $\mathcal{C}$. It considers quantum witnesses over two registers: a *constraint index* register of dimension R and a *value* register of dimension $\kappa := |\Sigma|^q$, i.e.

$$|\psi\rangle = \sum_{j \in [R], x \in \Sigma^q} a_{j,x} |j\rangle |x\rangle \ , \tag{4.3}$$

for some $a_{j,x} \in \mathbb{C}$ so that $\langle\psi|\psi\rangle = 1$. Before describing the protocol, we define two tests, $\mathsf{Density}$ and $\mathsf{QuasiCheck}$, that act on the quantum witness. We use them in tandem to ensure the witness is *rigid*.

Definition 102. $\mathsf{Density}$ *takes as input a quantum state, measures it in the Fourier basis, and accepts iff the observed basis element is $\frac{1}{\sqrt{R \cdot \kappa}} \sum_{j \in [R], x \in \Sigma^q} |j, x\rangle$.*

Definition 103. $\mathsf{QuasiCheck}_{\epsilon,\Delta}$ *takes as input a quantum state, measures the* constraint index *register in the computational basis, then sends the remaining quantum state to a superposition detector $\mathsf{SupDetect}_{\log\kappa,\epsilon,\Delta}$, and accepts iff the output of $\mathsf{SupDetect}_{\log\kappa,\epsilon,\Delta}$ is* 1.

Density is straightforward: it checks the overlap of a quantum state with the uniform superposition state. Let $A(\mathsf{V}, |\psi\rangle)$ be the acceptance probability of test V when input the quantum state $|\psi\rangle$.

Fact 104. $A(\mathsf{Density}, |\psi\rangle) = |\langle +|\psi\rangle|^2$. *Moreover, if* $|\psi\rangle$ *is* rigid, *then* $A(\mathsf{Density}, |\psi\rangle) = \frac{1}{\kappa}$.

$\mathsf{QuasiCheck}_{\epsilon,\Delta}$ collapses the *constraint index* register of the input, and ensures that there is no superposition over the *value* register. A quasirigid state perfectly satisfies this test.

Fact 105. *If* $|\psi\rangle$ *is* quasirigid, *then* $A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) = 1$. *For arbitrary* $|\psi\rangle$,

$$A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) \leq \sum_{j\in[R]} \left(\sum_{x\in\Sigma^q} |a_{j,x}|^2 \right) \cdot c_{j,\epsilon} \tag{4.4}$$

$$= (1-\Delta) + \Delta \cdot \sum_{j\in[R];c_{j,\epsilon}=1} \left(\sum_{x\in\Sigma^q} |a_{j,x}|^2 \right), \tag{4.5}$$

where $c_{j,\epsilon} = 1-\Delta$ *if* $\max_{x\in\Sigma^q} |a_{j,x}|^2 \leq (1-\epsilon)\sum_{x\in\Sigma^q} |a_{j,x}|^2$, *and* 1 *otherwise.*

In a formal sense, $\mathsf{QuasiCheck}_{\epsilon,\Delta}$ ensures that the input is close to a quasirigid state:

Lemma 106. *Suppose* $A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) = w$. *Then there exists a* quasirigid *state* $|\phi\rangle$ *such that* $|\langle\psi|\phi\rangle|^2 \geq (1-\epsilon)\frac{w-(1-\Delta)}{\Delta}$.

Proof. Choose a map $f : [R] \to \Sigma^q$ so that for every clause $j \in [R]$, $f(j)$ is an element of Σ^q that has the largest weight; i.e. $|a_{j,f(j)}|^2 = \max_{x\in\Sigma^q} |a_{j,x}|^2$. Consider the *quasirigid* state $|\phi\rangle = \frac{1}{\sqrt{\gamma}} \sum_{j\in[R]} a_{j,f(j)} |j\rangle |f(j)\rangle$, where $\gamma = \sum_{j\in[R]} |a_{j,f(j)}|^2$ is chosen so that $\langle\phi|\phi\rangle = 1$. Then $|\langle\psi|\phi\rangle|^2 = \left| \frac{1}{\sqrt{\gamma}} \sum_{j\in[R]} |a_{j,f(j)}|^2 \right|^2 = \gamma$. Finally, notice that

$$\gamma \geq \sum_{j\in[R];c_{j,\epsilon}=1} |a_{j,f(j)}|^2 \geq \sum_{j\in[R];c_{j,\epsilon}=1} (1-\epsilon) \sum_{x\in\Sigma^q} |a_{j,x}|^2 \geq (1-\epsilon)\cdot\frac{w-(1-\Delta)}{\Delta}. \qquad \square$$

Since quasirigid states only use a $\frac{1}{\kappa}$ fraction of basis elements, they will not test well on Density. In fact, no quantum state can do well on both Density and $\mathsf{QuasiCheck}_{\epsilon,\Delta}$:

Lemma 107. *Suppose* $A(\mathsf{Density}, |\psi\rangle) = w_{\mathsf{D}} \geq \frac{1}{\kappa}$ *and* $A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) = w_{\mathsf{Q}}$. *Then* $(w_{\mathsf{D}} - \frac{1}{\kappa})^2 + (1-\epsilon)\frac{w_{\mathsf{Q}} - (1-\Delta)}{\Delta} \leq 1$.

Proof. By Lemma 106, there exists a quasirigid state $|\phi\rangle$ such that

$$|\langle\psi|\phi\rangle|^2 \geq (1-\epsilon)\frac{w_{\mathsf{Q}} - (1-\Delta)}{\Delta}.$$

Let $|\phi\rangle = \sum_{j\in[R]} b_j |j\rangle |\sigma(j)\rangle$ for some function σ and amplitudes $\{b_j\}_{j\in[R]}$. Using Fact 101, we see that $\left| |\langle+|\psi\rangle|^2 - |\langle+|\phi\rangle|^2 \right|$ is at most $\sqrt{1-(1-\epsilon)\frac{w_{\mathsf{Q}}-(1-\Delta)}{\Delta}}$. Note that $|\langle+|\psi\rangle|^2 = w_{\mathsf{D}}$, and by Cauchy-Schwarz, $|\langle+|\phi\rangle|^2 = \frac{1}{R\cdot\kappa}\left|\sum_{j\in[R]} b_j\right|^2 \leq \frac{1}{\kappa}\sum_{j\in[R]}|b_j|^2 = \frac{1}{\kappa}$. Since $w_{\mathsf{D}} \geq \frac{1}{\kappa}$, we have $w_{\mathsf{D}} - \frac{1}{\kappa} \leq |\langle+|\psi\rangle|^2 - |\langle+|\phi\rangle|^2 \leq \sqrt{1-(1-\epsilon)\frac{w_{\mathsf{Q}}-(1-\Delta)}{\Delta}}$. The statement follows after rearranging terms. □

We show that the *best* quantum states for this pair of tests are close to *rigid* states. Intuitively, the quasirigid states that test well on Density have nearly uniform amplitudes, and so are close to rigid.

Claim 108. *Suppose* $A(\mathsf{Density}, |\psi\rangle) \geq \frac{1}{\kappa} - d_{\mathsf{D}}$ *and* $A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) \geq 1 - \Delta\cdot d_{\mathsf{Q}}$. *Then there exists a rigid state* $|\chi\rangle$ *such that* $|\langle\chi|\psi\rangle|^2 \geq 1 - \kappa d_{\mathsf{D}} - (\kappa+1)\sqrt{\epsilon + d_{\mathsf{Q}}}$.

Proof. By Lemma 106, there exists a quasirigid state $|\phi\rangle$ such that $|\langle\psi|\phi\rangle|^2 \geq (1-\epsilon)(1-d_{\mathsf{Q}})$. By Fact 101, we have $\left| |\langle\mu|\psi\rangle|^2 - |\langle\mu|\phi\rangle|^2 \right| \leq \sqrt{1-(1-\epsilon)(1-d_{\mathsf{Q}})} \leq \sqrt{\epsilon + d_{\mathsf{Q}}}$ for any unit vector $|\mu\rangle$. We use this in two places. First, since $|\langle+|\psi\rangle|^2 \geq \frac{1}{\kappa} - d_{\mathsf{D}}$, applying Fact 101 with $|\mu\rangle = |+\rangle$ implies $|\langle+|\phi\rangle|^2 \geq \frac{1}{\kappa} - d_{\mathsf{D}} - \sqrt{\epsilon + d_{\mathsf{Q}}}$. Now let $|\chi\rangle$ be the *rigid* state with the same computational basis elements as $|\phi\rangle$. Then

$$|\langle\chi|\phi\rangle|^2 = \kappa|\langle+|\phi\rangle|^2 \geq 1 - \kappa\left(d_{\mathsf{D}} + \sqrt{\epsilon + d_{\mathsf{Q}}}\right). \tag{4.6}$$

Applying Fact 101 using $|\mu\rangle = |\chi\rangle$ implies $|\langle\chi|\psi\rangle|^2 \geq 1 - \kappa\left(d_{\mathsf{D}} + \sqrt{\epsilon + d_{\mathsf{Q}}}\right) - \sqrt{\epsilon + d_{\mathsf{Q}}}$. □

We are now ready to prove Theorem 92.

Proof of Theorem 92. Suppose for any constant k and $\epsilon > 0$, there exists some Δ at least inverse polynomial in input size such that $\mathsf{SupDetect}_{k,\epsilon,\Delta}$ can be efficiently implemented. We construct the following QMA verifier that decides every GapCSP instance guaranteed by Theorem 99, choosing p_1, p_2, p_3, ϵ later on:

- With probability p_1, run $\mathsf{Density}$.
- With probability p_2, run $\mathsf{QuasiCheck}_{\epsilon,\Delta}$.
- With probability p_3, run the protocol $\mathsf{ConstraintCheck}$ guaranteed by Lemma 100.

Lemma 100 implies the existence of absolute constants $\mathrm{C_{YES}}, \xi, \kappa$. We choose ϵ and *distance thresholds* $\nu_{\mathrm{low}}, \nu_{\mathrm{high}}$ to be small positive constants that satisfy the following conditions:

$$\epsilon < \nu_{\mathrm{high}}^2 \leq \nu_{\mathrm{low}}^2 \leq 1 \quad \frac{\nu_{\mathrm{high}}}{\nu_{\mathrm{low}}} \leq \frac{\xi}{6(1-\mathrm{C_{YES}})} \quad \left(\kappa\nu_{\mathrm{low}} + (\kappa+1)\sqrt{\epsilon+\nu_{\mathrm{low}}}\right)^{1/2} \leq \frac{\xi}{2} \tag{4.7}$$

For example, this can be done by first making all constants equal and small enough to satisfy the right-most inequality, then reducing ν_{high} to satisfy the middle inequality, then reducing ϵ to satisfy the first inequality.

Now we choose the probabilities p_1, p_2, p_3. Let

$$p_1 = \frac{1}{Z} \qquad p_2 = \frac{(\nu_{\mathrm{low}} + \nu_{\mathrm{high}})(1-\epsilon)}{\Delta(\nu_{\mathrm{high}}^2 - \epsilon)Z} \qquad p_3 = \frac{\nu_{\mathrm{low}}}{2(1-\mathrm{C_{YES}})Z}\,, \tag{4.8}$$

where $Z := 1 + \frac{(\nu_{\mathrm{low}}+\nu_{\mathrm{high}})(1-\epsilon)}{\Delta(\nu_{\mathrm{high}}^2-\epsilon)} + \frac{\nu_{\mathrm{low}}}{2(1-\mathrm{C_{YES}})}$, so that the probabilities sum to 1.

Given a quantum witness $|\psi\rangle$, the verifier accepts with probability

$$p_1 \cdot A(\mathsf{Density}, |\psi\rangle) + p_2 \cdot A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) + p_3 \cdot A(\mathsf{ConstraintCheck}, |\psi\rangle)\,. \tag{4.9}$$

In completeness, we consider the quantum witness guaranteed by Lemma 100. Since this state is rigid, by Fact 104 and Fact 105, the verifier accepts with probability at least $\mathrm{P_{YES}} :=$

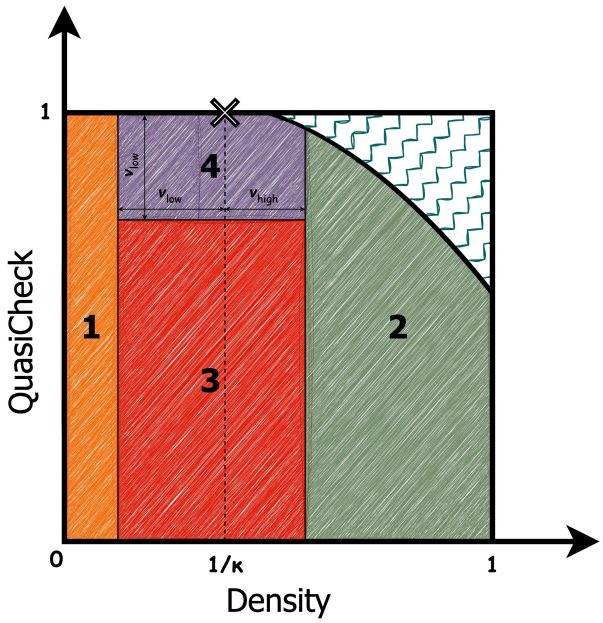

Figure 4.1: Plot of possible acceptance probabilities for the tests Density and $\mathsf{QuasiCheck}_{\epsilon,\Delta}$ when $\Delta = 1$. The numbers 1 to 4 correspond to the four cases in soundness in the proof of Theorem 92. The upper right corner is forbidden by Lemma 107. Rigid states live at the black "×" icon by Fact 104 and Fact 105.

$p_1 \cdot \frac{1}{\kappa} + p_2 \cdot 1 + p_3 \cdot \mathrm{C_{YES}}$.

We split soundness into four cases, depending on the acceptance probabilities of Density and QuasiCheck. We give a graphical picture in Figure 4.1.

1. In the first case, the witness doesn't test well enough on Density. If $A(\mathsf{Density}, |\psi\rangle) = \frac{1}{\kappa} - d$ for $d \geq \nu_{\mathrm{low}}$, then the verifier accepts with probability at most

$$\mathrm{P_{NO}} \leq p_1 \cdot (\frac{1}{\kappa} - d) + p_2 \cdot 1 + p_3 \leq \mathrm{P_{YES}} - p_1 \cdot \nu_{\mathrm{low}} + p_3 \cdot (1 - \mathrm{C_{YES}}) \tag{4.10}$$

$$= \mathrm{P_{YES}} - \frac{\nu_{\mathrm{low}}}{2Z}\,. \tag{4.11}$$

2. In the second case, the witness tests too well on Density, and so cannot test well enough on QuasiCheck. If $A(\mathsf{Density}, |\psi\rangle) = \frac{1}{\kappa} + d$ for $d \geq \nu_{\mathrm{high}}$, then by Lemma 107, $A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) \leq 1 - \Delta + \Delta \cdot \frac{1-d^2}{1-\epsilon}$:

$$\mathrm{P_{NO}} \leq p_1 \cdot \left(\frac{1}{\kappa} + d\right) + p_2 \cdot \left(1 - \Delta + \Delta \cdot \frac{1-d^2}{1-\epsilon}\right) + p_3 \tag{4.12}$$

$$\leq \mathrm{P}_{\mathrm{YES}} + p_1 \cdot d - p_2 \Delta \cdot \left(\frac{d^2 - \epsilon}{1 - \epsilon} \right) + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) \,. \tag{4.13}$$

The right-most expression depends on d through the term $p_1 \cdot d - p_2 \Delta \cdot \left(\frac{d^2 - \epsilon}{1 - \epsilon} \right) = p_1 \left(d - \frac{(\nu_{\mathrm{low}} + \nu_{\mathrm{high}})(d^2 - \epsilon)}{\nu_{\mathrm{high}}^2 - \epsilon} \right)$. This term is decreasing with d when $d \geq \nu_{\mathrm{high}}$; one can verify this by taking a partial derivative with respect to d. So the expression is maximized at $d = \nu_{\mathrm{high}}$, and so

$$\mathrm{P}_{\mathrm{NO}} \leq \mathrm{P}_{\mathrm{YES}} + p_1 \left(\nu_{\mathrm{high}} - \frac{(\nu_{\mathrm{low}} + \nu_{\mathrm{high}})(\nu_{\mathrm{high}}^2 - \epsilon)}{\nu_{\mathrm{high}}^2 - \epsilon} \right) + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) \tag{4.14}$$

$$= \mathrm{P}_{\mathrm{YES}} - \frac{\nu_{\mathrm{low}}}{2Z} \,. \tag{4.15}$$

3. In the third case, the witness doesn't test well enough on $\mathsf{QuasiCheck}$, and the acceptance probability of $\mathsf{Density}$ is not too high. If $A(\mathsf{Density}, |\psi\rangle) = \frac{1}{\kappa} + d$ for $-\nu_{\mathrm{low}} \leq d \leq \nu_{\mathrm{high}}$ and $A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) \leq 1 - \Delta \cdot \nu_{\mathrm{low}}$, then

$$\mathrm{P}_{\mathrm{NO}} \leq p_1 \cdot \left(\frac{1}{\kappa} + d \right) + p_2 \cdot (1 - \Delta \cdot \nu_{\mathrm{low}}) + p_3 \tag{4.16}$$

$$\leq \mathrm{P}_{\mathrm{YES}} + p_1 \cdot d - p_2 \Delta \cdot \nu_{\mathrm{low}} + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) \,. \tag{4.17}$$

By our choice of constants, $d \leq \nu_{\mathrm{high}} \leq \nu_{\mathrm{low}}$. Moreover, one can verify that $p_2 \Delta \geq 2 \cdot p_1$ whenever $\nu_{\mathrm{high}} \leq 1$. So $p_1 \cdot d - p_2 \Delta \cdot \nu_{\mathrm{low}} \leq -p_1 \cdot \nu_{\mathrm{low}}$, and $\mathrm{P}_{\mathrm{NO}} \leq \mathrm{P}_{\mathrm{YES}} - p_1 \cdot \nu_{\mathrm{low}} + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) = \mathrm{P}_{\mathrm{YES}} - \frac{\nu_{\mathrm{low}}}{2Z}$.

4. In the fourth case, the witness tests well on $\mathsf{QuasiCheck}$, and the acceptance probability of $\mathsf{Density}$ is not too low. So the witness must be close to *rigid*, and do worse on $\mathsf{ConstraintCheck}$. If $A(\mathsf{Density}, |\psi\rangle) = \frac{1}{\kappa} + d$ for $-\nu_{\mathrm{low}} \leq d \leq \nu_{\mathrm{high}}$ and $A(\mathsf{QuasiCheck}_{\epsilon,\Delta}, |\psi\rangle) \geq 1 - \Delta \cdot \nu_{\mathrm{low}}$, then by Claim 108, there is a rigid state $|\chi\rangle$ such that $|\langle \chi | \psi \rangle|^2 \geq 1 - \kappa d - (\kappa + 1)\sqrt{\epsilon + \nu_{\mathrm{low}}}$. We use Fact 101 with Π equal to the accepting projector of $\mathsf{ConstraintCheck}$; then $|A(\mathsf{ConstraintCheck}, |\psi\rangle) -$

$A(\mathsf{ConstraintCheck}, |\chi\rangle)| \leq (\kappa d + (\kappa+1)\sqrt{\epsilon+\nu_{\text{low}}})^{1/2}$. Putting this together,

$$\mathrm{P_{NO}} \leq p_1 \cdot \left(\frac{1}{\kappa} + d\right) + p_2 \cdot 1 + p_3 \cdot \left(\mathrm{C_{YES}} - \xi + \left(\kappa d + (\kappa+1)\sqrt{\epsilon+\nu_{\text{low}}}\right)^{1/2}\right) \tag{4.18}$$

$$\leq \mathrm{P_{YES}} + p_1 \cdot d + p_3 \cdot \left(-\xi + \left(\kappa d + (\kappa+1)\sqrt{\epsilon+\nu_{\text{low}}}\right)^{1/2}\right) \tag{4.19}$$

$$\leq \mathrm{P_{YES}} + p_1 \cdot \nu_{\text{high}} - p_3 \cdot \frac{\xi}{2} \tag{4.20}$$

$$= \mathrm{P_{YES}} + \frac{1}{Z}\left(\nu_{\text{high}} - \frac{\nu_{\text{low}} \cdot \xi}{4(1-\mathrm{C_{YES}})}\right) \tag{4.21}$$

$$\leq \mathrm{P_{YES}} - \frac{\nu_{\text{high}}}{2Z}\,. \tag{4.22}$$

So, this protocol succeeds with completeness $\mathrm{P_{YES}}$ and soundness at most $\mathrm{P_{YES}} - \frac{\nu_{\text{high}}}{2Z}$. □

4.5 Implications

4.5.1 QMA *with non-collapsing measurements equals* NEXP

The ability to make non-collapsing measurements implies efficient superposition detection.

Claim 109. *For any constant k and $\epsilon > 0$, a quantum computer with the ability to make* just one *non-collapsing measurement can efficiently implement* $\mathsf{SupDetect}_{k,\epsilon,2(x-x^2)}$ *where* $x = \min(\epsilon, \frac{1}{2^k})$.

Proof. Suppose the quantum computer is given a k-qubit state $|\psi\rangle$, and choose any $\epsilon > 0$. The computer first makes one non-collapsing measurement of $|\psi\rangle$. The computer then makes a (collapsing) measurement of $|\psi\rangle$ in the computational basis, and outputs 1 if and only if the measurements agree.

If $|\psi\rangle$ is truly a computational basis state, then the computer always outputs 1. Now suppose $|\langle e|\psi\rangle|^2 \leq 1-\epsilon$ for all k-qubit computational basis states $|e\rangle \in \mathcal{B}_k$. Let $|e_*\rangle$ be a state in $\mathcal{B}_k$ with the largest squared overlap with $|\psi\rangle$, and let $t = |\langle e_*|\psi\rangle|^2$. Then the

probability of outputting 1 can be upper-bounded as

$$\sum_{|e\rangle\in\mathcal{B}_k}\left(|\langle e|\psi\rangle|^2\right)^2 = |\langle e_*|\psi\rangle|^4 + \sum_{|e\rangle\in\mathcal{B}_k;|e\rangle\neq|e_*\rangle}|\langle e|\psi\rangle|^4 \tag{4.23}$$

$$\leq t^2 + \left(\sum_{|e\rangle\in\mathcal{B}_k;|e\rangle\neq|e_*\rangle}|\langle e|\psi\rangle|^2\right)^2 = t^2 + (1-t)^2\,, \tag{4.24}$$

which equals $1-2(t-t^2)$. This is maximized at the largest and smallest valid values of t. Observe that $t \geq \frac{1}{2^k}$ since $|e_*\rangle$ has the largest squared overlap with $|\psi\rangle$, and $t \leq 1-\epsilon$ by supposition. □

Corollary 93 is then a direct consequence of Theorem 92 and Claim 109. This resolves Question 1.

Aaronson et al. [2014] defined two complexity classes with non-collapsing measurement: CQP and naCQP (non-adaptive CQP). In naCQP, the non-collapsing measurement results can only be used at the *end* of the quantum computation (i.e. in classical post-processing). We remark that both Claim 109 and the proof of Theorem 92 satisfy this non-adaptive condition.

The original framing of Question 1 in Aaronson [2023] directly considers the power of QMA with a naCQP (or CQP) verifier. This has a technical issue which we discuss here. Aaronson et al. [2014] considered a *classical* oracle $\mathcal{O}$ that receives a quantum circuit and outputs non-collapsing measurement results. naCQP (CQP) is then defined as P with 1 (poly(n)) query to $\mathcal{O}$. If we use such a machine as a verifier, it equals NEXP only if $\mathcal{O}$ can also receive the *quantum* witness. To handle this issue, we consider QMA where the verifier can make a non-collapsing measurement on an arbitrary quantum state at *any* point in the computation. Here, Claim 109 uses one non-collapsing measurement, which is immediately followed by one collapsing measurement.

Aaronson et al. [2024b] instead defines the class PDQMA, where the verifier first runs a quantum circuit (with collapsing measurements), then makes *all* non-collapsing measure-

ments, and finally runs a classical circuit given the non-collapsing measurement results. In this model, Claim 109 is not valid as written, since non-collapsing measurements must occur after all collapsing measurements. However, by replacing the final collapsing measurement with a non-collapsing measurement, we recover $\mathsf{NEXP} \subseteq \mathsf{PDQMA}$ using *two* non-collapsing measurements. We remark that this is optimal: PDQMA with one non-collapsing measurement is equal to QMA, since the verifier's final measurement can always be collapsing. Aaronson et al. [2024b] uses $O(n \log n)$ non-collapsing measurements to prove $\mathsf{NEXP} \subseteq \mathsf{PDQMA}$.

4.5.2 $\mathsf{QMA}^+ = \mathsf{NEXP}$

If a quantum witness has *non-negative amplitudes*, a partial measurement in the computational basis will preserve this property. A BQP machine can implement superposition detection on any such quantum state:

Claim 110. *Consider the algorithm* Verify_k^+*, which given a* k*-qubit quantum state* $|\psi\rangle$*, measures* $|\psi\rangle$ *in the Fourier basis and outputs* 0 *if the observed basis component is* $\frac{1}{2^{k/2}} \sum_{i \in [2^k]} |i\rangle$ *and* 1 *otherwise. Then for every* $\epsilon > 0$*, there is a superposition detector* $\mathsf{SupDetect}_{k,\epsilon,\sqrt{\epsilon/2^k}}$ *such that for every state* $|\psi\rangle$ *with non-negative amplitudes in the computational basis:*

$$A(\mathsf{Verify}_k^+, |\psi\rangle) = A(\mathsf{SupDetect}_{k,\epsilon,\sqrt{\epsilon/2^k}}, |\psi\rangle) - \frac{1}{2^k}$$

.

Proof. Let the input be $|\psi\rangle = \sum_{i \in [2^k]} a_i |i\rangle$. Then $A(\mathsf{Verify}_k^+, |\psi\rangle) = 1 - \frac{1}{2^k} \left| \sum_{i \in [2^k]} a_i \right|^2$. Since all $a_i \geq 0$,

$$A(\mathsf{Verify}_k^+, |\psi\rangle) = 1 - \frac{1}{2^k} \Big(\sum_{i \in [2^k]} a_i \Big)^2 = 1 - \frac{1}{2^k} - \sum_{i,j \in [2^k]; i \neq j} a_i a_j \leq 1 - \frac{1}{2^k}\,, \tag{4.25}$$

and this is achieved if $|\psi\rangle$ is a computational basis state. Now, let $i^* = \mathrm{argmax}_{i \in [2^k]} a_i^2$.

Since $\sum_{i\in[2^k]} a_i^2 = 1$, $a_{i^*} \geq \frac{1}{2^{k/2}}$. If $a_{i^*}^2 \leq 1-\epsilon$, then

$$\sum_{i,j\in[2^k];i\neq j} a_i a_j \geq \frac{1}{2^{k/2}} \sum_{j\in[2^k],j\neq i^*} a_j \geq \frac{1}{2^{k/2}} \sqrt{\sum_{j\in[2^k],j\neq i^*} a_j^2} \geq \frac{\sqrt{\epsilon}}{2^{k/2}}\,. \qquad \square$$

Corollary 94 is then a direct consequence of Theorem 92 and Claim 110. Note that Verify_k^+ acts as a superposition detector, since the additive error $\frac{1}{2^k}$ in Claim 110 does not change the promise gap of the protocol in Theorem 92.

In QMA^+, one *only* considers quantum witnesses which admit efficient superposition detection by Claim 110. By contrast, Theorem 92 implies that this property cannot be tested on *all* quantum witnesses, assuming $\mathsf{QMA} \neq \mathsf{NEXP}$. In fact, it is information-theoretically *impossible*, assuming the density matrix is "all there is" in quantum mechanics:

Fact 111. *Fix k. Consider the set S_1 of k-qubit computational basis states and the set S_2 of k-qubit Fourier basis states. Then a uniformly random choice from S_1 has the same density matrix as a uniformly random choice from S_2. Despite this, $\mathsf{SupDetect}_{k,\epsilon,\Delta}$ distinguishes these two options for all $\epsilon \leq 1 - \frac{1}{2^k}$ and $\Delta > 0$.*

Does it help to have more quantum witnesses? Note that having *two* copies of a quantum state allows one to detect superposition: measure each copy in the computational basis, and output 1 iff the outcomes agree. However, in Theorem 92, the superposition detector is used on a quantum witness *after a partial measurement*. In completeness, even with poly(n) quantum witnesses, there is an inverse exponential chance of receiving two copies of the same post-measurement state. So as is, we cannot use the additional quantum witnesses to detect superposition unless we allow an inverse exponential promise gap, or *post-selection* Aaronson [2004]. We observe that Theorem 92 neatly recovers the results precise$\mathsf{QMA}(2) = \mathsf{NEXP}$ Pereszlényi [2012b] and post$\mathsf{QMA}(2) = \mathsf{NEXP}$ Kinoshita [2018], as well as the $\mathsf{QMA}(2)$ protocol with logarithmic-sized proofs for NP Blier and Tapp [2010].

In light of this, we suggest that lower bounds to $\mathsf{QMA}(2)$ must go beyond the QMA^+ framework, as the verifier must be able to handle quantum witnesses without directly de-

tecting superposition.

4.6 Open problems

1. Non-collapsing measurement is especially powerful with access to an inefficient quantum state: QMA with non-collapsing measurements equals NEXP, but QCMA with non-collapsing measurements is in PSPACE. Can one use this concept to construct a classical oracle separating QMA and QCMA? Similarly, how does the power of CQP/poly compare to the power of CQP/qpoly = ALL Aaronson [2018]?

2. The position of the non-collapsing measurement in a quantum computation seems to matter in Corollary 93: non-collapsing measurement is used *after* a collapsing partial measurement. This stymies one approach to prove QMA(2) = NEXP; see discussion in Section 4.5.2. What is the power of BQP (or QMA) with the ability to make non-collapsing measurements, but only *before* all collapsing measurements?

CHAPTER 5

INTERNALLY SEPARABLE PROOFS

This chapter is adapted (with minor modifications) from the work of Roozbeh Bassirian, Bill Fefferman, Itai Leigh, Kunal Marwaha, and Pei Wu. Quantum merlin-arthur with an internally separable proof. *CoRR*, abs/2410.19152, 2024a. doi:10.48550/ARXIV.2410.19152. URL `https://doi.org/10.48550/arXiv.2410.19152`.

5.1 Techniques and Contributions

Consider the complexity class $\mathsf{QMA}_{\mathsf{IS}}$, a variant to QMA where the quantum proof must have a fixed, separable *subsystem*: that is, after tracing out all qubits beyond the subsystem, a predetermined set of $O(1)$ qubits are *not* entangled with the rest of the subsystem. Unlike in $\mathsf{QMA}(2)$, this guarantee refers to *internal* separability: the two parts of the subsystem can be entangled, but only if the entanglement disappears when the subsystem is isolated. See Figure 5.1 for a visual comparison. We restrict one part of the subsystem to $O(1)$ qubits to zero in on the effect of *internal* entanglement structure; without the restriction, this class trivially contains $\mathsf{QMA}(2)$.

We first show that $\mathsf{QMA}_{\mathsf{IS}}$, even with inverse exponentially small gap, is contained in EXP:

Theorem 112 (Upper bound)**.** *For any polynomial p and $1 > c > s > 0$ such that $c - s > \frac{1}{2^{p(n)}}$, we have $\mathsf{QMA}_{\mathsf{IS}}^{c,s} \subseteq \mathsf{EXP}$, where c, s are the completeness and soundness parameters, respectively.*[1]

We also define $\mathsf{QMA}_{\mathsf{IS}}(2)$, where the verifier receives *two* unentangled proofs, each with the above guarantee. In contrast with Theorem 112, we find a constant-sized gap where

1. We formally define $\mathsf{QMA}_{\mathsf{IS}}$ in Section 5.4. In a $\mathsf{QMA}_{\mathsf{IS}}^{c,s}$ protocol, YES instances are accepted with probability at least c, and NO instances are accepted with probability at most s.

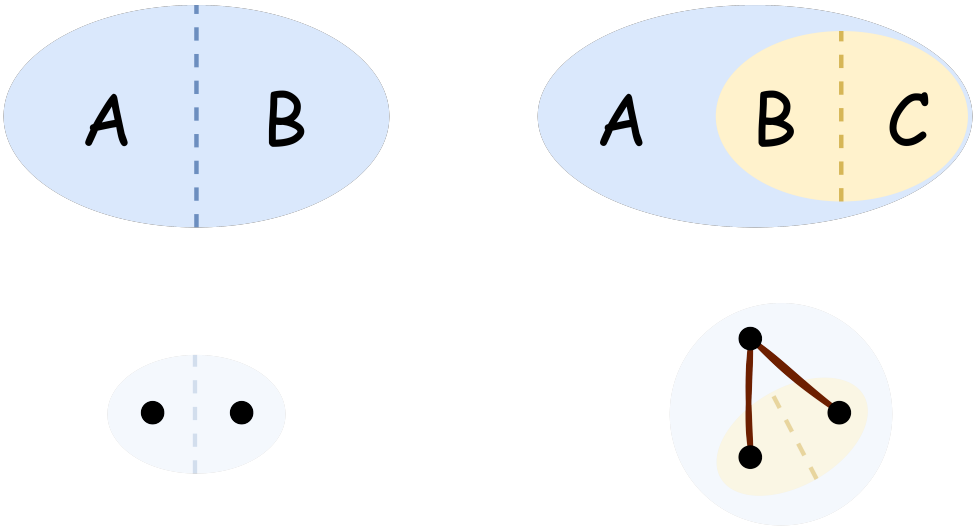

Figure 5.1: Cartoons of a separable state and a state with a separable subsystem. The left side depicts a separable state, where parts A and B share no entanglement. The right side depicts a state where *after* tracing out part A, parts B and C share no entanglement; this state is in general entangled across every bipartition. The lower cartoons each display a graph where every vertex is a part (A, B, or C), and an edge between two parts X, Y allows bipartite entanglement in the reduced state ρ_{XY}. On the right side, part A is entangled with both part B and part C.

$\mathsf{QMA}_{\mathsf{IS}}(2)$ equals NEXP. As a consequence, $\mathsf{QMA}_{\mathsf{IS}}$ and $\mathsf{QMA}_{\mathsf{IS}}(2)$ cannot be equal unless $\mathsf{EXP} = \mathsf{NEXP}$.

Theorem 113 (Lower bound)**.** *There exist constants* $1 > c > s > 0$ *where* $\mathsf{QMA}_{\mathsf{IS}}^{c,s}(2) = \mathsf{NEXP}$.

Corollary 114. *If* $\mathsf{EXP} \neq \mathsf{NEXP}$*, then* $\mathsf{QMA}_{\mathsf{IS}} \neq \mathsf{QMA}_{\mathsf{IS}}(2)$.

We summarize these results in Table 5.1. Theorem 113 also implies that a quantum verifier with a specific guarantee on entanglement structure is *exponentially* more powerful than a classical verifier.

Corollary 114 provides a new route to characterize the power of $\mathsf{QMA}(2)$. For any constant k, $\mathsf{QMA}_{\mathsf{IS}}(k)$ at large enough completeness-soundness gap is equal to $\mathsf{QMA}(k)$. This suggests that the guarantee of internal separability is not too strong.

Claim 115. *For any constant* k*, there exist constants* $1 > c > s > 0$ *so that* $\mathsf{QMA}_{\mathsf{IS}}^{c,s}(k) = \mathsf{QMA}(k)$.

By Claim 115 and Theorem 113, gap amplification of $\mathsf{QMA}_{\mathsf{IS}}(2)$ would imply $\mathsf{QMA}(2) = \mathsf{NEXP}$:

Corollary 116. *If* $\mathsf{QMA}_{\mathsf{IS}}(2)$ *exhibits gap amplification, then* $\mathsf{QMA}(2) = \mathsf{NEXP}$.

Although the same is true of $\mathsf{QMA}^+(2)$ Jeronimo and Wu [2023a], we now know that the class QMA^+ at some completeness-soundness gap is also equal to NEXP, while it is equal to QMA at a larger gap. As a consequence, QMA^+ *cannot* have gap amplification unless $\mathsf{QMA} = \mathsf{NEXP}$ Bassirian et al. [2024b]. This limits the techniques available to prove gap amplification of $\mathsf{QMA}^+(2)$ in order to show $\mathsf{QMA}(2) = \mathsf{NEXP}$.

By contrast, gap amplification of $\mathsf{QMA}_{\mathsf{IS}}(2)$ can be plausibly found by studying gap amplification of $\mathsf{QMA}_{\mathsf{IS}}$, since the latter implies no such complexity collapse. This raises the exciting possibility of using Corollary 116 to fully characterize $\mathsf{QMA}(2)$.

external structure / *internal structure*	GENERAL	SEPARABLE BIPARTITION
GENERAL	QMA	$\mathsf{QMA}(2)$
SEPARABLE SUBSYSTEM	$\mathsf{QMA}_{\mathsf{IS}} \subseteq \mathsf{EXP}$	$\mathsf{QMA}_{\mathsf{IS}}(2) = \mathsf{NEXP}$

Table 5.1: Categorizing quantum verification protocols by the guaranteed entanglement structures in the proof.

To show Theorem 112, we rely on a well-known duality of membership testing of a convex set and optimizing a convex function over this set. We use this duality to reduce any $\mathsf{QMA}_{\mathsf{IS}}$ problem to membership testing of the set of internally separable quantum states $\mathcal{W}_{\mathsf{IS}}$. In order to use tools from convex optimization, we represent quantum states as vectors in a generalized Bloch sphere. We use a non-ellipsoidal reduction that permits an error analysis matching our setup Liu [2007], Gharibian [2009].

We then show how to test membership of $\mathcal{W}_{\mathsf{IS}}$ in EXP. It is essential that only $O(1)$ qubits are unentangled from the rest of the subsystem. Our algorithm tests membership of two convex sets whose intersection forms $\mathcal{W}_{\mathsf{IS}}$; note that this requires testers with inverse exponential precision. The first set represents all Bloch vectors of quantum states, which

can be tested using exponentially many explicit conditions (e.g. Kimura [2003]). The second set represents all vectors whose reduced density matrix on the subsystem is separable. This can be tested by solving the Doherty-Parrilo-Spedalieri SDP Doherty et al. [2002, 2004], Navascués et al. [2009] which decides if a state has a *symmetric extension*. A careful error analysis shows that separability of a bipartite state ρ with local dimensions M, N can be tested to error δ in time $\widetilde{O}((2/\delta)^{9N}\text{poly}(M,N))$ Ioannou [2007]. When N is a fixed constant, this SDP runs in exponential time.

To prove Theorem 113, we consider $\mathsf{QMA}_{|\mathsf{IS}\rangle}$, a variant of $\mathsf{QMA}_{\mathsf{IS}}$ where the proof must also be a *pure state*. Note that unlike QMA and $\mathsf{QMA}(2)$, restricting $\mathsf{QMA}_{\mathsf{IS}}$ to pure states can change the power of the complexity class, as an internally separable ensemble of pure states may not decompose into an *ensemble* of internally separable pure states.[2] We then develop a $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ protocol for an NEXP-complete problem, building on the proof of $\mathsf{QMA}^+ = \mathsf{NEXP}$ Blier and Tapp [2010], Jeronimo and Wu [2023a], Bassirian et al. [2024b].

As in Jeronimo and Wu [2023a], Bassirian et al. [2024b], we consider a succinct constraint satisfaction problem (CSP) with constant promise gap, which is NEXP-hard by the PCP theorem Arora and Safra [1992], Arora et al. [1998a], Harsha [2004]. In the QMA^+ protocol, one enforces a *rigidity* property of the quantum proof: the state must be close to $\frac{1}{\sqrt{R}}\sum_{j\in[R]}|j\rangle\,|a_j\rangle$. It is easy to ensure that all constraints $j \in [R]$ are included in the state ("density"), but the positive-amplitude guarantee is used to ensure there is at most one assignment a_j per constraint ("quasirigidity"). We find a way to instead enforce quasirigidity using both internal separability *and* purity.

Lemma 117 (informal)**.** *There is a test A and a fixed quantum circuit C such that for any internally separable $|\psi\rangle$, if $A(|\psi\rangle) \geq 1-\epsilon$, then $C\,|\psi\rangle$ has $1 - O(\epsilon)$ squared overlap with a* quasirigid *state.*

2. For example, the same is true in QMA^+: this is the difference between doubly non-negative matrices and completely positive matrices.

In our $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ protocol, the verifier enforces tripartite proofs of the form

$$\sum_{j,a} z_{ja} \left|j\right\rangle \left|a, j\right\rangle \left|a\right\rangle .$$

This can be done by measuring all registers and accepting iff we see two copies of the same constraint j, and two copies of the same assignment a. We then observe that having multiple assignments per constraint adds entanglement to the second and third registers, even after tracing out the first register. Thus, if the state is *internally separable*, it has high overlap with a state with one assignment a per constraint j. Notably, this test is inspired by recasting the QMA^+ protocol in the language of "superposition detection" and non-collapsing measurement Bassirian and Marwaha [2024], Aaronson et al. [2024b], an unphysical operation related to a problem of Aaronson Aaronson [2023].

We then prove that $\mathsf{QMA}_{|\mathsf{IS}\rangle} \subseteq \mathsf{QMA}_{\mathsf{IS}}(2)$. Without loss of generality, the best proof Merlin can send in a $\mathsf{QMA}_{\mathsf{IS}}(2)$ protocol is a product state $\rho_1 \otimes \rho_2$, since separable states form the convex hull of the set of product states. The SWAP test accepts a product state with probability $\frac{1}{2}(1 + \mathrm{Tr}[\rho_1\rho_2])$: this is 1 only if ρ_1, ρ_2 are *pure*. The $\mathsf{QMA}_{\mathsf{IS}}(2)$ simulator thus applies a SWAP test with some probability, and otherwise applies the $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ circuit.

5.2 Related work

The complexity class $\mathsf{QMA}(2)$ One sign that $\mathsf{QMA}(2)$ is more powerful than QMA is that QMA with a $O(\log n)$-qubit proof is equal to BQP Marriott and Watrous [2005b]. By contrast, $\mathsf{QMA}(2)$ with two unentangled $O(\log n)$-qubit proofs can decide NP, although with inverse polynomially small completeness-soundness gap Aaronson et al. [2009], Beigi [2008], Blier and Tapp [2010], Chen and Drucker [2010], Chiesa and Forbes [2011], Gall et al. [2012], Pereszlényi [2012a]. Improving the completeness-soundness gap to a constant would likely imply $\mathsf{QMA}(2) = \mathsf{NEXP}$; see Jeronimo and Wu [2023a], Bassirian et al. [2024b], Jeronimo and Wu [2024] for a recent attempt using quantum proofs with non-negative amplitudes (QMA^+).

In general, one only needs two unentangled proofs since $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ Harrow and Montanaro [2013]; the proof of this result uses an iterated SWAP test. There are few complete problems known for $\mathsf{QMA}(2)$ Chailloux and Sattath [2012], Hayden et al. [2013], Gutoski et al. [2015]; one proposal is the *pure state marginal problem*, which is known to be QMA-hard and in $\mathsf{QMA}(2)$ Liu [2006], Liu et al. [2007], Liu [2007], Wei et al. [2010], Yu et al. [2021, 2022].

Convex optimization and quantum entanglement Some quantum papers build a Turing reduction using the duality between convex optimization and membership testing of a convex set, for example NP-hardness of separability testing Gurvits [2003], Ioannou [2007], Gharibian [2009], and QMA-hardness of the mixed state marginal problem Liu [2006], Liu et al. [2007], Liu [2007], Wei et al. [2010]. This reduction can be implemented in several ways, including the ellipsoid method Yudin and Nemirovskii [1976], Grötschel et al. [1988], random walks Bertsimas and Vempala [2004], and a perceptron-like algorithm Liu [2007]. One can use a hierarchy of semidefinite programs to detect entanglement Doherty et al. [2002, 2004], Navascués et al. [2009], Harrow et al. [2017] or optimize over separable states Brandão et al. [2011b], Brandão and Harrow [2013], Barak et al. [2017]. In some cases, carefully chosen ϵ-nets provide similar performance Shi and Wu [2015], Barak et al. [2014], Brandão and Harrow [2015].

Measuring quantum entanglement There are numerous ways to quantify entanglement; see the surveys Życzkowski and Bengtsson [2017], Gühne and Tóth [2009], Horodecki et al. [2009, 2024]. One approach considers the set of reachable states under transformations comprising only of local operations and classical communication (LOCC) Dür et al. [2000], Bennett et al. [2000], Vrana and Christandl [2015]. Other works classify multipartite entanglement by whether a *subsystem* is separable Thapliyal [1999], Brassard and Mor [2001]. Permutation-symmetric approaches include the Positive Partial Transpose (PPT) criterion Peres [1996] and quantum de Finetti theorems Caves et al. [2002], Christandl et al.

[2007]. Mysteriously, both approaches identify entangled states that "look" separable under LOCC transformations Tóth and Gühne [2009], Brandão et al. [2011b], Brandão and Harrow [2013], Lancien and Winter [2013], Li and Smith [2015]. There exist ensembles of low-entanglement quantum states that are computationally indistinguishable from maximally entangled quantum states Aaronson et al. [2024a].

5.3 Other implications

A complexity-theoretic framework to study multipartite entanglement. Despite much attention given to $\mathsf{QMA}(2)$, entanglement can be much richer than simply across a bipartition. For example, the W state

$$\frac{1}{\sqrt{3}}\left(|001\rangle + |010\rangle + |100\rangle\right)$$

and GHZ state $\frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right)$ are entangled in *inequivalent* ways: both states are entangled across every bipartition, yet the states are not equivalent under LOCC Dür et al. [2000]. The same paper observes that tracing out a qubit from the W state gives an entangled state, but the same action on the GHZ state destroys all entanglement.

$\mathsf{QMA}_{\mathsf{IS}}$ allows us to probe one type of multipartite (un)entanglement from the lens of computational complexity, defined as *partial semi-separability* in Brassard and Mor [2001]. It is intuitive to study the power of QMA with other kinds of multipartite entanglement guarantees. As a first step, we show that restricting to quantum proofs that are pure and *multiseparable*[3] Thapliyal [1999] can also decide NEXP:

Corollary 118. *There exist constants* $1 > c > s > 0$ *where* $\mathsf{QMA}^{c,s}$ *restricted to pure, multiseparable quantum proofs can decide* NEXP.

These results motivate a general question about the power of multipartite entanglement:

3. See Section 5.12 for a definition of multiseparability.

Question 119. *Which notions of* multipartite unentanglement *are computationally strong? Why?*

A unifying perspective: $\mathsf{QMA}(2)$ and purity testing. Many works Liu et al. [2007], Harrow and Montanaro [2013], Barak et al. [2017], Yu et al. [2022] use the fact that a protocol over separable states is equivalently a protocol over states with a purity constraint. This suggests that *purity testing* is at the heart of $\mathsf{QMA}(2)$'s power, even if it seems like a weaker primitive. In our case, $\mathsf{QMA}_{\mathsf{IS}}(2)$ is no more powerful than $\mathsf{QMA}_{\mathsf{IS}}$ with a purity test. The proof of Theorem 113 only uses the two unentangled quantum proofs to implement a SWAP test, ensuring that Merlin sends a *pure* quantum state. We show that optimizing over *purifications* of separable states is enough to decide NEXP, and even hard to approximate:

Corollary 120. *Given any $M \in \mathbb{R}^{n\times n}$, consider the optimization problem*

$$\max_{\rho\in\mathcal{W}_{|\mathsf{IS}\rangle}} \operatorname{Tr}[M\rho]\,, \tag{5.1}$$

where $\mathcal{W}_{|\mathsf{IS}\rangle}$ is the set of n-dimensional pure *states that the prover in $\mathsf{QMA}_{\mathsf{IS}}$ can send. Then there exists a constant $0 < \alpha < 1$ for which it is NP-hard to compute an α-approximation of this problem.*

By contrast, optimizing over separable states (of $\operatorname{poly}(n)$ dimension) is known to be NP-hard, but with gap decreasing with system size Gurvits [2003], Aaronson et al. [2009], Gharibian [2009], Blier and Tapp [2010], Pereszlényi [2012a], Gall et al. [2012]. If the gap in these problems could be made constant, then $\mathsf{QMA}(2) = \mathsf{NEXP}$.[4]

More generally, with mild conditions on a set of quantum states $\mathcal{W}$, we show how a $\mathsf{QMA}(2)$ protocol with each proof restricted to $\mathcal{W}$ can implement a QMA protocol with a proof restricted to *pure states* in $\mathcal{W}$. We use this as a recipe to generate other computational models where one quantum proof is provably weaker than two *unentangled* quantum proofs,

4. One must also assume that the problems can be made succinct.

up to standard complexity-theoretic assumptions. We give two examples in Section 5.11, from the pure quantum proofs in QMA^+ and the *proper states* of Aaronson et al. [2009], Beigi [2008]. Each set represents the pure states of some larger set $\mathcal{W}$. Then, $\mathsf{QMA}_{\mathcal{W}}$ can be decided with a semidefinite program, but $\mathsf{QMA}_{\mathcal{W}}(2)$ can restrict to the more capable pure states. This technique can likely generate other examples that demonstrate the power of unentanglement.

5.4 Our setup

Let $\mathcal{D}(X)$ be the set of density matrices over register X. Let $\mathsf{SEP}(X,Y)$ be the set of separable density matrices over registers X, Y. Let $Sym(X_1, \ldots, X_k)$ be the symmetric subspace over registers $X_1, \ldots, X_k$. Let $XY|Z$ refer to the bipartition with registers X, Y in one part and Z in the other.

QMA with a restricted set of proofs. Fix any set of quantum states $\mathcal{W}$. Let $\mathsf{QMA}_{\mathcal{W}}(m)$ as $\mathsf{QMA}(m)$ where Merlin *must* send m unentangled proofs from $\mathcal{W}$. Here we modify both completeness and soundness. Only modifying completeness sometimes has no effect, as seen in Grilo et al. [2014].

Definition 121 ($\mathsf{QMA}_{\mathcal{W}}(m)$)**.** *Let $c, s : \mathbb{N} \to \mathbb{R}^+$ be polynomial-time computable functions, and m, p be polynomials. Fix an arbitrary set of quantum states $\mathcal{W}$.*

A promise problem $A = (A_{yes}, A_{no})$ is in $\mathsf{QMA}^{c,s}_{\mathcal{W}}(m)$ if and only if there exists a polynomial q and a polynomial-time uniform family of quantum circuits $\{Q_n\}$, where Q_n takes as input a string $x \in \{0,1\}^n$, m $p(n)$-qubit quantum states in $\mathcal{W}$, and $q(n)$ scratch qubits in state $|0\rangle^{\otimes q(n)}$, such that:

- *completeness: If $x \in A_{yes}$, there is a set of m $p(n)$-qubit quantum states $\rho_1, \ldots, \rho_m \in \mathcal{W}$ such that Q_n accepts $|x\rangle\langle x| \otimes (\rho_1 \otimes \cdots \otimes \rho_m) \otimes |0\rangle\langle 0|^{\otimes q(n)}$ with probability at least $c(n)$.*

- *soundness: If $x \in A_{no}$, then for any set of m $p(n)$-qubit quantum states $\rho_1, \ldots, \rho_m \in \mathcal{W}$, Q_n accepts $|x\rangle\langle x| \otimes (\rho_1 \otimes \cdots \otimes \rho_m) \otimes |0\rangle\langle 0|^{\otimes q(n)}$ with probability at most $s(n)$.*

Remark 122. *Following the convention in Jeronimo and Wu [2023a], we drop the c, s when the completeness-soundness gap is allowed to be any absolute constant:*

$$\mathsf{QMA}_{\mathcal{W}}(m) \stackrel{\text{def}}{=} \bigcup_{c-s=\Omega(1)} \mathsf{QMA}^{c,s}_{\mathcal{W}}(m).$$

However, note that $\mathsf{QMA}^{c,s}_{\mathcal{W}}(m)$ is not in general equal to $\mathsf{QMA}^{c',s'}_{\mathcal{W}}(m)$ Bassirian et al. [2024b].

This work studies tripartite proofs. Label the three registers A, B, C. Consider two families of quantum states, where $k \in \mathbb{N}$ is the dimension of register C:

- ${\mathcal{W}_{\mathsf{IS}}}^k = \{\rho \mid \mathrm{Tr}_A[\rho] \in \mathsf{SEP}(B, C)\}$: all mixed states that are *internally separable* on B, C.
- ${\mathcal{W}_{|\mathsf{IS}\rangle}}^k = \{|\psi\rangle \mid \mathrm{Tr}_A[|\psi\rangle\langle\psi|] \in \mathsf{SEP}(B, C)\}$: all *purifications* of separable states on B, C.

We denote $\mathcal{W}_{\mathsf{IS}} = \bigcup_{k\in\mathbb{N}} {\mathcal{W}_{\mathsf{IS}}}^k$ and $\mathcal{W}_{|\mathsf{IS}\rangle} = \bigcup_{k\in\mathbb{N}} {\mathcal{W}_{|\mathsf{IS}\rangle}}^k$ to allow the dimension of register C to be any fixed number. This generates the following complexity classes:[5]

$$\mathsf{QMA}_{\mathsf{IS}}(m) \stackrel{\text{def}}{=} \mathsf{QMA}_{\mathcal{W}_{\mathsf{IS}}}(m), \tag{5.2}$$

$$\mathsf{QMA}_{|\mathsf{IS}\rangle} \stackrel{\text{def}}{=} \mathsf{QMA}_{\mathcal{W}_{|\mathsf{IS}\rangle}}. \tag{5.3}$$

We will only consider $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ with one proof, as we prove $\mathsf{QMA}_{|\mathsf{IS}\rangle} = \mathsf{NEXP}$ in Section 5.6. This implies $\mathsf{QMA}_{|\mathsf{IS}\rangle}(m) = \mathsf{NEXP}$ for all $m \geq 1$.

5. In the definition for $\mathsf{QMA}_{\mathsf{IS}}$ and $\mathsf{QMA}_{|\mathsf{IS}\rangle}$, the underlying Hilbert spaces of registers A, B, C are omitted. To be more accurate, we could instead define $\mathsf{QMA}^{a,b,k}_{\mathsf{IS}}, \mathsf{QMA}^{a,b,k}_{|\mathsf{IS}\rangle}$, where a, b correspond to the number of qubits in registers A and B (some polynomial depending on n), and $k \in \mathbb{N}$ is some constant representing the number of qubits for register C. However, to ease the notation, we put this information in $\mathcal{W}_{\mathsf{IS}}$ and $\mathcal{W}_{|\mathsf{IS}\rangle}$.

We occasionally select another property P to choose a set of quantum states which we denote $\mathcal{W}_P$. When this occurs, we use the shorthand $\mathsf{QMA}_P(m) \stackrel{\text{def}}{=} \mathsf{QMA}_{\mathcal{W}_P}(m)$. We use the label $|P\rangle$ to signify that the set $\mathcal{W}_{|P\rangle}$ is pure.

Convex optimization. In the area of convex optimization, a convex set K is typically defined as a subset of $\mathbb{R}^d$ that is equipped with the standard (Euclidean) inner product.[6] Let $B_\ell(\vec{p})$ be the ball of radius ℓ around point $\vec{p} \in \mathbb{R}^d$. Some algorithms use special properties of K:

- *compactness*: By Heine-Borel, it is equivalent for K to be closed and bounded.
- *well-bounded $\vec{p}$-centeredness*: There is a known radius R such that $K \subseteq B_R(\vec{0})$. Moreover, there is a known point $\vec{p}$ and radius r such that $B_r(\vec{p}) \subseteq K$.

To work with this formalism, we represent a quantum state in $\mathcal{D}(\mathbb{C}^M)$ by its generalized Bloch vector in $\mathbb{R}^{M^2-1}$; see Kimura [2003] for a review. For example, fix any set $\{\sigma_i\}_{i\in[M^2-1]}$ of traceless Hermitian generators of $SU(M)$ satisfying $\mathrm{Tr}\big[\sigma_i\sigma_j\big] = 2\delta_{ij}$. Then we may decompose ρ as

$$\rho = \frac{\mathbb{I}}{M} + \frac{1}{2}\sum_{i=1}^{M^2-1} \vec{r}_i\sigma_i \,. \tag{5.4}$$

We call $\vec{r} \in \mathbb{R}^{M^2-1}$ the (generalized) *Bloch vector* of ρ. Note that a Bloch vector uniquely specifies a quantum state. We can thus identify a set of quantum states $J \subseteq \mathcal{D}(\mathbb{C}^M)$ with the set of associated Bloch vectors $K \subseteq \mathbb{R}^{M^2-1}$. We only consider problems where M is a power of 2. Thus, choose the generators to be the nontrivial $\log M$-qubit Pauli words, each multiplied by the normalizing factor $2^{(1-\log M)/2}$ to preserve the identity $\mathrm{Tr}\big[\sigma_i\sigma_j\big] = 2\delta_{ij}$.

Let $S(K,\epsilon)$ be the set of points x that are at most ϵ-far from K; i.e. $B_\epsilon(x) \cap K$ is nonempty. Let $S(K,-\epsilon)$ be the set of points $x \in K$ that are at least distance ϵ from the boundary; i.e. $B_\epsilon(x) \subseteq K$. We always have $S(K,-\epsilon) \subseteq K \subseteq S(K,\epsilon)$.

6. For a self-contained introduction to convex optimization, see the textbook Grötschel et al. [1988].

We work with two of the five "basic problems" Grötschel et al. [1988] in convex optimization. The first problem we use decides the maximum value of a linear function on a convex set.

Definition 123 (Weak Validity (WVAL_ϵ(K)) Grötschel et al. [1988], Liu [2007], Gharibian [2009])**.** *Let $K \subseteq \mathbb{R}^d$ be a convex, compact, and well-bounded $\vec{p}$-centered set. Then, given $\vec{c} \in \mathbb{Q}^d$ with $\|\vec{c}\|_2 = 1$, and $\gamma, \epsilon \in \mathbb{Q}$ such that $\epsilon > 0$, decide the following:*

- *If there exists $\vec{y} \in S(K, -\epsilon)$ with $\vec{c} \cdot \vec{y} \geq \gamma + \epsilon$, then output YES.*
- *If for all $\vec{x} \in S(K, \epsilon)$, we have $\vec{c} \cdot \vec{x} \leq \gamma - \epsilon$, then output NO.*[7]

The second problem we use decides if an arbitrary point is in the convex set:

Definition 124 (Weak Membership (WMEM_β(K)) Grötschel et al. [1988], Liu [2007], Gharibian [2009])**.** *Let $K \subseteq \mathbb{R}^d$ be a convex, compact, and well-bounded $\vec{p}$-centered set. Then given $y \in \mathbb{Q}^d$ and parameter $\beta \in \mathbb{Q}$ such that $\beta > 0$, decide the following:*

- *If $y \in S(K, -\beta)$, then output YES.*
- *If $y \notin S(K, \beta)$, then output NO.*

We conclude this section with a few facts about Bloch vector representations of quantum states. To start, the norm of the Bloch vector is always bounded by a constant:

Claim 125 (Folklore)**.** *Every Bloch vector has norm at most $R = \sqrt{2}$.*

Proof. Fix any quantum state ρ, and let $\vec{r}$ be the associated Bloch vector. Then

$$1 \geq \mathrm{Tr}\left[\rho^2\right] = \frac{1}{M} + \frac{1}{2} \sum_{i=1}^{M^2-1} \vec{r}_i{}^2 \geq \frac{1}{2}\|\vec{r}\|^2 . \qquad \square$$

Moreover, Bloch vectors with small enough norm are separable across every bipartition:

7. We use the naming convention of Grötschel et al. [1988] where WVAL is a decision problem. As in Liu [2007], Gharibian [2009], we formulate WVAL and WMEM as promise problems instead of multi-output problems.

Fact 126 (Gurvits and Barnum [2002])**.** *Fix any quantum state $\rho \in \mathcal{D}(\mathbb{C}^M)$, and let $\vec{r}$ be the associated Bloch vector. If $\|\vec{r}\|^2 \leq \frac{1}{M^2}$, then ρ is separable across every bipartition.*

Corollary 127. *The set of Bloch vectors representing separable states in $\mathcal{D}(\mathbb{C}^M, \mathbb{C}^N)$ is well-bounded $\vec{0}$-centered with $r = \frac{1}{M^2N^2}$ and $R = \sqrt{2}$.*

Finally, the Euclidean distance of two vectors is upper-bounded by a constant times the trace distance of their associated matrices:

Claim 128. *Consider two vectors $\vec{r}, \vec{t} \in \mathbb{R}^{M^2-1}$, and let ρ and τ be the associated $M \times M$ matrices according to Eq.* (5.4)*. Then $\|\vec{r} - \vec{t}\|_2 \leq \sqrt{2}\|\rho - \tau\|_1$.*

Proof. Note that $\|\rho - \tau\|_1 \geq \|\rho - \tau\|_2$ by monotonicity of Schatten norms. Moreover,

$$\|\rho - \tau\|_2 = \sqrt{\frac{1}{4} \sum_{i,j=1}^{M^2-1} (\vec{r}_i - \vec{t}_i)(\vec{r}_j - \vec{t}_j) \operatorname{Tr}\left[\sigma_i \sigma_j\right]} = \sqrt{\frac{1}{2} \sum_{i=1}^{M^2-1} (\vec{r}_i - \vec{t}_i)^2} = \frac{1}{\sqrt{2}} \|\vec{r} - \vec{t}\|_2 \,.$$

□

5.5 Optimization over internally separable states: $\mathsf{QMA}_{\mathsf{IS}} \subseteq \mathsf{EXP}$

In this section, we show how to decide any $\mathsf{QMA}_{\mathsf{IS}}$ problem in exponential time. This has three steps:

$$\mathsf{QMA}_{\mathsf{IS}} \overset{(1)}{\leq_m} \mathrm{WVAL}(\mathcal{W}_{\mathsf{IS}}) \overset{(2)}{\leq_T} \mathrm{WMEM}(\mathcal{W}_{\mathsf{IS}}) \overset{(3)}{\subseteq} \mathsf{EXP} \,. \tag{5.5}$$

First, we build a many-one reduction[8] from any problem in $\mathsf{QMA}_{\mathsf{IS}}$ to the convex optimization problem $\mathrm{WVAL}(\mathcal{W}_{\mathsf{IS}})$. Then, we use a Turing reduction from $\mathrm{WVAL}(\mathcal{W}_{\mathsf{IS}})$ to the membership testing problem $\mathrm{WMEM}(\mathcal{W}_{\mathsf{IS}})$. We finally show that $\mathrm{WMEM}(\mathcal{W}_{\mathsf{IS}})$ can be decided in

8. Recall that a *Turing* reduction from problem A to problem B is an algorithm to solve problem A using an oracle for problem B. A *many-one* reduction is a special type of Turing reduction where the oracle can only be called once, and the oracle's output must be returned by the algorithm.

EXP.[9]

For any problem in $\mathsf{QMA}_{\mathsf{IS}}$, there is a $\mathsf{QMA}_{\mathsf{IS}}$ protocol and polynomial p such that for input size n, the quantum proof in $\mathcal{W}_{\mathsf{IS}}$ has $p(n)$ qubits. In this section, we identify $\mathcal{W}_{\mathsf{IS}}$ with its subset on $p(n)$ qubits, i.e. the set of allowed quantum proofs in this protocol. By Section 5.4, we can interchange $\mathcal{W}_{\mathsf{IS}}$ with its projection to the generalized Bloch sphere (a subset of $\mathbb{R}^{4^{p(n)}-1}$) in exponential time. Therefore, in this section, we may view $\mathcal{W}_{\mathsf{IS}}$ as a subset of real Euclidean space.

5.5.1 Showing that $\mathcal{W}_{\mathsf{IS}}$ is well-behaved

We first verify that $\mathcal{W}_{\mathsf{IS}}$ has all the properties we need in our reduction.

Claim 129. *$\mathcal{W}_{\mathsf{IS}}$ is a convex, well-bounded $\vec{0}$-centered set with radii $R = \sqrt{2}, r = \frac{1}{4^{p(n)}}$.*

Proof. We first check convexity. Consider $\rho_1, \rho_2 \in \mathcal{W}_{\mathsf{IS}}$. Since both $\mathrm{Tr}_A[\rho_1]$ and $\mathrm{Tr}_A[\rho_2]$ are separable, then so is $\mathrm{Tr}_A[p\rho_1 + (1-p)\rho_2]$. So the state $p\rho_1 + (1-p)\rho_2 \in \mathcal{W}_{\mathsf{IS}}$. This establishes that $\mathcal{W}_{\mathsf{IS}}$ is convex when viewed as density matrices. Since the transformation from density matrix to Bloch vector is linear, $\mathcal{W}_{\mathsf{IS}}$ is also convex when viewed as Bloch vectors.

By Claim 125, all Bloch vectors have norm at most $\sqrt{2}$. Moreover, by Fact 126, every Bloch vector in $\mathbb{R}^{M^2-1}$ with norm at most $\frac{1}{M^2}$ is separable across every bipartition. This implies that they are separable across the $AB|C$ bipartition, and are thus internally separable. So $\mathcal{W}_{\mathsf{IS}}$ is well-bounded $\vec{0}$-centered with $r = \frac{1}{4^{p(n)}}$ and $R = \sqrt{2}$. □

It will be useful to represent $\mathcal{W}_{\mathsf{IS}}$ as the intersection of two sets $\mathcal{W}_{\mathsf{IS}} = K_1 \cap K_2$. Let K_1 be the set of all Bloch vectors corresponding to density matrices on $p(n)$ qubits; i.e. the projection of $\mathcal{D}(\mathbb{C}^{2^{p(n)}})$ onto $\mathbb{R}^{4^{p(n)}-1}$. Let K_2 be the set of all vectors in $\mathbb{R}^{4^{p(n)}-1}$ whose associated vector on the subsystem represents a separable state (up to normalization), and where all other coordinates have magnitude at most 2. This works for the following reason:

9. To be precise, step (1) runs in exponential time and creates an exponentially long output. Steps (2) and (3) each run in polynomial time with respect to input size, which is exponential in the original $\mathsf{QMA}_{\mathsf{IS}}$ problem's input length.

Remark 130. *When the Bloch vector encoding uses the generators $\{\sigma_i\}$ associated with Pauli words, the Bloch vector of a* subsystem *is a subset of the coordinates of the original Bloch vector, multiplied to a global normalization.*[10] *In particular, $\vec{r}_i = \mathrm{Tr}[\rho\sigma_i]$ corresponds to an element of the Bloch vector of the subsystem if and only if σ_i is $\mathbb{I}$ on every qubit beyond the subsystem.*

Claim 131. *K_1 and K_2 are convex, compact, well-bounded $\vec{0}$-centered sets, with associated radii $R_1 = \sqrt{2}, R_2 = 4^{p(n)}, r = \frac{1}{4^{p(n)}}$.*

Proof. We start with convexity. The set of all $p(n)$-qubit states is convex; since the transformation from density matrix to Bloch vector is linear, K_1 is also convex. The set of separable states is convex, by the same argument, so are the associated Bloch vectors. K_2 is the set of all vectors which match a separable Bloch vector on a fixed set of coordinates, intersected with the solid hypercube $[-2, 2]^{4^{p(n)}-1}$. Both sets are convex (and K_2 is non-empty), so K_2 is convex.

All Bloch vectors have norm at most $\sqrt{2}$. Furthermore, vectors in K_2 have norm at most $\sqrt{2 \cdot (4^{p(n)} - 1)}$. Since we have $\mathcal{W}_{\mathsf{IS}} \subseteq K_1$ and $\mathcal{W}_{\mathsf{IS}} \subseteq K_2$, both K_1 and K_2 contain a $\vec{0}$-centered ball of radius $r = \frac{1}{4^{p(n)}}$ by Claim 129.

The set of all density matrices is compact (e.g. Życzkowski and Bengtsson [2017]). Since the transformation from density matrix to Bloch vectors is linear, K_1 is also compact. By Heine-Borel, K_2 is compact iff it is closed and bounded. It remains to show closure. The set of Bloch vectors representing separable states is closed (e.g. Gharibian [2009]). Note that K_2 is the Cartesian product of this set with the closed set $[-2, 2]$ on each other coordinate. Since the product of closed sets is closed, K_2 is closed. □

Claim 132. *$\mathcal{W}_{\mathsf{IS}}$ is compact.*

Proof. By Heine-Borel, $\mathcal{W}_{\mathsf{IS}}$ is compact since it is closed and bounded. $\mathcal{W}_{\mathsf{IS}}$ is bounded

10. The normalizing factor is exactly sqrt of the dimension of register A, to preserve the identity $\mathrm{Tr}[\sigma_i\sigma_j] = 2\delta_{ij}$.

because the norm is always at most R. $\mathcal{W}_{\mathsf{IS}}$ is closed because it is the intersection of two closed sets K_1, K_2 (Claim 131). □

5.5.2 Proving the reduction

The first step is straightforward. Recall the definition of WVAL from Section 5.4. Observe that the maximum acceptance probability of any $\mathsf{QMA}_{\mathsf{IS}}$ protocol with verifier circuit V can be written as $\max_{\rho \in \mathcal{W}_{\mathsf{IS}}} \mathrm{Tr}[V\rho]$. Thus, we can decide this problem if we can decide the value $\max_{\rho \in \mathcal{W}_{\mathsf{IS}}} \mathrm{Tr}[V\rho]$ with good enough precision.

Claim 133. *Every problem in* $\mathsf{QMA}_{\mathsf{IS}}$ *with completeness-soundness gap* Δ *has an exponential time reduction to* $\mathrm{WVAL}_\epsilon(\mathcal{W}_{\mathsf{IS}})$*, where* $\epsilon = \Omega(\frac{\Delta}{2^{\mathrm{poly}(n)}})$.

Proof. Suppose the verification circuit is V, and let the proof be a state on $p(n)$ qubits. We represent V as a vector $\vec{c}$, so the acceptance probability of V on ρ is affine to the inner product of $\vec{c}$ with the Bloch vector representation of ρ. The quantity $\mathrm{Tr}[V\rho]$ can be decomposed via Eq. (5.4) as

$$\mathrm{Tr}[V\rho] = \frac{1}{2^{p(n)}} \mathrm{Tr}[V] + \frac{1}{2} \sum_{i=1}^{4^{p(n)}-1} \vec{r}_i \cdot \mathrm{Tr}[V\sigma_i] \,, \tag{5.6}$$

where $\vec{r} \in \mathbb{R}^{4^{p(n)}-1}$ is the Bloch vector representation of ρ. Let $\vec{d}$ be the vector where $\vec{d_i} = \mathrm{Tr}[V\sigma_i]$, and choose $\vec{c} \stackrel{\mathrm{def}}{=} \frac{\vec{d}}{\|\vec{d}\|}$. We now argue that there is a small enough ϵ so that deciding $\mathrm{WVAL}_\epsilon(\mathcal{W}_{\mathsf{IS}})$ with $\vec{c}$ decides the $\mathsf{QMA}_{\mathsf{IS}}$ problem.

Since V is generated by a polynomial-time Turing machine, there is a polynomial q such that each matrix element of V has norm at most $2^{q(n)}$. So, each vector element $\vec{d_i} = \mathrm{Tr}[V\sigma_i]$ has norm at most $2^{p(n)+q(n)}$. This implies that $\|\vec{d}\|$ is at most $2^{2p(n)+q(n)}$.

Suppose the $\mathsf{QMA}_{\mathsf{IS}}$ problem has soundness s. Then in soundness, for any state $\rho \in \mathcal{W}_{\mathsf{IS}}$,

we have $\mathrm{Tr}[V\rho] \leq s$, or equivalently,

$$\vec{c} \cdot \vec{r} = \frac{2}{\|\vec{d}\|} \left(\mathrm{Tr}[V\rho] - \frac{1}{2^{p(n)}} \mathrm{Tr}[V] \right) \leq \frac{2}{\|\vec{d}\|} \left(s - \frac{1}{2^{p(n)}} \mathrm{Tr}[V] \right) . \tag{5.7}$$

The $\mathsf{QMA}_{\mathsf{IS}}$ problem has completeness $s + \Delta$ by definition. In this case, there exists $\rho \in \mathcal{W}_{\mathsf{IS}}$ where $\mathrm{Tr}[V\rho] \geq s + \Delta$, or equivalently,

$$\vec{c} \cdot \vec{r} \geq \frac{2}{\|\vec{d}\|} \left(s - \frac{1}{2^{p(n)}} \mathrm{Tr}[V] \right) + \frac{2}{\|\vec{d}\|} \Delta \, . \tag{5.8}$$

Define the parameter

$$\gamma \stackrel{\text{def}}{=} \frac{2}{\|\vec{d}\|} \left(s - \frac{1}{2^{p(n)}} \mathrm{Tr}[V] \right) + \frac{1}{\|\vec{d}\|} \Delta .$$

We verify this problem can be solved by the *weak* optimization problem $\mathrm{WVAL}_{\epsilon/4^{p(n)+1}}(\mathcal{W}_{\mathsf{IS}})$ for

$$\epsilon \stackrel{\text{def}}{=} \min \left(\frac{1}{2 \cdot 4^{p(n)}}, \frac{\Delta}{2\|\vec{d}\|} \right) . \tag{5.9}$$

- In the soundness case, consider $\vec{y} \in S(\mathcal{W}_{\mathsf{IS}}, \epsilon)$. By definition, there exists a vector $\vec{x} \in \mathcal{W}_{\mathsf{IS}}$ such that $\|\vec{y} - \vec{x}\| \leq \epsilon$. So by Cauchy-Schwarz, for unit vector $\vec{c}$, we have

$$\vec{c} \cdot \vec{y} = \vec{c} \cdot \vec{x} + \vec{c} \cdot (\vec{y} - \vec{x}) \leq \vec{c} \cdot \vec{x} + \|\vec{c}\| \|\vec{y} - \vec{x}\| \leq \vec{c} \cdot \vec{x} + \epsilon \, . \tag{5.10}$$

  Since for any $\vec{x} \in \mathcal{W}_{\mathsf{IS}}$, we have $\vec{c} \cdot \vec{x} \leq \gamma - \frac{\Delta}{\|\vec{d}\|} \leq \gamma - 2\epsilon$ by Eq. (5.7), the above inequality implies $\vec{c} \cdot \vec{y} \leq \gamma - \epsilon$.

- In completeness, there is an $\vec{r} \in \mathcal{W}_{\mathsf{IS}}$ such that $\vec{c} \cdot \vec{r} \geq \gamma + \frac{\Delta}{\|\vec{d}\|} \geq \gamma + 2\epsilon$ by Eq. (5.8). We split into two cases, depending on the norm of $\vec{r}$:

  - If $\|\vec{r}\| \leq \epsilon \leq \frac{1}{2 \cdot 4^{p(n)}}$, then it must be in $S(\mathcal{W}_{\mathsf{IS}}, -\frac{1}{2 \cdot 4^{p(n)}}) \subseteq S(\mathcal{W}_{\mathsf{IS}}, -\epsilon)$. This is

because by Claim 129, all vectors of norm at most $\frac{1}{4^{p(n)}}$ are in $\mathcal{W}_{\mathsf{IS}}$.

– Otherwise, define $\vec{z} \stackrel{\text{def}}{=} \vec{r} \cdot \frac{\|\vec{r}\|-\epsilon}{\|\vec{r}\|}$. Since $\vec{0} \in \mathcal{W}_{\mathsf{IS}}$, $\vec{z} \in \mathcal{W}_{\mathsf{IS}}$ by convexity. Moreover,

$$\vec{c} \cdot \vec{z} = \vec{c} \cdot \vec{r} + \vec{c} \cdot (\vec{z} - \vec{r}) \geq \vec{c} \cdot \vec{r} - \epsilon \geq \gamma + \epsilon \,. \tag{5.11}$$

Finally, by a geometric argument[11], $B_\ell(\vec{z}) \subseteq \mathcal{W}_{\mathsf{IS}}$ for $\ell = \frac{1}{2} \cdot \frac{\epsilon}{\|\vec{r}\|} \cdot \frac{1}{4^{p(n)}} \geq \frac{\epsilon}{4 \cdot 4^{p(n)}}$. So then $\vec{z} \in S(\mathcal{W}_{\mathsf{IS}}, -\frac{\epsilon}{4 \cdot 4^{p(n)}})$. □

The second step uses a duality between optimization over a convex set K and membership testing of K. In our case, we need a reduction from optimization to membership testing. We use a reduction that runs in exponential time given our setup. The runtime of this reduction depends on *the encoding size* of the convex set K, denoted by $\langle K \rangle$. This is defined as the dimension plus the number of bits required to specify $r, R, \vec{p}$ Gurvits [2003], Gharibian [2009]. In our case, $\langle \mathcal{W}_{\mathsf{IS}} \rangle$ is at most $O(2^{\text{poly}(n)})$.

Theorem 134 ([Liu, 2007, Proposition 2.8])**.** *Let $K \subseteq \mathbb{R}^d$ be a convex, compact, and well-bounded $\vec{p}$-centered set with associated radii (R, r). Given an instance $\Pi = (K, c, \gamma, \epsilon)$ of $\text{WVAL}_\epsilon(K)$ with $0 < \epsilon < 1$, there exists an algorithm which runs in time $\text{poly}(\langle K \rangle, R, \lceil 1/\epsilon \rceil)$, and solves Π using an oracle for $\text{WMEM}_\beta(K)$ with $1/\beta = O(\text{poly}(d, \epsilon, R, 1/r))$.*

Putting the first two steps together, we obtain the following:

Corollary 135. *Every problem in $\mathsf{QMA}_{\mathsf{IS}}$ with completeness-soundness gap Δ has an exponential time reduction to $\text{WMEM}_\beta(\mathcal{W}_{\mathsf{IS}})$, where $\beta = \Omega(\frac{\Delta}{2^{\text{poly}(n)}})$.*

In the third step, we show that $\text{WMEM}(\mathcal{W}_{\mathsf{IS}})$ can be solved in exponential time. We do this by testing membership of convex sets K_1, K_2 whose intersection $K_1 \cap K_2 = \mathcal{W}_{\mathsf{IS}}$.

Theorem 136 ([Grötschel et al., 1988, Section 4.7])**.** *For each $i \in \{1, 2\}$, let $K_i \subseteq \mathbb{R}^d$ be a convex, compact, and well-bounded $\vec{p_i}$-centered set with associated radii (R_i, r_i). Suppose we*

11. Consider a $\vec{0}$-centered convex set K with inner radius r. Then for any vector $\vec{x} \in K$ and $0 \leq \alpha \leq 1$, the set K contains a ball centered at $(1-\alpha)\vec{x}$ of radius at least $\frac{1}{2}\alpha r$.

are given an $r_3 \in \mathbb{Q}$ *such that* $K_1 \cap K_2$ *contains a ball with radius* r_3*. Then* $K_1 \cap K_2$ *is a well-bounded convex body, and given an instance* $\Pi = (K_1 \cap K_2, y, \beta)$ *of* $\mathrm{WMEM}_\beta(K)$*, there exists an algorithm which runs in constant time and solves* Π *using an oracle for* $\mathrm{WMEM}_\zeta(K_1)$ *and* $\mathrm{WMEM}_\zeta(K_2)$*, with* $\zeta = \beta \cdot \frac{r_3}{2\min(R_1,R_2)}$.

To decide $\mathrm{WMEM}(K_2)$, we must test that each coordinate of the vector has magnitude at most 2, and that the vector's projection to the subsystem (as in Remark 130) is proportional to a Bloch vector of a separable state. Fortunately, since one part of the subsystem has $O(1)$ qubits, we can test the latter using an algorithm that runs in time polynomial in dimension, i.e. exponential time.

There are many known algorithms to test whether a Bloch vector is separable; see Ioannou [2007] for a (slightly dated) comparison. In this work, we rely on an algorithm which searches over *symmetric extensions* of a state. This algorithm takes an input $\rho \in \mathcal{D}(X, Y)$ and tries to find an *extension*; i.e. a state $\sigma \in \mathcal{D}(X, Y_1, \ldots, Y_k)$ such that the reduced density matrix of σ on X and any Y register is equal to ρ. A state is separable if and only if it is k-extendable for all k, so this test always succeeds at large enough k (although this may not be polynomial in the number of qubits). This algorithm was initiated by works of Doherty-Parrilo-Spedalieri Doherty et al. [2002, 2004] and the quantum de Finetti theorem Caves et al. [2002], Christandl et al. [2007], but see follow-ups Navascués et al. [2009], Brandão and Christandl [2012], Harrow et al. [2017] which find ways to quantify and reduce the value of k required to test separability.

Claim 137 ([Ioannou, 2007, Corollary 1 and Section 3.5])**.** *Let* $\mathcal{S}_{N_1,N_2}$ *be the set of separable states of dimension* $N_1 \times N_2$*. Then there is an algorithm to decide* $\mathrm{WMEM}_\beta(\mathcal{S}_{N_1,N_2})$ *(with respect to the trace distance) in time* $\mathrm{poly}((1/\beta)^{9N_2}, N_1, N_2)$.

By Claim 128, we know that Claim 137 also holds with respect to Euclidean distance of the real vectors. As a consequence:

Theorem 138. *We can decide* $\mathrm{WMEM}_\zeta(K_2)$ *in exponential time, for any* ζ *that is at least an inverse exponential in* n*.*

Proof. Let $M = 2^{p(n)}$ be such that $K_2 \subseteq \mathbb{R}^{M^2-1}$. Consider any vector $\vec{r} \in \mathbb{R}^{M^2-1}$. We first trivially check if any coordinate is more than 2; if it is, we stop and report that $\vec{r} \notin K_2$. Otherwise, we consider the projection of $\vec{r}$ to the BC subsystem, which is a vector on fewer dimensions. By Claim 137, we can check this reduced vector is separable up to error $\beta \stackrel{\text{def}}{=} \frac{\zeta}{2M}$ in time $\mathrm{poly}((2M/\zeta)^{9\dim(C)}, \dim(B), \dim(C))$. Since $\dim(C)$ is at most some uniform constant, this runtime is $\mathrm{poly}(1/\zeta, M)$, i.e. exponential in n.

We remark on the error tolerance. Let $|^{BC}$ represent the projection onto the BC subsystem. By Remark 130, K_2 is the direct sum of $K_2|^{BC}$ and a solid hypercube (for the other coordinates).

- In the YES case, the input $\vec{x} \in S(K_2, -\zeta)$. So $K_2|^{BC}$ contains a ball centered at $\vec{x}|^{BC}$ of radius ζ: $\vec{x}|^{BC} \in S(K_2|^{BC}, -\zeta)$. This is correctly decided since Claim 137 has error at most $\frac{\zeta}{M}$, and the normalization factor is at most M.

- In the NO case, the input $\vec{x} \notin S(K_2, \zeta)$. We can check each coordinate of $\vec{x}$ is at most 2 up to bit precision, i.e. inverse double exponential in n. If this check succeeds, then $\vec{x}|^{BC}$ must be at least distance $\frac{\zeta}{2}$ from $K_2|^{BC}$: $\vec{x}|^{BC} \notin S(K_2|^{BC}, \frac{\zeta}{2})$. This is correctly decided since Claim 137 has error at most $\frac{\zeta}{2M}$, and the normalization factor is at most M. □

To decide $\mathrm{WMEM}(K_1)$, we must test whether a vector is a true Bloch vector. This amounts to ensuring that the associated density matrix is positive semi-definite, as it is already trace-1 and Hermitian. There are several ways to test this property; for example, we can use Claim 137 with $N_1 = M$ and $N_2 = 1$.

Fact 139. *We can decide* $\mathrm{WMEM}_\zeta(K_1)$ *in exponential time, for any* ζ *that is at least an inverse exponential in* n*.*

Remark 140. *In a model of computation that supports exact arithmetic, we can analytically verify that a vector is a true Bloch vector Kimura [2003]. Choose* $\vec{r} \in \mathbb{R}^{M^2-1}$ *and a set* $\{\sigma_i\}_{i\in[M^2-1]}$ *of traceless Hermitian generators of* $SU(M)$ *satisfying* $\mathrm{Tr}\left[\sigma_i\sigma_j\right] = 2\delta_{ij}$*. Let* P *be the* $M \times M$ *matrix*

$$P \stackrel{\text{def}}{=} \frac{\mathbb{I}}{M} + \frac{1}{2}\sum_{i=1}^{M^2-1} \vec{r}_i\sigma_i\,. \tag{5.12}$$

Then P *is positive semidefinite if and only if* $a_k \geq 0$ *for all* $k \in [M]$*, where* $a_0 = 1$*, and*

$$k \cdot a_k = \sum_{q=1}^{k} (-1)^{q-1}\,\mathrm{Tr}[P^q]a_{k-q}\,. \tag{5.13}$$

We can now state the main result of this section:

Theorem 141. *For all* $0 \leq c < s \leq 1$ *where* $c - s = \Omega(\frac{1}{2^{\mathrm{poly}(n)}})$*, we have* $\mathsf{QMA}_{\mathsf{IS}}^{c,s} \subseteq \mathsf{EXP}$*.*

Proof. Recall that $\mathcal{W}_{\mathsf{IS}} = K_1 \cap K_2$, where K_1 and K_2 are convex, compact, well-bounded $\vec{0}$-centered sets with associated radii $R_1 = \sqrt{2}, R_2 = 4^{p(n)}, r = \frac{1}{4^{p(n)}}$ by Claim 131. Both $\mathrm{WMEM}_\zeta(K_1)$ and $\mathrm{WMEM}_\zeta(K_2)$ can be solved in time exponential in n, for any ζ at least an inverse exponential in n, by Fact 139 and Theorem 138. Moreover, $\mathcal{W}_{\mathsf{IS}}$ contains a $\vec{0}$-centered ball of radius $r_3 = \frac{1}{4^{p(n)}}$. Thus, by Theorem 136, we can solve $\mathrm{WMEM}_\beta(\mathcal{W}_{\mathsf{IS}})$ for any $\beta \geq \zeta \cdot 4^{p(n)+1} \geq \zeta \cdot \frac{2\min(R_1,R_2)}{r_3}$; i.e. any β at least inverse exponential in n.

Consider a problem in $\mathsf{QMA}_{\mathsf{IS}}$ with completeness-soundness gap Δ. By Corollary 135, there is an exponential-time reduction to $\mathrm{WMEM}_\beta(\mathcal{W}_{\mathsf{IS}})$, where $\beta = \Omega(\frac{\Delta}{2^{\mathrm{poly}(n)}})$. When $\Delta = \Omega(\frac{1}{2^{\mathrm{poly}(n)}})$, so is β, and this problem can be decided in exponential time. □

Remark 142. *It is likely that Theorem 141 can be improved to a completeness-soundness gap that is inverse* double *exponential in* n*. This would require two changes. First, the second step needs a reduction that runs in* $\mathrm{poly}(\log 1/\epsilon)$ *time; this may be satisfied by the shallow-cut ellipsoid method Grötschel et al. [1988], Liu [2007]. Second, membership testing*

of K_2 needs a program that runs in $\mathrm{poly}(\log 1/\epsilon)$ *time; this can plausibly be done using an improvement to the Doherty-Parrilo-Spedalieri SDP Harrow et al. [2017].*

5.6 Purity and internal separability: $\mathsf{QMA}_{|\mathsf{IS}\rangle} = \mathsf{NEXP}$

In this section we prove there exist constants c, s representing completeness and soundness for which $\mathsf{QMA}^{c,s}_{|\mathsf{IS}\rangle}$ (the variant of $\mathsf{QMA}^{c,s}_{\mathsf{IS}}$ restricted to pure quantum proofs) can decide NEXP.

The starting point of our approach is the Blier-Tapp protocol to decide the NP-hard 3-COLORING graph problem Blier and Tapp [2010] in $\mathsf{QMA}_{\log}(2)$, i.e. $\mathsf{QMA}(2)$ with two unentangled quantum proofs each on $O(\log n)$ qubits. In their protocol, the classical NP proof is encoded in $O(\log n)$-qubit quantum states of the form $|\psi\rangle = \sum_{v\in V} |v\rangle\,|c_v\rangle$, where V is the set of vertices of the graph and c_v is a color assigned to vertex v. So $|\psi\rangle$ encodes a coloring to the graph. With two copies of a quantum proof in this form, the verifier can measure both in the computational basis, obtaining two random vertices v, w with their assigned colors c_v, c_w. If the two measured vertices are neighbors, one can check whether or not the color constraint is satisfied on this edge. When the two proofs are *unentangled*, Arthur can verify that the proofs have the "correct" form: (i) each vertex is assigned exactly one color, and (ii) the coloring is consistent across the two proofs.

The catch is that the protocol decides NP with an *inverse polynomial* completeness-soundness gap. Thus, a straightforward application of the same protocol (for a succinct version of 3-colorability Papadimitriou and Yannakakis [1986]) decides NEXP with *inverse exponential* completeness-soundness gap. This is too small a gap for $\mathsf{QMA}(2)$.[12] What causes this small gap in the Blier-Tapp protocol? The first source is that when the graph is not 3-colorable, there may be only *one* edge that violates the coloring, so it is very difficult to sample this edge. This can be overcome by taking the PCP version of 3-COLORING.

12. Intuitively, Arthur most often sees two vertices which are not connected by an edge and so he cannot check the validity of the coloring. In the $O(\log n)$-qubit version he saw adjacent vertices with an inverse-polynomial probability, but now this occurs with an inverse-exponential probability.

The second source is that for *sparse* graphs, the probability of that two vertices chosen at random are adjacent is inverse polynomial in the number of vertices. One may try to fix this with a *dense* graph; however, since one can efficiently approximate 3-COLORING on *dense* graphs Arora et al. [1999], we can no longer use a PCP to fix the first issue.

To simultaneously overcome both issues in the Blier-Tapp protocol and appropriately scale up to NEXP, Jeronimo and Wu Jeronimo and Wu [2023a] consider a different succinct gapped constraint satisfaction problem (CSP). Their protocol follows a similar outline.[13] A key part of the protocol in Jeronimo and Wu [2023a] is to enforce the same "correct" form of each quantum proof, as used in the Blier-Tapp protocol. This form was later named *rigid* in Bassirian et al. [2024b]:

Definition 143 (Rigid and quasirigid states)**.** *A bipartite state* $|\psi\rangle \in \mathbb{C}^R \otimes \mathbb{C}^\kappa$ *is called quasirigid if it is of the form* $|\psi\rangle = \sum_{v\in[R]} \alpha_v |v\rangle |c_v\rangle$ *for some complex numbers* $\{\alpha_v\}_{v\in[R]}$ *and indices* $\{c_v \in [\kappa]\}_{v\in[R]}$*, and moreover called* rigid *if all* $\alpha_v = \frac{1}{\sqrt{R}}$.

However, Jeronimo and Wu [2023a] only managed to enforce the *rigidity* of each quantum proof by requiring additional structure, i.e. the proofs are unentangled, *and* have non-negative amplitudes in the computational basis. Bassirian et al. [2024b] show how to use only the non-negative amplitudes requirement (i.e. without imposing unentanglement) to get the same result. This is done using two distinct tests: (i) check that each vertex is assigned exactly one color (i.e. *quasirigid* or *valid*), and (ii) check that every vertex is represented in the proof with nontrivial weight.

In this section we go the other direction: we enforce that a quantum proof is *rigid* for pure states with our specific entanglement structure of *internal separability*. We closely follow the recipe of Bassirian et al. [2024b], Bassirian and Marwaha [2024], adjusting some of their claims (especially [Bassirian et al., 2024b, Lemma 10]) to our scenario.

13. Each quantum proof, instead of encoding each vertex and a chosen color, now encodes a each constraint and a list of assigned values to its corresponding variables. For ease of discussion, we continue to call the registers "vertex" and "color" registers. This is just a difference in notation, as each constraint only uses a constant number of variables.

5.6.1 A warm-up with the computational basis

In Bassirian and Marwaha [2024], Bassirian and Marwaha create a framework to unify the previous results. They observed that a variant of QMA, where Arthur is equipped with a fantastical operation they call a "superposition detector", can decide NEXP. This operation inputs a quantum state, and decides whether the quantum state is a computational basis element, or ϵ-far from any such element. It empowers Arthur to enforce condition (i) (quasirigidity) — i.e. ensuring the proof has one color per vertex. For condition (ii), Arthur needn't any special equipment. He tests that all vertices are in the proof with similar weight by projecting to the uniform superposition state $|+\rangle$. While this test is simple, it comes at a price: the honest proof state passes it only with probability $1/\kappa$, strictly less than 1. This creates some difficulty when choosing the probabilities of running each test in a way that decides the problem, but it typically can be overcome.

A superposition detector can be easily implemented given two copies of $|\psi\rangle$: measure each copy in the computational basis, and accept iff the outcomes agree. But in Bassirian and Marwaha [2024], this operation is applied to a single quantum proof after a partial measurement in the computational basis. Nonetheless, Bassirian and Marwaha [2024] shows how to efficiently implement this operation if the proof has a particular form, which includes states with non-negative amplitudes (showing the main result of Bassirian et al. [2024b] within the framework).

Here, we warm up by considering two entanglement guarantees that are stronger than $\mathcal{W}_{|\mathsf{IS}\rangle}$. We show how to decide NEXP with each entanglement guarantee using the framework of Bassirian and Marwaha [2024]. We progressively relax the guarantees, eventually working with $\mathcal{W}_{|\mathsf{IS}\rangle}$ in the next subsection.

We start with a set of tripartite proofs — where the first register may contain $O(\mathrm{poly}(n))$ qubits, and the other two registers are of constant size — that roughly corresponds to having "two copies" of the color register. Precisely, the set is $\mathcal{W}_{|\mathrm{dup}\rangle} \stackrel{\mathrm{def}}{=} \{|\psi\rangle \mid |\psi\rangle = \sum_i \alpha_i |i\rangle |\phi_i\rangle |\phi_i\rangle\}$. Thus, after measuring any outcome $|i\rangle$, the reduced state is some $|\phi_i\rangle |\phi_i\rangle$.

Since the reduced state has two copies, we can implement a superposition detector and enforce condition (i). Thus, a quantum proof from this state must be close to $\sum_v \alpha_v |v\rangle |c_v\rangle |c_v\rangle$ (where $|c_v\rangle$ is a computational basis state). We may convert such a state to a *quasirigid* state by applying a CNOT on the last two registers. From here, the framework of Bassirian and Marwaha [2024] shows how to choose the test probabilities to enforce a rigid state, and thus decide NEXP.

We now slightly relax the guarantee by considering the set of computational basis purifications of separable states: $\mathcal{W}_{|\text{comp}\rangle} = \{|\psi\rangle \mid |\psi\rangle = \sum_i \alpha_i |i\rangle |\mu_i\rangle |\nu_i\rangle\}$. The difference between $\mathcal{W}_{|\text{dup}\rangle}$ and $\mathcal{W}_{|\text{comp}\rangle}$ is that $\mathcal{W}_{|\text{comp}\rangle}$ contains states where $|\mu_i\rangle$ and $|\nu_i\rangle$ are different. We claim that a similar argument as before shows $\mathsf{QMA}_{|\text{comp}\rangle} = \mathsf{NEXP}$.

Claim 144 (Inspired by [Bassirian and Marwaha, 2024, Claim 21])**.** *Consider a quantum state of the form $|\mu\rangle |\nu\rangle$, where each register has k qubits. For $0 < \epsilon \leq 1/2$, there is an efficient quantum protocol that decides if (YES) $|\mu\rangle |\nu\rangle$ is two copies of a computational basis element $|e\rangle |e\rangle$, or (NO) ϵ-far in trace distance from any such basis element, with completeness and soundness gap $2x(1-x)$ for $x = \min(\epsilon, \sqrt{1-\epsilon}, \sqrt{\frac{1-\epsilon}{2^k}})$.*

Proof. Let $|\mu\rangle = \sum_j \beta_j |j\rangle$ and $|\nu\rangle = \sum_j \delta_j |j\rangle$, where $\sum_j |\beta_j|^2 = \sum_j |\delta_j|^2 = 1$. The protocol measures each state in the computational basis and accepts iff they agree. This succeeds with probability

$$\sum_j |\beta_j \delta_j|^2 . \tag{5.14}$$

In the YES case, $|\mu\rangle = |\nu\rangle$ both are equal to some computational basis element $|j\rangle$, and so the protocol succeeds with probability 1.

In the NO case, $|\beta_j \delta_j|^2 = |\langle \mu, \nu | j, j\rangle|^2 \leq 1 - \epsilon$ for every computational basis element $|j\rangle$. We suppose the probability of success is at least $1 - 2\epsilon$, since otherwise there is a completeness-soundness gap of 2ϵ. Let j^* be a value that maximizes $|\beta_j \delta_j|^2$ (break ties

arbitrarily). Then $|\beta_{j^*}\delta_{j^*}|^2 \geq \frac{1-2\epsilon}{2^k}$. So the probability of success is

$$\sum_j |\beta_j \delta_j|^2 = |\beta_{j^*}\delta_{j^*}|^2 + \sum_{j \neq j^*} |\beta_j \delta_j|^2 \tag{5.15}$$

$$\leq |\beta_{j^*}\delta_{j^*}|^2 + \Big(\sum_{j \neq j^*} |\beta_j| \cdot |\delta_j|\Big)^2 \tag{5.16}$$

$$\leq |\beta_{j^*}\delta_{j^*}|^2 + \Big(\sum_{j \neq j^*} |\beta_j|^2\Big) \cdot \Big(\sum_{j \neq j^*} |\delta_j|^2\Big) \tag{5.17}$$

$$\leq |\beta_{j^*}\delta_{j^*}|^2 + (1 - |\beta_{j^*}|^2) \cdot (1 - |\delta_{j^*}|^2)\,. \tag{5.18}$$

The second inequality follows by Cauchy-Schwarz. We next invoke the AM-GM inequality: for non-negative x, y, we have $2\sqrt{xy} \leq x + y$. This implies

$$(1-x)(1-y) = 1 - x - y + xy \leq 1 - 2\sqrt{xy} + xy = (1 - \sqrt{xy})^2\,. \tag{5.19}$$

So the probability of acceptance is at most $|\beta_{j^*}\delta_{j^*}|^2 + (1 - |\beta_{j^*}\delta_{j^*}|)^2 = 1 - 2|\beta_{j^*}\delta_{j^*}| \cdot (1 - |\beta_{j^*}\delta_{j^*}|)$. This expression is maximized at the smallest and largest values of $|\beta_{j^*}\delta_{j^*}|$. By definition of the NO case, $|\beta_{j^*}\delta_{j^*}| \leq \sqrt{1-\epsilon}$, and by assumption, $|\beta_{j^*}\delta_{j^*}| \geq \sqrt{\frac{1-\epsilon}{2^k}}$. □

Since we can still implement the superposition detector with a single proof from $\mathcal{W}_{|\text{comp}\rangle}$, we may use the framework of Bassirian and Marwaha [2024] to enforce rigidity of our proof (up to a CNOT gate), and so $\mathsf{QMA}_{|\text{comp}\rangle} = \mathsf{NEXP}$.

5.6.2 Handling purifications in an arbitrary basis

In $\mathcal{W}_{|\text{comp}\rangle}$, after a computational basis measurement in the first register, the post measurement state is *always* separable. We took advantage of this structure to implement a superposition detector in Claim 144. Unfortunately, showing that $\mathsf{QMA}_{|\mathsf{IS}\rangle} = \mathsf{NEXP}$ is not quite as simple. The main difference between $\mathcal{W}_{|\mathsf{IS}\rangle}$ and $\mathcal{W}_{|\text{comp}\rangle}$ is that in $\mathcal{W}_{|\mathsf{IS}\rangle}$, the purification register can be oriented in any basis. For example, the below state belongs to $\mathcal{W}_{|\mathsf{IS}\rangle}$,

but not $\mathcal{W}_{|\mathrm{comp}\rangle}$:

$$\tfrac{1}{\sqrt{2}}(|+\rangle|0\rangle|0\rangle + |-\rangle|1\rangle|1\rangle) = \tfrac{1}{2}(|0\rangle(|00\rangle + |11\rangle) + |1\rangle(|00\rangle - |11\rangle)). \tag{5.20}$$

To amend this issue, we add data to the second register to mimic the behavior of states in $\mathcal{W}_{|\mathrm{comp}\rangle}$.

Consider states on four registers $\mathbb{C}^R \otimes (\mathbb{C}^\kappa \otimes \mathbb{C}^R) \otimes \mathbb{C}^\kappa$, where $R = 2^{\mathrm{poly}(n)}$ and κ is some constant. Let p_1, p_2 be probabilities that we set later. Our protocol is as follows:

- *Enforce "quasirigidity".* With probability p_1 run MatchCheck — measure all the registers in the computational basis, and make sure it is of the form $|v\rangle|c\rangle|v\rangle|c\rangle$ (i.e. the first pair of outcomes *matches* the second pair of outcomes).

- *Enforce "density".* With probability p_2 run $\mathrm{CNOT}_{1,3}$ (i.e. applied to the first and third register), then $\mathrm{CNOT}_{2,4}$, and check Density — measure the first two registers in the Hadamard (X) basis and accept if all the qubits are $|+\rangle$.

- *Estimate CSP value.* Otherwise run $\mathrm{CNOT}_{1,3}$ and $\mathrm{CNOT}_{2,4}$, and then run the constraint tests of Bassirian et al. [2024b] on the first two registers, which check that the encoded classical assignment satisfies the CSP.

Why does this work? It turns out that MatchCheck combined with the internal separability guarantee enforces the state (after applying CNOTs) to be close to *quasirigid*. Adding the Density check (after CNOTs) enforces it to be close to *rigid*. However, to show this, we must reprove adjusted versions of lemmas from Bassirian et al. [2024b], Bassirian and Marwaha [2024].

Consider an arbitrary tripartite state

$$|\psi\rangle \stackrel{\mathrm{def}}{=} \sum_{v,c,w,d} a_{vc,wd} |v\rangle_A |c,w\rangle_B |d\rangle_C \ , \tag{5.21}$$

and let $\rho_{B,C} = \mathrm{tr}_A(|\psi\rangle\langle\psi|)$ be the reduced density matrix after tracing out register A. Then the success probability of MatchCheck is exactly $\sum_{vc} |a_{vc,vc}|^2$.

Now consider $|\psi\rangle$ after applying $\mathrm{CNOT}_{1,3}$ and $\mathrm{CNOT}_{2,4}$:

$$\mathrm{CNOT}_{1,3}\mathrm{CNOT}_{2,4} |\psi\rangle = \sum_{vc,wd} a_{vc,wd} |v\rangle |c\rangle |w \oplus v\rangle |d \oplus c\rangle \ . \tag{5.22}$$

The last two registers are $|0\rangle |0\rangle$ exactly when $(v, c) = (w, d)$, which happens with *exactly* the probability that MatchCheck succeeds. So conditioned on the last two registers being $|0\rangle |0\rangle$, the new state is proportional to

$$\left(\sum_{vc} a_{vc,vc} |v\rangle |c\rangle \right) |0\rangle |0\rangle \ . \tag{5.23}$$

It remains to show that the new state is quasirigid. Suppose MatchCheck passes with high probability; then for a typical $|v\rangle_A$ in state Eq. (5.21), the two color registers are the roughly in superposition of $|cc\rangle$ for allowed colors c. If the superposition consists more than one color, it adds some entanglement between the B and C registers. Since we have that state $\rho_{B,C}$ is *separable*, then $|\psi\rangle$ must be close to quasirigid.

To prove this claim formally, we require a technical lemma:

Lemma 145. *Consider a state $\rho \in \mathcal{D}(\mathbb{C}^m, \mathbb{C}^n)$ that is separable. Then for any $w, y \in [m]$ and $x, z \in [n]$, we have*

$$|\langle wx| \rho |yz\rangle| \leq \sqrt{\min(\langle wx| \rho |wx\rangle, \langle wz| \rho |wz\rangle, \langle yx| \rho |yx\rangle, \langle yz| \rho |yz\rangle)}, \tag{5.24}$$

and moreover,

$$|\langle wx| \rho |yz\rangle| \leq \frac{1}{2} \cdot \min(\langle wx| \rho |wx\rangle + \langle yz| \rho |yz\rangle, \langle wz| \rho |wz\rangle + \langle yx| \rho |yx\rangle). \tag{5.25}$$

Proof. Fix a separable decomposition of ρ of the form $\rho = \sum_\iota p_\iota M_\iota \otimes N_\iota$, where all $p_\iota \geq 0$,

$\sum_\iota p_\iota = 1$, and M_ι, N_ι are trace-1 and positive semidefinite. Choose any $w, y \in [m]$ and $x, z \in [n]$. Then

$$|\langle wx| \rho |yz\rangle| \leq \left| \sum_\iota p_\iota \langle w| M_\iota |y\rangle \cdot \langle x| N_\iota |z\rangle \right| \leq \sum_\iota p_\iota |\langle w| M_\iota |y\rangle| \cdot |\langle x| N_\iota |z\rangle| \,. \tag{5.26}$$

Recall that PSD matrices A decompose as $A = B^\dagger B$, and so by Cauchy-Schwarz,

$$|\langle i| A |k\rangle|^2 = \left| (B|i\rangle)^\dagger (B|k\rangle) \right|^2 \leq \langle i| B^\dagger B |i\rangle \cdot \langle k| B^\dagger B |k\rangle = \langle i| A |i\rangle \cdot \langle k| A |k\rangle \,, \tag{5.27}$$

So then $|\langle wx| \rho |yz\rangle| \leq \sum_\iota p_\iota \sqrt{\langle w| M_\iota |w\rangle \cdot \langle y| M_\iota |y\rangle \cdot \langle x| N_\iota |x\rangle \cdot \langle z| N_\iota |z\rangle}$.

For any PSD matrix A and vector $|\psi\rangle$, we have $0 \leq \langle \psi| A |\psi\rangle \leq \mathrm{Tr}(A)$. All M_ι, N_ι are trace-1, so

$$|\langle wx| \rho |yz\rangle| \leq \sum_\iota p_\iota \sqrt{\langle w| M_\iota |w\rangle \cdot \langle z| N_\iota |z\rangle} \tag{5.28}$$

$$\leq \sqrt{\sum_\iota p_\iota \langle w| M_\iota |w\rangle \cdot \langle z| N_\iota |z\rangle} \tag{5.29}$$

$$= \sqrt{\langle wz| \rho |wz\rangle} \,, \tag{5.30}$$

where the second inequality is Jensen's. The other bounds follow mutatis mutandis.

Moreover, we can use the AM-GM inequality; i.e. for non-negative a, b, we have $\sqrt{ab} \leq \frac{a+b}{2}$. So

$$|\langle wx| \rho |yz\rangle| \leq \frac{1}{2} \sum_\iota p_\iota \left(\langle w| M_\iota |w\rangle \cdot \langle x| N_\iota |x\rangle + \langle y| M_\iota |y\rangle \cdot \langle z| N_\iota |z\rangle \right) \tag{5.31}$$

$$= \frac{\langle wx| \rho |wx\rangle + \langle yz| \rho |yz\rangle}{2} \,. \tag{5.32}$$

Similarly, $|\langle wx| \rho |yz\rangle| \leq \frac{1}{2} \cdot (\langle wz| \rho |wz\rangle + \langle yx| \rho |yx\rangle)$. □

Lemma 146. *Suppose* MatchCheck *succeeds with probability* $1 - \epsilon$*, and* $\rho_{B,C}$ *is separa-*

ble. Then there is a state $|\phi\rangle$ *that is* quasirigid *after applying* $CNOT_{1,3}CNOT_{2,4}$ *such that* $|\langle\phi|\psi\rangle|^2 \geq 1-(\kappa+1)\epsilon$.[14]

Proof. Label the computational basis elements of $|\psi\rangle$ as $a_{vc,wd}|v\rangle|c,w\rangle|d\rangle$ as in Eq. (5.21). Let $\sigma : [R] \to [\kappa]$ be a function that picks out the largest amplitude per vertex, i.e. where $|a_{v\sigma(v),v\sigma(v)}|^2 \geq |a_{vd,vd}|^2$ for all $d \neq \sigma(v)$. For convenience, denote $\sigma_v \stackrel{\text{def}}{=} \sigma(v)$.

Consider the matrix elements of the reduced density matrix of $|\psi\rangle$ after tracing out the first register:

$$\langle c,v,d|\rho_{B,C}|e,w,f\rangle = \sum_x a_{xc,vd}a^\dagger_{xe,wf}\,. \tag{5.33}$$

We use this expression to upper-bound the weight on colors not chosen by σ:

$$\sum_v \sum_{d;d\neq\sigma_v} |a_{vd,vd}|^2 \leq \sum_v \sum_{d;d\neq\sigma_v} |a_{v\sigma_v,v\sigma_v}|\cdot|a^\dagger_{vd,vd}| \tag{5.34}$$

$$= \sum_v \sum_{d;d\neq\sigma_v} |a_{v\sigma_v,v\sigma_v}\cdot a^\dagger_{vd,vd}| \tag{5.35}$$

$$= \sum_v \sum_{d;d\neq\sigma_v} \left(|\langle\sigma_v,v,\sigma_v|\rho_{B,C}|d,v,d\rangle - \sum_{x;x\neq v} a_{x\sigma_v,v\sigma_v}\cdot a^\dagger_{xd,vd}| \right) \tag{5.36}$$

$$\leq \left(\sum_v \sum_{d;d\neq\sigma_v} |\langle\sigma_v,v,\sigma_v|\rho_{B,C}|d,v,d\rangle| \right) \tag{5.37}$$

$$+ \sum_v \sum_{d;d\neq\sigma_v} \sum_{x;x\neq v} |a_{x\sigma_v,v\sigma_v}|\cdot|a^\dagger_{xd,vd}|\,. \tag{5.38}$$

The second term is actually quite small. By the AM-GM inequality,

$$|a_{x\sigma_v,v\sigma_v}|\cdot|a^\dagger_{xd,vd}| \leq \frac{1}{2}\cdot\left(|a_{x\sigma_v,v\sigma_v}|^2 + |a^\dagger_{xd,vd}|^2\right)\,. \tag{5.39}$$

14. Recall that κ is the dimension of the third register. Although this lemma resembles [Bassirian and Marwaha, 2024, Lemma 18], note that ϵ has a different meaning in this paper.

Taking $|a_{xd,vd}|^2$ over all three sums (v,d,x) is at most $\sum_{(x,c)\neq(v,d)} |a_{xc,vd}|^2 = 1-(1-\epsilon) = \epsilon$. Similarly, summing $|a_{x\sigma_v,v\sigma_v}|^2$ over sums (v,x) contributes at most ϵ, so at most $\kappa\epsilon$ over sums (v,d,x). In total, the second term is at most $\frac{\kappa+1}{2}\epsilon$.

For the first term, since $\rho_{B,C}$ is separable, we can use one of the bounds proven in Lemma 145, i.e.

$$|\langle\sigma_v,v,\sigma_v|\rho_{B,C}|d,v,d\rangle| \leq \frac{1}{2}\cdot\left(\langle\sigma_v,v,d|\rho_{B,C}|\sigma_v,v,d\rangle + \langle d,v,\sigma_v|\rho_{B,C}|d,v,\sigma_v\rangle\right). \quad (5.40)$$

So then

$$\sum_v \sum_{d;d\neq\sigma_v} |a_{vd,vd}|^2 \leq \frac{\kappa+1}{2}\epsilon \quad (5.41)$$

$$+\frac{1}{2}\left(\sum_v \sum_{d;d\neq\sigma_v} \langle\sigma_v,v,d|\rho_{B,C}|\sigma_v,v,d\rangle + \langle d,v,\sigma_v|\rho_{B,C}|d,v,\sigma_v\rangle\right) \quad (5.42)$$

$$\leq \frac{\kappa+1}{2}\epsilon + \frac{1}{2}\left(\sum_v \sum_{d;d\neq\sigma_v} \sum_x |a_{x\sigma_v,vd}|^2 + |a_{xd,v\sigma_v}|^2\right). \quad (5.43)$$

The remaining sum is at most $\sum_{(x,c)\neq(v,d)} |a_{xc,vd}|^2 = \epsilon$, so the upper bound is at most $\frac{\kappa+2}{2}\epsilon \leq \kappa\epsilon$.

Consider the state $|\phi\rangle = \frac{1}{\sqrt{\gamma}}\sum_v a_{v\sigma_v,v\sigma_v}|v\rangle|\sigma_v\rangle|v\rangle|\sigma_v\rangle$, where $\gamma = \sum_v |a_{v\sigma_v,v\sigma_v}|^2$ is chosen so that $\langle\phi|\phi\rangle = 1$. Note that $|\phi\rangle$ is quasirigid after applying $\mathrm{CNOT}_{1,3}$ and $\mathrm{CNOT}_{2,4}$. Moreover, $|\langle\phi|\psi\rangle|^2 = \gamma$. Putting everything together,

$$\gamma = \sum_v |a_{v\sigma_v,v\sigma_v}|^2 = \mathbf{Pr}[\mathsf{MatchCheck}\text{ succeeds}] - \sum_v \sum_{d;d\neq\sigma_v} |a_{vd,vd}|^2 \geq 1-\epsilon-\kappa\epsilon. \qquad \square$$

We also use the following fact in several places:

Fact 147 (Fuchs and van de Graaf [1998])**.** *Let $0 \leq \Pi \leq \mathbb{I}$ be a positive semi-definite matrix, and let $|\psi_1\rangle$ and $|\psi_2\rangle$ be quantum states such that $|\langle\psi_1|\psi_2\rangle|^2 \geq 1-d$. Then $|\langle\psi_1|\Pi|\psi_1\rangle -$*

$\langle\psi_2|\,\Pi\,|\psi_2\rangle\,| \leq \sqrt{d}$.

We are now ready to show that a quantum proof passing both tests must be close to quasirigid:

Claim 148. *Suppose* Density *and* MatchCheck *succeed on* $|\psi\rangle$ *with probability* $\frac{1}{\kappa} - d_{\mathsf{D}}$ *and* $1 - d_{\mathsf{M}}$*, respectively. Then there exists a state* $|\chi\rangle$ *that is rigid after applying* $CNOT_{1,3}CNOT_{2,4}$ *such that*

$$|\,\langle\chi|\psi\rangle\,|^2 \geq 1 - \kappa d_{\mathsf{D}} - (\kappa+1)\sqrt{(\kappa+1)d_{\mathsf{M}}}\,. \tag{5.44}$$

Proof. By Lemma 146, there exists a state $|\phi\rangle = \sum_v \alpha_v\,|v\rangle\,|c_v, v\rangle\,|c_v\rangle$ such that

$$|\,\langle\psi|\phi\rangle\,|^2 \geq 1 - (\kappa+1)d_{\mathsf{M}}\,. \tag{5.45}$$

By Fact 147, for any unit vector $|\mu\rangle$, we have $\big|\,|\,\langle\mu|\psi\rangle\,|^2 - |\,\langle\mu|\phi\rangle\,|^2\big| \leq \sqrt{(\kappa+1)d_{\mathsf{M}}}$. We use this in two places. Let $\big|+'\big\rangle = \mathrm{CNOT}_{1,3}\mathrm{CNOT}_{2,4}\,|+\rangle\,|0\rangle\,|0\rangle$. First, since $|\,\big\langle+'\big|\psi\big\rangle\,|^2 \geq \frac{1}{\kappa} - d_{\mathsf{D}}$, applying Fact 147 with $|\mu\rangle = \big|+'\big\rangle$ implies $|\,\big\langle+'\big|\phi\big\rangle\,|^2 \geq \frac{1}{\kappa} - d_{\mathsf{D}} - \sqrt{(\kappa+1)d_{\mathsf{M}}}$. Now let $|\chi\rangle = \frac{1}{\sqrt{R}}\sum_v |v\rangle\,|c_v, v\rangle\,|c_v\rangle$ be the state with the same computational basis elements as $|\phi\rangle$. Then

$$|\,\langle\chi|\phi\rangle\,|^2 = \left|\frac{1}{\sqrt{R}}\sum_{v\in[R]}\alpha_v\right|^2 = \kappa\left|\frac{1}{\sqrt{\kappa R}}\sum_{v\in[R]}^{R}\alpha_v\right|^2 \tag{5.46}$$

$$= \kappa|\,\big\langle+'\big|\phi\big\rangle\,|^2 \geq 1 - \kappa\left(d_{\mathsf{D}} + \sqrt{(\kappa+1)d_{\mathsf{M}}}\right)\,. \tag{5.47}$$

Applying Fact 147 using $|\mu\rangle = |\chi\rangle$ then implies

$$|\,\langle\chi|\psi\rangle\,|^2 \geq 1 - \kappa\left(d_{\mathsf{D}} + \sqrt{(\kappa+1)d_{\mathsf{M}}}\right) - \sqrt{(\kappa+1)d_{\mathsf{M}}} = 1 - \kappa d_{\mathsf{D}} - (\kappa+1)\sqrt{(\kappa+1)d_{\mathsf{M}}}\,.$$

□

Recall that in the YES case, the Density test succeeds with probability strictly less than

1. A conniving Merlin in the NO case could try to take advantage of this fact. To mitigate this issue, we show that if the Density test succeeds with probability higher than in the YES case, the success of the MatchCheck test is penalized. We prove a tradeoff between the success of these two tests:

Lemma 149. *Suppose* Density *and* MatchCheck *succeed on state* $|\psi\rangle$ *with probability* $w_{\mathsf{D}} \geq \frac{1}{\kappa}$ *and* w_{M}*, respectively. Then* $(w_{\mathsf{D}} - \frac{1}{\kappa})^2 + (\kappa+1)w_{\mathsf{M}} \leq \kappa + 1$.

Proof. By Lemma 146, there exists a $|\phi\rangle = \sum_{v\in[R]} \alpha_v |v\rangle |c_v, v\rangle |c_v\rangle$ such that $|\langle\psi|\phi\rangle|^2 \geq 1 - (\kappa+1)(1-w_{\mathsf{M}})$. Using Fact 147, we see that $\left| |\langle +'|\psi\rangle|^2 - |\langle +'|\phi\rangle|^2 \right|$ is at most $\sqrt{(\kappa+1)(1-w_{\mathsf{M}})}$. Note that by assumption, $|\langle +'|\psi\rangle|^2 = w_{\mathsf{D}} \geq \frac{1}{\kappa}$, and by Cauchy-Schwarz, $|\langle +'|\phi\rangle|^2 = \frac{1}{R\cdot\kappa}\left|\sum_{v\in[R]} \alpha_v\right|^2 \leq \frac{1}{\kappa}\sum_{v\in[R]} |\alpha_v|^2 = \frac{1}{\kappa}$. Combining these claims, we have

$$w_{\mathsf{D}} - \frac{1}{\kappa} \leq |\langle +'|\psi\rangle|^2 - |\langle +'|\phi\rangle|^2 \leq \sqrt{(\kappa+1)(1-w_{\mathsf{M}})}\,. \tag{5.48}$$

The lemma follows after rearranging the terms. □

After following the recipe in Bassirian and Marwaha [2024] and adjusting many tools from there, we have in Claim 148 and Lemma 149 two tests that enforce *rigidity* of the quantum proof. These statements allow us to set the probability of running each test (i.e. p_1, p_2) in the protocol. We defer the full calculation to Section 5.13. Once this is done, we reach our goal in this section:

Theorem 150. *There exist constants* $1 > c > s > 0$ *such that* $\mathsf{QMA}^{c,s}_{|\mathsf{IS}\rangle} = \mathsf{NEXP}$.

5.7 Two copies enforce purity: $\mathsf{QMA}_{|\mathsf{IS}\rangle} \subseteq \mathsf{QMA}_{\mathsf{IS}}(2)$

We can use two unentangled proofs to implement a SWAP test. For product states, this ensures the unentangled proofs are *pure*. Consider a convex set of quantum states $\mathcal{W}$ with the property that nearly-pure states in $\mathcal{W}$ are always close to a state in rank1($\mathcal{W}$), i.e. a

pure state in $\mathcal{W}$. When this is the case, we show that any protocol in $\mathsf{QMA}_{\mathrm{rank1}(\mathcal{W})}$ can be simulated in $\mathsf{QMA}_{\mathcal{W}}(2)$.

Definition 151 (Continuity condition of $\mathcal{W}$)**.** *A set of quantum states $\mathcal{W}$ has the* continuity condition *if there exists a number $0 < a \leq 1$ and a continuous, strictly monotonic function $f : [0, a] \to [0, 1]$ with $f(0) = 0$ satisfying the following property: For any $\rho \in \mathcal{W}$, if the maximum eigenvalue of ρ is $1 - x$, then there is some $|\psi\rangle \in \mathrm{rank1}(\mathcal{W})$ where $\langle\psi| \rho |\psi\rangle \geq 1 - f(x)$.*

The continuity condition guarantees that $\rho \in \mathcal{W}$ is ϵ-close to a pure state only if it is $f(\epsilon)$-close to a pure state *in* $\mathcal{W}$. Although this condition is satisfied by many sets $\mathcal{W}$, for example the set of separable states, there exist pathological sets without this condition.[15]

Lemma 152. *Consider any protocol V in $\mathsf{QMA}_{\mathrm{rank1}(\mathcal{W})}$ with completeness-soundness gap Δ. Suppose also that $\mathcal{W}$ has the continuity condition with function f. Then $\mathsf{QMA}_{\mathcal{W}}(2)$ can simulate V with completeness-soundness gap $\Omega(\Delta \cdot f^{-1}(\frac{\Delta^2}{4}))$.*

Proof. The quantum proof in $\mathsf{QMA}_{\mathcal{W}}(2)$ is some product state $\rho_1 \otimes \rho_2$ where $\rho_1, \rho_2 \in \mathcal{W}$. The verifier runs the following procedure, where we fix the value of p later on:

- With probability p, run the SWAP test.
- Otherwise, choose $i \in \{1, 2\}$ uniformly at random and run V on ρ_i.

The SWAP test passes with probability $\frac{1}{2}(1 + \mathrm{Tr}[\rho_1\rho_2])$. Notice that if this is 1, then $\rho_1 = \rho_2$ is a pure state; i.e. $\rho_1, \rho_2 \in \mathrm{rank1}(\mathcal{W})$. In completeness, Merlin sends an honest quantum proof, so the procedure accepts with probability at least $p + (1 - p)(s + \Delta)$, where s is the soundness guaranteed by V. We'll split the analysis of soundness into cases depending on the maximum overlap of ρ_1 and ρ_2 with a pure state.

15. For example, let $P_d \subseteq \mathcal{D}(\mathbb{C}^d)$ be the set of states with overlap at least $1 - \frac{1}{d}$ with some pure state, and define $\mathcal{G} \stackrel{\mathrm{def}}{=} \bigcup_d (\mathcal{D}(\mathbb{C}^d) \setminus P_d)$. Then $\mathcal{G}$ has states that are arbitrarily close to pure, yet $\mathrm{rank1}(\mathcal{G})$ is empty.

Let $\rho_1 = \sum_i p_i \left|\psi_i\right\rangle\!\left\langle\psi_i\right|$ and $\rho_2 = \sum_j q_j \left|\phi_j\right\rangle\!\left\langle\phi_j\right|$ where $\{|\psi_i\rangle\}$ and $\{|\phi_j\rangle\}$ are orthonormal sets, and the p_i's and q_i's are in decreasing order. Suppose $p_1 q_1 = 1 - \epsilon$. Then each of p_1, q_1 is at least $1 - \epsilon$.

1. The first case is when $\epsilon \geq 1/3$. Here, the proofs are very far from pure, and so the SWAP test fails. Notice that $\mathrm{Tr}[\rho_1\rho_2] = \sum_j q_j \left\langle\phi_j\right| \rho_1 \left|\phi_j\right\rangle \leq \max_{|\chi\rangle} \left\langle\chi\right| \rho_1 \left|\chi\right\rangle = p_1$. Similarly, $\mathrm{Tr}[\rho_1\rho_2] \leq q_1$. So then $\mathrm{Tr}[\rho_1\rho_2]^2 \leq \min(p_1, q_1)^2 \leq p_1 q_1 = 1 - \epsilon$. Thus, the SWAP test succeeds with probability $\frac{1}{2}(1 + \mathrm{Tr}[\rho_1\rho_2]) \leq \frac{1}{2}(1 + \sqrt{2/3})$. So the completeness-soundness gap is at least

$$p + (1-p)(s+\Delta) - \left(p \cdot \frac{1}{2}(1+\sqrt{2/3}) + (1-p)\right) \tag{5.49}$$

$$= p \cdot \frac{1}{2}(1-\sqrt{2/3}) + (1-p)(s+\Delta-1)\,. \tag{5.50}$$

   The gap is at least $\delta > 0$ when the probability p is at least

$$p > \frac{1 - (s+\Delta) + \delta}{1 - (s+\Delta) + \frac{1}{2}(1-\sqrt{2/3})}\,. \tag{5.51}$$

2. Otherwise, $\epsilon < 1/3$. We will need the probability of the SWAP test in this case, whose proof is deferred:

   Claim 153. *When $\epsilon < 1/3$, the SWAP test accepts with probability at most $1 - \frac{1}{2}\epsilon + \frac{1}{2}\epsilon^2$.*

   We next analyze the acceptance probability of the other test. Suppose $i = 1$; the $i = 2$ case is the same by symmetry. Let f be the polynomial associated with the continuity condition of $\mathcal{W}$. Since $p_1 \geq 1 - \epsilon$, ρ_1 has squared overlap at least $1 - f(\epsilon)$ with a state in rank1($\mathcal{W}$). Since one state is pure, the squared overlap is equal to the fidelity.[16] Fuchs-van de Graaf Fuchs and van de Graaf [1998] implies that the trace distance is at most $\sqrt{f(\epsilon)}$. So this test accepts with probability at most $s + \sqrt{f(\epsilon)}$.

16. We use the "squared" version of fidelity, i.e. $F(\rho,\sigma) \stackrel{\text{def}}{=} \left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]\right)^2$.

Altogether, the procedure accepts with probability at most

$$p \cdot (1 - \frac{1}{2}\epsilon + \frac{1}{2}\epsilon^2) + (1-p)(s + \sqrt{f(\epsilon)}), \tag{5.52}$$

and so the completeness-soundness gap is at least

$$\frac{1}{2}p \cdot \epsilon(1-\epsilon) + (1-p)(\Delta - \sqrt{f(\epsilon)}). \tag{5.53}$$

Choose $\gamma > 0$ where $f(\gamma) = \frac{\Delta^2}{4}$.[17] We further divide into two subcases:

(a) When $\epsilon \leq \gamma$, the proof is close to a state in rank1($\mathcal{W}$), so the gap is at least $(1-p)\frac{\Delta}{2}$.

(b) Otherwise, $\gamma < \epsilon < 1/3$. Then the proofs are far from pure and fail the SWAP test, even if $f(\epsilon) = 1$. The gap is at least $\delta > 0$ when the probability p is at least

$$p > \frac{1 - \Delta + \delta}{1 - \Delta + \frac{1}{2}\gamma(1-\gamma)}. \tag{5.54}$$

If we set $\delta = \frac{1}{4}\min(1 - \sqrt{2/3}, \gamma(1-\gamma))$, then we may choose $0 < p < 1$ that satisfies both Eqs. (5.51) and (5.54), producing a completeness-soundness gap $\Omega(\min(\delta, (1-p)\Delta)) = \Omega(\delta\Delta)$. So the completeness-soundness gap is $\Omega(\Delta \cdot f^{-1}(\frac{\Delta^2}{4}))$. □

We finish the proof of Lemma 152 by proving Claim 153.

Proof of Claim 153. Recall that $\mathrm{Tr}[\rho_1\rho_2] = \sum_{i,j} p_i q_j |\langle\psi_i|\phi_j\rangle|^2$. Then

$$\mathrm{Tr}[\rho_1\rho_2] \leq p_1 q_1 |\langle\psi_1|\phi_1\rangle|^2 + \sum_{j\neq 1} q_j |\langle\psi_1|\phi_j\rangle|^2 + \sum_{i\neq 1} p_i |\langle\psi_i|\phi_1\rangle|^2 + \left(\sum_{i\neq 1} p_i\right)\left(\sum_{j\neq 1} q_j\right). \tag{5.55}$$

Let $|\langle\psi_1|\phi_1\rangle|^2 = 1 - z$. We simplify each term in this upper bound:

17. If this is not possible for any $\gamma \leq 1/3$, then set $\gamma = 1/3$, and ignore case (b).

- The first term is $(1-\epsilon)(1-z)$.
- The value $|\langle\psi_1|\phi_j\rangle|^2$ when $j \neq 1$ is at most $1 - |\langle\psi_1|\phi_1\rangle|^2 = z$. So the second term is at most $z(\sum_{j\neq 1} q_j) \leq z\epsilon$.
- By symmetry, the third term is at most $z\epsilon$.
- Since $p_1, q_1 \geq 1-\epsilon$, the last term is at most ϵ^2.

Altogether, we have

$$\mathrm{Tr}[\rho_1\rho_2] \leq (1-\epsilon)(1-z) + 2z\epsilon + \epsilon^2 = 1 - \epsilon + (3\epsilon - 1)z + \epsilon^2 . \tag{5.56}$$

Since $\epsilon < 1/3$, this upper bound is maximized at $z = 0$. So the SWAP test accepts with probability $\frac{1}{2}(1 + \mathrm{Tr}[\rho_1\rho_2]) \leq 1 - \frac{1}{2}\epsilon + \frac{1}{2}\epsilon^2$. □

We now use Lemma 152 to prove that $\mathsf{QMA}_{|\mathsf{IS}\rangle} \subseteq \mathsf{QMA}_{\mathsf{IS}}(2)$.

Claim 154. $\mathsf{QMA}_{|\mathsf{IS}\rangle} \subseteq \mathsf{QMA}_{\mathsf{IS}}(2)$*, and so* $\mathsf{QMA}_{\mathsf{IS}}(2) = \mathsf{NEXP}$.

Proof. Ideally, we would like to show that $\mathcal{W}_{\mathsf{IS}}$ has the continuity condition. However, we are unable to do this when the first register in the proof has too few qubits. Nevertheless, we show that a subset of $\mathcal{W}_{\mathsf{IS}}$ has the continuity condition, and moreover, the protocol for NEXP (Section 5.6) can use quantum proofs in this subset of $\mathcal{W}_{\mathsf{IS}}$. Since the NEXP-hard problem is complete for $\mathsf{QMA}_{|\mathsf{IS}\rangle}$, every problem in $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ can be decided in $\mathsf{QMA}_{\mathsf{IS}}(2)$.

Label the registers of the quantum proof A, B, C, and denote the dimension of any register X as $\dim(X)$. Let ${\mathcal{W}_{\mathsf{IS}}}^* \subseteq \mathcal{W}_{\mathsf{IS}}$ be the set of tripartite states in $\mathcal{W}_{\mathsf{IS}}$ where the registers also satisfy $\dim(A) \geq \dim(B) \cdot \dim(C)$. We show in Section 5.14 that the protocol of Section 5.6 can use a quantum proof in $\mathcal{W}_{|\mathsf{IS}\rangle}$ obeying this register size inequality; thus, $\mathsf{NEXP} \subseteq \mathsf{QMA}_{\mathrm{rank1}({\mathcal{W}_{\mathsf{IS}}}^*)}$.

We now show that the continuity condition holds for ${\mathcal{W}_{\mathsf{IS}}}^*$. Consider any such quantum proof ρ. Suppose the largest eigenvalue-eigenvector pair of ρ is $(\lambda_1, |\psi\rangle)$. The fidelity of

ρ and $|\psi\rangle$ is exactly $\langle\psi|\,\rho\,|\psi\rangle = \lambda_1$. Since fidelity is non-decreasing under CPTP maps, the fidelity of $\mathrm{Tr}_A[\rho]$ and $\mathrm{Tr}_A[|\psi\rangle\langle\psi|]$ is at least λ_1. Since $|\psi\rangle$ purifies $\mathrm{Tr}_A[|\psi\rangle\langle\psi|]$, and $\dim(A) \geq \dim(B)\cdot\dim(C)$, Uhlmann's theorem implies the existence of a purification $\big|\chi_\rho\big\rangle$ of $\mathrm{Tr}_A[\rho]$ that maintains fidelity, i.e. $|\,\big\langle\psi\big|\chi_\rho\big\rangle\,|^2 \geq \lambda_1$.

Note that by construction, $\big|\chi_\rho\big\rangle \in \mathrm{rank1}(\mathcal{W}_{\mathsf{IS}}{}^*)$. Moreover,

$$\big\langle\chi_\rho\big|\,\rho\,\big|\chi_\rho\big\rangle \geq \lambda_1 \big\langle\chi_\rho\big|\,|\psi\rangle\langle\psi|\,\big|\chi_\rho\big\rangle \geq \lambda_1^2\,. \tag{5.57}$$

We verify that $\mathcal{W}_{\mathsf{IS}}$ has the continuity condition with $f(x) = 2x$. Suppose $\lambda_1 = 1-x$; then the squared overlap with $\big|\chi_\rho\big\rangle$ is at least $(1-x)^2 = 1-(2x-x^2) \geq 1-2x$. Lemma 152 then implies that $\mathsf{NEXP} \subseteq \mathsf{QMA}_{\mathcal{W}_{\mathsf{IS}}{}^*}(2)$, which is trivially contained in $\mathsf{QMA}_{\mathcal{W}_{\mathsf{IS}}}(2) = \mathsf{QMA}_{\mathsf{IS}}(2)$. □

We show in Section 5.10 that $\mathsf{QMA}_{\mathsf{IS}}(2)$ at large enough completeness-soundness gap can be simulated by $\mathsf{QMA}(2)$. As a consequence:

Corollary 155. *Gap amplification of* $\mathsf{QMA}_{\mathsf{IS}}(2)$ *implies* $\mathsf{QMA}(2) = \mathsf{NEXP}$.

However, by Theorem 112, the single-copy variant $\mathsf{QMA}_{\mathsf{IS}} \subseteq \mathsf{EXP}$ at all completeness-soundness gaps at least inverse exponential in input size. Unlike QMA^+, gap amplification of $\mathsf{QMA}_{\mathsf{IS}}$ does not obviously imply any complexity collapse. In other words, $\mathsf{QMA}_{\mathsf{IS}}(2)$ has exactly the desired features of Jeronimo and Wu [2023a] while bypassing the issue identified in Bassirian et al. [2024b].

5.8 Open questions

1. Does the complexity class $\mathsf{QMA}_{\mathsf{IS}}$ exhibit gap amplification? This is a good starting point to study gap amplification of $\mathsf{QMA}_{\mathsf{IS}}(2)$, which by Corollary 116 implies $\mathsf{QMA}(2) = \mathsf{NEXP}$.

2. The proof of Theorem 113 uses two quantum proofs only to implement a purity test. Is there a set of proofs $\mathcal{W}$ such that $\mathsf{QMA}_{\mathcal{W}}(2)$ is strictly more powerful than

$\mathsf{QMA}_{\mathrm{rank1}(\mathcal{W})}$? If so, there is a setting where *unentanglement* is useful for something beyond purity testing.

3. Is there a reduction from WVAL to WMEM that preserves a constant gap? If so, we can convert our results on hardness of approximation to the associated testing problems. For example, this would imply NP-hardness of *testing* whether a state is a purification of a separable state or $\Omega(1)$-far from any such state.

4. How essential is the continuity condition (Definition 151) to show $\mathsf{QMA}_{\mathrm{rank1}(\mathcal{W})} \subseteq \mathsf{QMA}_{\mathcal{W}}(2)$? For example, does all of $\mathcal{W}_{\mathsf{IS}}$ satisfy the continuity condition? A positive answer would simplify the proof of Claim 154 by bypassing the argument in Section 5.14. Is there a simple characterization of sets which do *not* satisfy the continuity condition?

5. One problem related to QMA and purity testing is the *pure state marginal problem*. Here, one decides if there exists a *pure* state that is consistent with a given set of reduced density matrices. It is known to be in $\mathsf{QMA}(2)$ and at least QMA-hard, whereas the mixed variant is QMA-complete Liu [2006], Broadbent and Grilo [2022]. Can one exactly characterize the complexity of this problem?

6. A *disentangler* is a hypothetical quantum channel that maps the set of states to separable states and is approximately surjective. As noted by John Watrous (and communicated in Aaronson et al. [2009]), the existence of an efficient disentangler implies $\mathsf{QMA} = \mathsf{QMA}(2)$. We remark that an analogous map from states (or even separable states) to *internally separable* states would imply $\mathsf{QMA}(2) = \mathsf{NEXP}$. Does such a channel exist? In fact, disentanglers that separate $O(1)$ qubits from the rest of the state exist by the quantum de Finetti theorem Caves et al. [2002], Christandl et al. [2007], but they separate a register from the entire system, not a particular subsystem.

7. Our work initiates a complexity-theoretic study of a notion of *multipartite* unentan-

glement. As asked in Question 119, which notions are computationally powerful? For example, what is the power of $\mathsf{QMA}_{\mathcal{W}}$ when $\mathcal{W}$ contains states with a multipartite entanglement structure distinct from internal separability?

5.9 Characterizing QMA(2) as a purity test

The proof of $\mathsf{QMA}_{\mathsf{IS}}(2) = \mathsf{NEXP}$ uses the two unentangled proofs *only* to implement a purity test. We suggest this is not coincidence: many works Harrow and Montanaro [2013], Barak et al. [2017], Yu et al. [2021, 2022] take advantage of the fact that every protocol over separable states is equivalently a protocol over states with a pure subsystem. We formalize this equivalence in a self-contained way.

From Harrow and Montanaro [2013], the optimal success probability of a $\mathsf{QMA}(2)$ verifier V is

$$h_{sep}(V) \stackrel{\text{def}}{=} \max_{\rho \in \mathsf{SEP}(X,Y)} \mathrm{Tr}[V\rho].$$

By convexity, we only need to optimize over pure product states $\rho = |\psi_X\rangle\langle\psi_X| \otimes |\psi_Y\rangle\langle\psi_Y|$. In fact, we only need to optimize over $\rho \in \mathcal{D}(X,Y)$ such that $\mathrm{Tr}_Y[\rho]$ is pure, since these states contain all pure product states:

Claim 156. *Consider a state $\rho \in \mathcal{D}(X,Y)$ where $\mathrm{Tr}_Y[\rho]$ is pure. Then ρ is a product state.*

Proof. Diagonalize $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$. Since $\mathrm{Tr}_Y[\rho] = \sum_i p_i \mathrm{Tr}_Y[|\phi_i\rangle\langle\phi_i|]$ is some pure state $|\psi\rangle\langle\psi|$, all $\mathrm{Tr}_Y[|\phi_i\rangle\langle\phi_i|]$ must be proportional to $|\psi\rangle\langle\psi|$. A pure state has a pure subsystem if and only if it is separable, so $|\phi_i\rangle = |\psi\rangle |\chi_i\rangle$ for all i. So $\rho = |\psi\rangle\langle\psi| \otimes \sum_i p_i |\chi_i\rangle\langle\chi_i|$ is a product state. □

Fact 157. *Consider a state $\rho = |\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|$. Then its reduced density matrix $|\phi\rangle\langle\phi|$ is pure.*

So $h_{sep}(V)$ is equivalently an optimization over all states $\rho \in \mathcal{D}(X,Y)$ where $\mathrm{Tr}_Y[\rho]$ is pure. In other words, $h_{sep}(V)$ is exactly

$$\max_{\rho \in \mathcal{D}(X,Y)} \mathrm{Tr}[V\rho] \tag{5.58}$$

$$\text{such that} \quad rank(\mathrm{Tr}_Y[\rho]) = 1\,. \tag{5.59}$$

Thus, the optimal acceptance probability of a $\mathsf{QMA}(2)$ protocol can be written as an SDP with a rank-1 constraint (i.e. pure *subsystem*).

At the same time, any SDP with a rank-1 constraint can be converted to an SDP over separable states. For example, we may rewrite the above optimization as

$$\begin{aligned} &\max_{\rho \in \mathsf{SEP}(XY,Y')} \mathrm{Tr}[V\rho] \\ \text{such that} \quad &\mathrm{Tr}_X[\rho] \in Sym(Y,Y')\,. \end{aligned} \tag{5.60}$$

Eq. (5.60) specifies the SWAP test on (Y, Y') to accept with probability 1. This is because optimizing over separable states is equivalent to optimizing over product states, and for product states, the SWAP test passes with probability 1 iff the states are rank-1 (i.e. pure).

One can incorporate Eq. (5.60) into a new verifier V', which with probability p applies the SWAP test on (Y, Y'), and otherwise applies V. Then $\max_{\rho \in \mathsf{SEP}(XY,Y')} \mathrm{Tr}\left[V'\rho\right]$ is affine to $h_{sep}(V)$, with precision depending on p. This is essentially the content of the product test of Harrow and Montanaro [2013].

We also describe how $\mathsf{QMA}(2)$ essentially applies "purity testing" to extendable states. Extendable states can be thought of as "weakly" separable.

Definition 158 (k-extendable bipartite state)**.** *Fix $k > 1$. A state $\rho_{AB} \in \mathcal{D}(A, B)$ is k-extendable with respect to B if there exists a state $\sigma \in \mathcal{D}(A, B_1, \ldots, B_k)$ that is invariant on permutation of the subsystems $\{B_1, \ldots, B_k\}$ and satisfies $\mathrm{Tr}_{B_2,\ldots,B_k}[\sigma] = \rho_{AB}$.*

Fact 159. *Every separable state is k-extendable for all $k > 1$.*

Proof. Any separable state $\rho = \sum_i p_i \rho_{A,i} \otimes \rho_{B,i}$ can be extended to $\sigma = \sum_i p_i \rho_{A,i} \otimes \rho_{B,i}^{\otimes k}$. □

In fact, a state is separable *if and only if* it is k-extendable for all k Doherty et al. [2004].

For *pure* states, extendability is equivalent to separability. As a result, any optimization over pure separable states is equivalently an optimization over pure k-extendable states for any $k > 1$.

Fact 160. *Any* pure *bipartite state* $|\psi\rangle \in \mathcal{D}(A, B)$ *is 2-extendable with respect to* B *if and only if it is separable.*

Proof. The backwards direction follows from Fact 159, so we focus on the forwards direction. Consider the 2-extension $\sigma \in \mathcal{D}(A, B_1, B_2)$, which can be diagonalized as $\sigma = \sum_i p_i |\phi_i\rangle\langle\phi_i|$. Note that $\mathrm{Tr}_{B_2}[\sigma] = |\psi\rangle\langle\psi|$ is pure; thus, $\sigma = |\psi\rangle\langle\psi| \otimes \sum_i p_i |\chi_i\rangle\langle\chi_i|$ by Claim 156. But since σ is symmetric over (B_1, B_2), $\mathrm{Tr}_{B_1}[\sigma] = |\psi\rangle\langle\psi|$ is *also* pure. This implies that $\mathrm{Tr}_{B_1}[|\psi\rangle\langle\psi|] \otimes \sum_i p_i |\chi_i\rangle\langle\chi_i|$ must be pure, and thus $|\psi\rangle$ must be separable. □

By contrast, optimization over *mixed* extendable states is equivalent to that over arbitrary states:

Claim 161. *An optimization problem over arbitrary states can be converted to one over* poly(n)*-extendable states, and vice versa.*

Proof. Consider the optimization $\max_{\rho \in \mathcal{D}(A)} \mathrm{Tr}[V\rho]$. This is trivially equivalent to optimizing $\mathrm{Tr}[(V_A \otimes \mathbb{I}_B)\rho]$ over states $\rho \in \mathcal{D}(A, B)$ that are k-extendable with respect to B.

Fix any $k = \mathrm{poly}(n)$. Consider optimizing $\mathrm{Tr}[W\rho]$ over states $\rho \in \mathcal{D}(A, B)$ that are k-extendable with respect to B. This is equivalent to the following optimization:

$$\begin{aligned} &\max_{\rho \in \mathsf{SEP}(A, B_1, \ldots, B_k)} \mathrm{Tr}\big[(W_{A,B_1} \otimes \mathbb{I}_{B_2,\ldots,B_k})\rho\big] \\ \text{such that} \quad &\mathrm{Tr}_A[\rho] \in Sym(B_1, \ldots, B_k)\,. \end{aligned} \tag{5.61}$$

Moreover, one can implement the constraint by projecting ρ to the symmetric subspace, and accepting only if the projection succeeds. □

From this perspective, the power of unentanglement in $\mathsf{QMA}(2)$ is exactly the ability to test purity of extendable states. See Table 5.2 for a summary of these claims.

optimization over	arbitrary states	poly(n)-extendable states	separable states
mixed states allowed	QMA	QMA	QMA(2)
pure states only	QMA	QMA(2)	QMA(2)

Table 5.2: Categorizing quantum verification protocols by the purity and extendability of the quantum proof.

5.10 On gap amplification of $\mathsf{QMA}_{\mathsf{IS}}$ and $\mathsf{QMA}_{|\mathsf{IS}\rangle}$

Both $\mathsf{QMA}_{\mathsf{IS}}$ and $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ can trivially simulate any QMA protocol by using only part A of the quantum proof, i.e. ignoring the proof's separable subsystem. But at large enough completeness-soundness gap, both $\mathsf{QMA}_{\mathsf{IS}}$ and $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ can be simulated in QMA. This relies on the following claim:

Claim 162. *Consider any tripartite state $|\psi\rangle$ supported on registers A, B, C. Then there exists a state $|\chi\rangle \in \mathcal{W}_{|\mathsf{IS}\rangle}$ such that that $|\langle\chi|\psi\rangle|^2 \geq \frac{1}{\dim(C)}$.*

Proof. Schmidt decompose $|\psi\rangle$ on the $AB|C$ bipartition as $|\psi\rangle = \sum_i \alpha_i |i\rangle_{AB} |\phi_i\rangle_C$. By pigeonhole principle, there is an i^* such that $|\alpha_{i^*}|^2$ is at least 1 divided by the number of Schmidt vectors. The number of Schmidt vectors is at most $\min(\dim(B), \dim(C)) \leq \dim(C)$, so $|\alpha_{i^*}|^2 \geq \frac{1}{\dim(C)}$. Then the state $|\chi\rangle \stackrel{\text{def}}{=} |i^*\rangle_{AB} |\phi_{i^*}\rangle_C$ has at least $\frac{1}{\dim(C)}$ squared overlap with $|\psi\rangle$. □

Claim 163. *There exist constants c, s such that $\mathsf{QMA}^{c,s}_{\mathsf{IS}} \subseteq \mathsf{QMA}$ and $\mathsf{QMA}^{c,s}_{|\mathsf{IS}\rangle} \subseteq \mathsf{QMA}$.*

Proof. This proof follows Grilo et al. [2014], Bassirian et al. [2024b]. We focus on $\mathsf{QMA}_{\mathsf{IS}}$; the proof for $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ is identical. Consider any $\mathsf{QMA}^{c,s}_{\mathsf{IS}}$ protocol with verifier V. We directly run V in QMA.

By Claim 162, any quantum proof can be written as $|\psi\rangle = \sqrt{\ell}\,|g\rangle + (\sqrt{1-\ell})\,|b\rangle$, where $|g\rangle$ is a unit vector in $\mathcal{W}_{|\mathsf{IS}\rangle}$, $|b\rangle$ is some other unit vector, $0 < \ell \leq 1$ is a positive uniform constant, and $\langle b|g\rangle = 0$. Recall also that $\mathcal{W}_{|\text{comp}\rangle} \subseteq \mathcal{W}_{|\mathsf{IS}\rangle} \subseteq \mathcal{W}_{\mathsf{IS}}$.

- The completeness proof is unchanged, so the simulator accepts with probability at least c.
- In soundness, the simulator accepts with probability at most

$$\langle\psi| V |\psi\rangle = \ell \langle g| V |g\rangle + (1-\ell) \langle b| V |b\rangle + \sqrt{\ell(1-\ell)} \cdot (\langle b| V |g\rangle + \langle g| V |b\rangle) \tag{5.62}$$

$$\leq \ell \langle g| V |g\rangle + (1-\ell) \langle b| V |b\rangle + 2\sqrt{\ell(1-\ell)} \cdot \sqrt{\langle b| V |b\rangle \cdot \langle g| V |g\rangle} \tag{5.63}$$

$$\leq \ell \cdot s + (1-\ell) \cdot 1 + 2\sqrt{\ell(1-\ell)} \cdot \sqrt{s}\,. \tag{5.64}$$

The first inequality follows by Cauchy-Schwarz since V is PSD (i.e. $V = LL^\dagger$).

For any positive constant ℓ, there exist constant c, s where $1 - \ell + \ell s + 2\sqrt{s \cdot \ell(1-\ell)} < c$,[18] and so the protocol successfully distinguishes completeness and soundness. □

Remark 164. *As in Grilo et al. [2014], Bassirian et al. [2024b], Claim 163 naturally extends to* $\mathsf{QMA}(k)$*. For any constant* k*, there exist constants* c, s *such that* $\mathsf{QMA}_{\mathsf{IS}}^{c,s}(k) \subseteq \mathsf{QMA}(k)$*.*

We also prove that $\mathsf{QMA}_{|\mathsf{IS}\rangle} = \mathsf{NEXP}$ in Section 5.6. As a result, $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ does not have gap amplification unless $\mathsf{QMA} = \mathsf{NEXP}$. Note the same is true for the class QMA^+ Jeronimo and Wu [2023a], Bassirian et al. [2024b].

Nonetheless, $\mathsf{QMA}_{\mathsf{IS}}$ may exhibit gap amplification. The naive strategy of parallel repetition fails, since a partial measurement can destroy the guarantee of internal separability. But it is possible that other sophisticated methods (e.g. Marriott and Watrous [2005b], Harrow and Montanaro [2013]) could be used to prove such a statement.

18. For example, $c = 1 - \frac{\ell^2}{4}$ and $s = \frac{\ell}{4}$ is sufficient.

5.11 Other computational models where unentanglement is powerful

5.11.1 Towards a new understanding of QMA^+

Lemma 152 clarifies recent work on the complexity class $\mathsf{QMA}^+(k)$ Jeronimo and Wu [2023a], Bassirian et al. [2024b]. This class is defined as $\mathsf{QMA}^+(k) \stackrel{\text{def}}{=} \mathsf{QMA}_{\mathcal{W}_{|+\rangle}}(k)$, where $\mathcal{W}_{|+\rangle} \stackrel{\text{def}}{=} \{\sum_i a_i |i\rangle \mid (\forall i)\ a_i \geq 0\}$ is the set of quantum states with non-negative amplitudes in the computational basis.

Jeronimo and Wu found that $\mathsf{QMA}^+(2)$ at *large* enough completeness-soundness gap can be simulated by $\mathsf{QMA}(2)$, but can decide NEXP at *small* enough completeness-soundness gap Jeronimo and Wu [2023a]. They conjecture that $\mathsf{QMA}(2) = \mathsf{NEXP}$, since this would follow if $\mathsf{QMA}^+(2)$ exhibits gap amplification. However, Bassirian et al. [2024b] found a barrier to this approach, as gap amplification of QMA^+ would imply the unlikely $\mathsf{QMA} = \mathsf{NEXP}$. One interpretation of this result is that the power of QMA^+ comes from restricting quantum proofs to $\mathcal{W}_{|+\rangle}$. This suggests the following question:

Question 165. *Is there a set of quantum proofs $\mathcal{W}$ where gap amplification of both $\mathsf{QMA}_{\mathcal{W}}$ and $\mathsf{QMA}_{\mathcal{W}}(2)$ implies $\mathsf{QMA}(2) = \mathsf{NEXP}$, and no other complexity collapse?*

We remark that Question 165 is answered affirmatively by the set $\mathcal{W}_{\mathsf{IS}}$. Notably, a slight adjustment of QMA^+ also affirmatively answers Question 165. Expand the set of quantum proofs to *density matrices* with non-negative elements: $\mathcal{W}_+ \stackrel{\text{def}}{=} \{\rho \mid (\forall i,j)\ \rho_{ij} \geq 0\}$. As before, $\mathsf{QMA}_{\mathcal{W}_+}(k)$ at large enough completeness-soundness gap can be simulated by $\mathsf{QMA}(k)$. Moreover, $\mathsf{QMA}_{\mathcal{W}_+} \subseteq \mathsf{EXP}$, as the non-negative constraints can be added to an exponential-size semidefinite program. Finally, we can use Lemma 152 to show $\mathsf{QMA}_{\mathcal{W}_+}(2) = \mathsf{NEXP}$:

Claim 166. $\mathsf{QMA}_{\mathcal{W}_+}(2)$ *with a constant completeness-soundness gap can decide* NEXP.

Proof. First, note that any rank-1 density matrix with non-negative elements is in fact a non-negative amplitude state; thus, $\mathcal{W}_{|+\rangle} = \text{rank1}(\mathcal{W}_+)$. By Lemma 152, if $\mathcal{W}_+$ has the continuity condition, then $\mathsf{QMA}^+ = \mathsf{QMA}_{\mathcal{W}_{|+\rangle}} \subseteq \mathsf{QMA}_{\mathcal{W}_+}(2)$. The result follows since $\mathsf{QMA}^+ = \mathsf{NEXP}$ Bassirian et al. [2024b].

We now verify that $\mathcal{W}_+$ has the continuity condition. By a standard fact of linear algebra, the maximum eigenvector of any $\rho \in \mathcal{W}_+$ can be chosen to have non-negative amplitudes, i.e. in $\mathcal{W}_{|+\rangle} = \text{rank1}(\mathcal{W}_+)$. So then the continuity condition holds with $f(x) = x$. □

Lemma 152 provides an interpretation for the result of Bassirian et al. [2024b]. The complexity class $\mathsf{QMA}_{\mathcal{W}_+}$ restricts quantum states as *density matrices*. Here, we know that two unentangled proofs are more powerful than one: $\mathsf{QMA}_{\mathcal{W}_+} \neq \mathsf{QMA}_{\mathcal{W}_+}(2)$ unless $\mathsf{EXP} = \mathsf{NEXP}$. This directly stems from the SWAP test (i.e. purity test) in Lemma 152. Thus, the power of QMA^+ comes not from restricting quantum proofs to non-negative elements, but from *further* restricting $\mathcal{W}_+$ to *pure states*. Because of this, QMA^+ cannot have gap amplification (unless $\mathsf{QMA} = \mathsf{NEXP}$), but $\mathsf{QMA}_{\mathcal{W}_+}$ plausibly can.

5.11.2 Proper states

Proper states are states of the form $|\psi\rangle = \sum_{i=1}^{d} \pm\frac{1}{\sqrt{d}} |i\rangle$. It was shown in Aaronson et al. [2009], Beigi [2008] that optimizing over proper states of polynomial dimension is NP-hard with a constant gap. We show how this can be used to build a computational model where having one quantum proof is strictly weaker than having two quantum proofs, assuming $\mathsf{P} \neq \mathsf{NP}$.

Consider the set of real density matrices with equal values on the diagonal; i.e.

$$\mathcal{W}_{\text{proper}} \stackrel{\text{def}}{=} \{\rho \in \mathbb{R} \mid (\forall i, j)\ \rho_{ii} = \rho_{jj}\}. \tag{5.65}$$

Consider a variant of QMA allowing only $O(\log n)$ space, and the quantum proof restricted to states in $\mathcal{W}_{\text{proper}}$. Then this class can be decided in P, since there is an efficient semidefinite

program to calculate the maximum acceptance probability.

On the other hand, note that the pure states of $\mathcal{W}_{\text{proper}}$ are exactly the proper states of Aaronson et al. [2009], Beigi [2008]. By Lemma 152, the associated variant of $\mathsf{QMA}(2)$ can simulate NP, assuming that $\mathcal{W}_{\text{proper}}$ satisfies the continuity condition. We verify this now:

Claim 167. *$\mathcal{W}_{\text{proper}}$ has the continuity condition.*

Proof. Consider any element $\rho \in \mathcal{W}_{\text{proper}}$. Since ρ is real, we can always choose the maximum eigenvector to be real-valued; let it be $|\psi\rangle \stackrel{\text{def}}{=} \sum_i a_i |i\rangle$. Let the maximum eigenvalue is λ_1; then $\rho - \lambda_1 |\psi\rangle\langle\psi|$ is a PSD matrix, and so all diagonal values are non-negative. Thus,

$$\sum_i |1/n - \lambda_1 a_i^2| = \sum_i 1/n - \lambda_1 a_i^2 = \text{Tr}[\rho - \lambda_1 |\psi\rangle\langle\psi|] = 1 - \lambda_1 \,. \tag{5.66}$$

Let X be the random variable with value $n\lambda_1 a_i^2$. Then $\mathbb{E}[X] = \lambda_1 \sum_i a_i^2 = \lambda_1$. Since $0 \leq X \leq 1$, $Var(X) \leq \lambda_1(1-\lambda_1)$. By Chebyshev, $\mathbf{Pr}[X \leq \lambda_1 - k\sqrt{\lambda_1(1-\lambda_1)}] \leq \mathbf{Pr}[X \leq \lambda_1 - k\sigma] \leq \frac{1}{k^2}$.

We investigate the squared overlap of $|\psi\rangle$ with any proper state, which is maximized at value $\frac{1}{n}(\sum_i |a_i|)^2$. We lower-bound this quantity by ignoring the small terms from Chebyshev's inequality:

$$\frac{1}{n}\left(n \cdot \left(1 - \frac{1}{k^2}\right) \cdot \sqrt{\frac{\lambda_1 - k\sqrt{\lambda_1(1-\lambda_1)}}{n \cdot \lambda_1}}\right)^2 = \left(1 - \frac{1}{k^2}\right)^2 \cdot \left(1 - k\sqrt{\frac{1-\lambda_1}{\lambda_1}}\right) . \tag{5.67}$$

Let $c = \frac{1-\lambda_1}{\lambda_1}$. Then choose $k = c^{-1/6}$. The squared overlap value is then $(1 - c^{1/3})^3$. So, Lemma 152 holds with any polynomial dominating $g(x) = 1 - (1 - (\frac{x}{1-x})^{1/3})^3$. For example, we may use the Taylor polynomial of g around 0, or $f(x) = 4x^{1/4}$ for $x \in [0, 2^{-8}]$. □

In fact, assuming there is a PCP to scale up the NP-hardness result of Aaronson et al. [2009], Beigi [2008] to NEXP-hardness, we have $\mathsf{QMA}_{\mathcal{W}_{\text{proper}}} \subseteq \mathsf{EXP}$, yet $\mathsf{QMA}_{\mathcal{W}_{\text{proper}}}(2) =$

NEXP. Just as before, we can view the power of proper states as arising from restricting $\mathcal{W}_{\text{proper}}$ to pure states.

However, this class does not provide a good approach to prove QMA(2) = NEXP. Unlike QMA^+ and $\mathsf{QMA}_{\mathsf{IS}}$, it is not clear how to simulate $\mathsf{QMA}_{\mathcal{W}_{\text{proper}}}$ in QMA for any completeness-soundness gap. For example, any computational basis state in $\mathcal{D}(\mathbb{C}^d)$ has only $\frac{1}{d}$ squared overlap with every proper state.

5.12 QMA with a multiseparable quantum proof

We investigate the power of quantum proofs with another restriction of multipartite entanglement, called *multiseparability* by Thapliyal [1999]. These are tripartite states where after tracing *any* register, the reduced density matrix is separable. Label the registers A, B, C as before. Then

$$\mathcal{W}_{|\text{multisep}\rangle} = \{|\psi\rangle \mid \text{Tr}_A[|\psi\rangle\langle\psi|] \in \mathsf{SEP}(B,C), \text{Tr}_B[|\psi\rangle\langle\psi|] \in \mathsf{SEP}(A,C) \tag{5.68}$$

$$, \text{Tr}_C[|\psi\rangle\langle\psi|] \in \mathsf{SEP}(A,B)\}. \tag{5.69}$$

Let $\mathsf{QMA}_{|\text{multisep}\rangle} \stackrel{\text{def}}{=} \mathsf{QMA}_{\mathcal{W}_{|\text{multisep}\rangle}}$. Then this class can also decide NEXP at some gap:

Theorem 168. *There exist constants $1 > c > s > 0$ where $\mathsf{QMA}^{c,s}_{|\text{multisep}\rangle} = \mathsf{NEXP}$.*

Proof sketch. The proof is almost identical to the proof of Theorem 113.

We consider states on *five* registers $\mathbb{C}^R \otimes (\mathbb{C}^R \otimes \mathbb{C}^\kappa) \otimes (\mathbb{C}^R \otimes \mathbb{C}^\kappa)$, where $R = 2^{\text{poly}(n)}$ and κ is some constant. Our protocol is as follows:

- With probability p_1, measure all the registers in the computational basis, and make sure it is of the form $|v\rangle\,|v\rangle\,|c\rangle\,|v\rangle\,|c\rangle$ (i.e. the vertices match and the colors match).
- With probability p_2, run $\text{CNOT}_{1,2}$ (i.e. applied to the first and third register), then $\text{CNOT}_{1,4}$, then $\text{CNOT}_{3,5}$, and check $\mathsf{Density}$ on the first and third registers.
- Otherwise, run $\text{CNOT}_{1,2}$ and $\text{CNOT}_{1,4}$ and $\text{CNOT}_{3,5}$, and then run the constraint tests of Bassirian et al. [2024b] on the first and third registers.

Our quasirigid state before applying the CNOT gates is of the form

$$\sum_v a_v\,|v\rangle\,|v,\sigma(v)\rangle\,|v,\sigma(v)\rangle\,.$$

We remark that these states are in fact *multiseparable*. Since each basis element is uniquely

determined by the value in any one of the registers, tracing out any one register gives a separable state. The rest of the proof goes through virtually unchanged. □

Remarkably, pure states that are multiseparable can be characterized in another way.

Definition 169. *A tripartite state $|\psi\rangle$ has a* generalized *Schmidt decomposition if it can be written in the form*

$$|\psi\rangle = \sum_i \sqrt{p_i}\,|a_i\rangle\,|b_i\rangle\,|c_i\rangle\ , \tag{5.70}$$

where each set $\{|a_i\rangle\}, \{|b_i\rangle\}, \{|c_i\rangle\}$ is an orthonormal set.

Note that every tripartite state can be recursively Schmidt-decomposed, but this does not imply that every set $\{|a_i\rangle\}, \{|b_i\rangle\}, \{|c_i\rangle\}$ is simultaneously an orthonormal set.

Fact 170 (Thapliyal [1999])**.** *A state $|\psi\rangle$ has a generalized Schmidt decomposition iff $|\psi\rangle \in \mathcal{W}_{|\text{multisep}\rangle}$.*

Corollary 171. *There exists a constant $\alpha < 1$ above which it is* NP*-hard to α-approximately optimize linear functions of* poly(n)*-dimension states with generalized Schmidt decompositions.*

It is fruitful to further understand the connection between Corollary 171 and other NP-hardness results about testing properties of tensors.

One may also investigate the power of *mixed* multiseparable states $\mathcal{W}_{\text{multisep}}$. It is likely that Lemma 152 can be used to show $\mathsf{QMA}_{\text{multisep}}(2) = \mathsf{NEXP}$. The one copy-variant $\mathsf{QMA}_{\text{multisep}}$ can simulate $\mathsf{QMA}(2)$ by ignoring any one register of the proof, but its exact power remains unclear.

5.13 Computing probabilities of the $\mathsf{QMA}_{|\mathsf{IS}\rangle}$ protocol for NEXP

Here, we set the probability of running each test in the protocol of Section 5.6. This largely follows [Bassirian and Marwaha, 2024, Section 3.3] although with slightly different constants. We rely on the following claim, paraphrased from Bassirian et al. [2024b]:

Lemma 172 (Bassirian et al. [2024b])**.** *There exist constants* $0 < \xi \leq \mathrm{C}_{\mathrm{YES}} < 1$ *and a* QMA *protocol* ConstraintCheck *with completeness* $\mathrm{C}_{\mathrm{YES}}$ *and soundness* $\mathrm{C}_{\mathrm{YES}} - \xi$ *that decides* NEXP*,* assuming *the quantum proof is a* rigid *state.*

Lemma 172 implies the existence of absolute constants $\mathrm{C}_{\mathrm{YES}}, \xi, \kappa$. We choose *distance thresholds* $\nu_{\mathrm{low}}, \nu_{\mathrm{high}}$ to be small positive constants that satisfy the following conditions:

$$0 \leq \nu_{\mathrm{high}} \leq \frac{\nu_{\mathrm{low}}}{\kappa}\,, \quad \frac{\nu_{\mathrm{high}}}{\nu_{\mathrm{low}}} \leq \frac{\xi}{6(1-\mathrm{C}_{\mathrm{YES}})}\,, \quad \left(\kappa\nu_{\mathrm{low}} + (\kappa+1)\sqrt{(\kappa+1)\nu_{\mathrm{low}}}\right)^{1/2} \leq \frac{\xi}{2}\,. \tag{5.71}$$

This can be done by first making both constants equal and small enough to satisfy the last inequality, then reducing ν_{high} to satisfy the first two inequalities.

Now we choose the probabilities p_1, p_2, p_3. Let

$$p_1 = \frac{1}{Z} \qquad p_2 = \frac{(\nu_{\mathrm{low}} + \nu_{\mathrm{high}})}{\nu_{\mathrm{high}}^2 Z} \qquad p_3 = \frac{\nu_{\mathrm{low}}}{2(1-\mathrm{C}_{\mathrm{YES}})Z}\,, \tag{5.72}$$

where $Z \stackrel{\mathrm{def}}{=} 1 + \frac{(\nu_{\mathrm{low}}+\nu_{\mathrm{high}})}{\nu_{\mathrm{high}}^2} + \frac{\nu_{\mathrm{low}}}{2(1-\mathrm{C}_{\mathrm{YES}})}$, so that the probabilities sum to 1.

Given a quantum proof $|\psi\rangle$, the verifier accepts with probability

$$p_1 \cdot \mathbf{Pr}[\mathsf{Density} \text{ succeeds}] \tag{5.73}$$

$$+ p_2\, \mathbf{Pr}[\mathsf{MatchCheck} \text{ succeeds}] \tag{5.74}$$

$$+ p_3 \cdot \mathbf{Pr}[\mathsf{ConstraintCheck} \text{ succeeds}]\,. \tag{5.75}$$

In completeness, we consider the quantum proof guaranteed by Lemma 172. Since this state

is rigid, the verifier accepts with probability at least $\mathrm{P}_{\mathrm{YES}} \stackrel{\text{def}}{=} p_1 \cdot \frac{1}{\kappa} + p_2 \cdot 1 + p_3 \cdot \mathrm{C}_{\mathrm{YES}}$. For soundness, we analyze four separate cases:

1. In the first case, the proof doesn't test well enough on Density. If

$$\mathbf{Pr}[\textsf{Density succeeds}] = \frac{1}{\kappa} - d$$

for $d \geq \nu_{\mathrm{low}}$, then the verifier accepts with probability at most

$$\mathrm{P}_{\mathrm{NO}} \leq p_1 \cdot (\frac{1}{\kappa} - d) + p_2 \cdot 1 + p_3 \tag{5.76}$$

$$\leq \mathrm{P}_{\mathrm{YES}} - p_1 \cdot \nu_{\mathrm{low}} + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) = \mathrm{P}_{\mathrm{YES}} - \frac{\nu_{\mathrm{low}}}{2Z}\,. \tag{5.77}$$

2. In the second case, the proof tests too well on Density, and so cannot test well enough on MatchCheck. If $\mathbf{Pr}[\textsf{Density succeeds}] = \frac{1}{\kappa} + d$ for $d \geq \nu_{\mathrm{high}}$, then by Lemma 149 we can upper bound the success probability of MatchCheck as $\mathbf{Pr}[\textsf{MatchCheck succeeds}] \leq \frac{k+1-d^2}{k+1}$. Then

$$\mathrm{P}_{\mathrm{NO}} \leq p_1 \cdot \left(\frac{1}{\kappa} + d\right) + p_2 \cdot \frac{\kappa + 1 - d^2}{\kappa + 1} + p_3 \tag{5.78}$$

$$\leq \mathrm{P}_{\mathrm{YES}} + p_1 \cdot d - p_2 \cdot \frac{d^2}{\kappa + 1} + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}})\,. \tag{5.79}$$

The right-most expression depends on d through the term

$$p_1 \cdot d - p_2 \cdot \frac{d^2}{\kappa + 1} = p_1 \left(d - \frac{(\nu_{\mathrm{low}} + \nu_{\mathrm{high}})d^2}{(\kappa + 1)\nu_{\mathrm{high}}^2}\right).$$

This term is decreasing with d when $d > \frac{\kappa+1}{2} \cdot \frac{\nu_{\mathrm{high}}}{1+\nu_{\mathrm{low}}/\nu_{\mathrm{high}}} \geq \frac{\nu_{\mathrm{high}}}{2}$. So the expression

is maximized at $d = \nu_{\text{high}}$, and then

$$\mathrm{P}_{\mathrm{NO}} \leq \mathrm{P}_{\mathrm{YES}} + p_1 \left(\nu_{\text{high}} - \frac{(\nu_{\text{low}} + \nu_{\text{high}})\nu_{\text{high}}^2}{\nu_{\text{high}}^2} \right) + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) = \mathrm{P}_{\mathrm{YES}} - \frac{\nu_{\text{low}}}{2Z} \,. \tag{5.80}$$

3. In the third case, the proof doesn't test well enough on MatchCheck, and the acceptance probability of Density is not too high. If $\mathbf{Pr}[\text{Density succeeds}] = \frac{1}{\kappa} + d$ for $-\nu_{\text{low}} \leq d \leq \nu_{\text{high}}$ and $\mathbf{Pr}[\text{MatchCheck succeeds}] \leq 1 - \nu_{\text{low}}$, then

$$\mathrm{P}_{\mathrm{NO}} \leq p_1 \cdot \left(\frac{1}{\kappa} + d \right) + p_2 \cdot (1 - \nu_{\text{low}}) + p_3 \tag{5.81}$$

$$\leq \mathrm{P}_{\mathrm{YES}} + p_1 \cdot d - p_2 \cdot \nu_{\text{low}} + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) \,. \tag{5.82}$$

   By our choice of constants, $d \leq \nu_{\text{high}} \leq \nu_{\text{low}}$. Moreover, one can verify that $p_2 \geq 2 \cdot p_1$ whenever $\nu_{\text{high}} \leq 1$. So $p_1 \cdot d - p_2 \cdot \nu_{\text{low}} \leq -p_1 \cdot \nu_{\text{low}}$, and $\mathrm{P}_{\mathrm{NO}} \leq \mathrm{P}_{\mathrm{YES}} - p_1 \cdot \nu_{\text{low}} + p_3 \cdot (1 - \mathrm{C}_{\mathrm{YES}}) = \mathrm{P}_{\mathrm{YES}} - \frac{\nu_{\text{low}}}{2Z}$.

4. In the fourth case, the proof tests well on MatchCheck, and the acceptance probability of Density is not too low. So the proof must be close to *rigid*, and do worse on ConstraintCheck. If $\mathbf{Pr}[\text{Density succeeds}] = \frac{1}{\kappa} + d$ for $-\nu_{\text{low}} \leq d \leq \nu_{\text{high}}$ and $\mathbf{Pr}[\text{MatchCheck succeeds}] \geq 1 - \nu_{\text{low}}$, then by Claim 148, there is a rigid state $|\chi\rangle$ such that

$$|\langle \chi | \psi \rangle|^2 \geq 1 - \kappa d - (\kappa + 1)\sqrt{(\kappa + 1)\nu_{\text{low}}} \,. \tag{5.83}$$

   We use Fact 147 with Π equal to the accepting projector of ConstraintCheck; then we have

$$|\mathbf{Pr}[\text{ConstraintCheck succeeds}] - \mathbf{Pr}[\text{ConstraintCheck succeeds}]| \tag{5.84}$$

$$\leq \left(\kappa d + (\kappa+1)\sqrt{(\kappa+1)\nu_{\text{low}}}\right)^{1/2}. \tag{5.85}$$

Putting this all together,

$$\mathrm{P}_{\text{NO}} \leq p_1 \cdot \left(\frac{1}{\kappa} + d\right) + p_2 \cdot 1 + p_3 \cdot \left(\mathrm{C}_{\text{YES}} - \xi + \left(\kappa d + (\kappa+1)\sqrt{(\kappa+1)\nu_{\text{low}}}\right)^{1/2}\right) \tag{5.86}$$

$$\leq \mathrm{P}_{\text{YES}} + p_1 \cdot d + p_3 \cdot \left(-\xi + \left(\kappa d + (\kappa+1)\sqrt{(\kappa+1)\nu_{\text{low}}}\right)^{1/2}\right) \tag{5.87}$$

$$\leq \mathrm{P}_{\text{YES}} + p_1 \cdot \nu_{\text{high}} - p_3 \cdot \frac{\xi}{2} \tag{5.88}$$

$$= \mathrm{P}_{\text{YES}} + \frac{1}{Z}\left(\nu_{\text{high}} - \frac{\nu_{\text{low}} \cdot \xi}{4(1 - \mathrm{C}_{\text{YES}})}\right) \tag{5.89}$$

$$\leq \mathrm{P}_{\text{YES}} - \frac{\nu_{\text{high}}}{2Z}. \tag{5.90}$$

So, this protocol succeeds with completeness P_{YES} and soundness at most $\mathrm{P}_{\text{YES}} - \frac{\nu_{\text{high}}}{2Z}$.

5.14 Padding argument for the proof of $\mathsf{QMA}_{\mathsf{IS}}(2) = \mathsf{NEXP}$

Here, we verify that the protocol in Section 5.6 goes through even when the quantum proofs are from rank1($\mathcal{W}_{\mathsf{IS}}{}^*$); i.e. states in $\mathcal{W}_{|\mathsf{IS}\rangle}$ obeying the register size inequality in the proof of Claim 154. We modify the protocol as follows. We prepend a scratch qudit (using index variable z) to the first register to satisfy the register size inequality. In both Density and MatchCheck, we also measure the scratch qudit in the standard basis, and reject if it does not output 0_z. In the other test, we ignore the scratch qudit.

For this modified protocol, we prove that Lemma 146, Claim 148, and Lemma 149 hold with the same parameters. We may then choose the test probabilities exactly as in Section 5.13, as the analysis only depends on these three statements.

Lemma 173 (Padded version of Lemma 146)**.** *Suppose* MatchCheck *succeeds with probability* $1-\epsilon$*, and* $\rho_{B,C}$ *is separable. Then there is a state* $|0_z\rangle|\phi\rangle$ *that is* quasirigid *after applying* $CNOT_{1,3}CNOT_{2,4}$ *such that* $|\langle 0_z,\phi|\psi\rangle|^2 \geq 1-(\kappa+1)\epsilon$.

Proof. Label the computational basis elements of $|\psi\rangle$ as $a_{zvc,wd}|z\rangle|v\rangle|c,w\rangle|d\rangle$. Let $\sigma : [R] \to [\kappa]$ be a function that picks out the largest amplitude per vertex when the scratch qudit $z=0$, i.e. where $|a_{0_z,v\sigma(v),v\sigma(v)}|^2 \geq |a_{0_z,vd,vd}|^2$ for all $d \neq \sigma(v)$. For convenience, denote $\sigma_v \stackrel{\text{def}}{=} \sigma(v)$.

Consider the matrix elements of the reduced density matrix of $|\psi\rangle$ after tracing out the first register:

$$\langle c,v,d|\,\rho_{B,C}\,|e,w,f\rangle = \sum_z \sum_x a_{z,xc,vd} a^\dagger_{z,xe,wf}\,. \tag{5.91}$$

We use this expression to upper-bound the weight on colors not chosen by σ:

$$\sum_v \sum_{d;d\neq\sigma_v} |a_{0_z,vd,vd}|^2 \leq \sum_v \sum_{d;d\neq\sigma_v} |a_{0_z,v\sigma_v,v\sigma_v}| \cdot |a^\dagger_{0_z,vd,vd}| \tag{5.92}$$

$$= \sum_v \sum_{d;d\neq\sigma_v} |a_{0_z,v\sigma_v,v\sigma_v} \cdot a^\dagger_{0_z,vd,vd}|\,. \tag{5.93}$$

We break this upper bound into three parts, by rewriting $|a_{0_z,v\sigma_v,v\sigma_v} \cdot a^\dagger_{0_z,vd,vd}|$ as

$$| \langle \sigma_v, v, \sigma_v | \rho_{B,C} | d, v, d \rangle - \sum_{z \neq 0} \sum_x a_{z,x\sigma_v,v\sigma_v} \cdot a_{z,xd,vd} \tag{5.94}$$

$$- \sum_{x; x \neq v} a_{0_z,x\sigma_v,v\sigma_v} \cdot a^\dagger_{0_z,xd,vd} | \tag{5.95}$$

$$\leq | \langle \sigma_v, v, \sigma_v | \rho_{B,C} | d, v, d \rangle | + \sum_{z \neq 0} \sum_x |a_{z,x\sigma_v,v\sigma_v}| \cdot |a^\dagger_{z,xd,vd}| \tag{5.96}$$

$$+ \sum_{x; x \neq v} |a_{0_z,x\sigma_v,v\sigma_v}| \cdot |a^\dagger_{0_z,xd,vd}| \,. \tag{5.97}$$

1. For the first term, since $\rho_{B,C}$ is separable, we use one of the bounds proven in Lemma 145:

$$| \langle \sigma_v, v, \sigma_v | \rho_{B,C} | d, v, d \rangle | \leq \frac{1}{2} \cdot \big(\langle \sigma_v, v, d | \rho_{B,C} | \sigma_v, v, d \rangle \tag{5.98}$$

$$+ \langle d, v, \sigma_v | \rho_{B,C} | d, v, \sigma_v \rangle \big) \,. \tag{5.99}$$

So then

$$\sum_v \sum_{d; d \neq \sigma_v} | \langle \sigma_v, v, \sigma_v | \rho_{B,C} | d, v, d \rangle \leq \tag{5.100}$$

$$\frac{1}{2} \left(\sum_v \sum_{d; d \neq \sigma_v} \langle \sigma_v, v, d | \rho_{B,C} | \sigma_v, v, d \rangle + \langle d, v, \sigma_v | \rho_{B,C} | d, v, \sigma_v \rangle \right) \tag{5.101}$$

$$\leq \frac{1}{2} \left(\sum_v \sum_{d; d \neq \sigma_v} \sum_z \sum_x |a_{z,x\sigma_v,vd}|^2 + |a_{z,xd,v\sigma_v}|^2 \right) \,. \tag{5.102}$$

2. The second term appears because of the scratch qudit. This term is small because $z \neq 0$:

$$\sum_v \sum_{d; d \neq \sigma_v} \sum_{z \neq 0} \sum_x |a_{z,x\sigma_v,v\sigma_v}| \cdot |a^\dagger_{z,xd,vd}| \leq \sum_{z \neq 0} \sum_{v,x,d} |a_{z,xd,vd}|^2 \tag{5.103}$$

The sum of the first and second term are upper-bounded by

$$\sum_{(z,x,c)\neq(0,v,d)} |a_{z,xc,vd}|^2 = \epsilon.$$

3. For the third term, we use the AM-GM inequality:

$$|a_{0_z,x\sigma_v,v\sigma_v}| \cdot |a^\dagger_{0_z,xd,vd}| \leq \frac{1}{2} \cdot \left(|a_{0_z,x\sigma_v,v\sigma_v}|^2 + |a^\dagger_{0_z,xd,vd}|^2 \right) . \tag{5.104}$$

Taking $|a_{0_z,xd,vd}|^2$ over all three sums (v,d,x) is at most $\sum_{(z,x,c)\neq(0,v,d)} |a_{z,xc,vd}|^2 = 1-(1-\epsilon) = \epsilon$. Similarly, summing $|a_{0_z,x\sigma_v,v\sigma_v}|^2$ over sums (v,x) contributes at most ϵ, so at most $(\kappa-1)\epsilon$ over sums $(v, d\neq\sigma_v, x)$. In total, this term is at most $\frac{\kappa}{2}\epsilon$.

Combining each of the three terms, we have

$$\sum_v \sum_{d;d\neq\sigma_v} |a_{0_z,vd,vd}|^2 \leq \epsilon + \frac{\kappa}{2}\epsilon = \frac{\kappa+2}{2}\epsilon \leq \kappa\epsilon\,, \tag{5.105}$$

where the last inequality holds because $\kappa \geq 2$ (i.e. we need at least two colors).

Consider the state $|\phi\rangle = \frac{1}{\sqrt{\gamma}} \sum_v a_{0_z,v\sigma_v,v\sigma_v} |v\rangle |\sigma_v\rangle |v\rangle |\sigma_v\rangle$, where $\gamma = \sum_v |a_{0_z,v\sigma_v,v\sigma_v}|^2$ is chosen so that $\langle\phi|\phi\rangle = 1$. Note that $|\phi\rangle$ is quasirigid after applying $\mathrm{CNOT}_{1,3}$ and $\mathrm{CNOT}_{2,4}$. Moreover, $|\langle 0_z,\phi|\psi\rangle|^2 = \gamma$. Putting everything together,

$$\gamma = \sum_v |a_{0_z,v\sigma_v,v\sigma_v}|^2 = \mathbf{Pr}[\mathsf{MatchCheck}\text{ succeeds}] - \sum_v \sum_{d;d\neq\sigma_v} |a_{0_z,vd,vd}|^2 \geq 1-\epsilon-\kappa\epsilon\,.$$

□

Claim 174 (Padded version of Claim 148)**.** *Suppose* Density *and* MatchCheck *succeed on* $|\psi\rangle$ *with probability* $\frac{1}{\kappa} - d_{\mathsf{D}}$ *and* $1-d_{\mathsf{M}}$*, respectively. Then there is a state* $|0_z\rangle|\chi\rangle$ *that is rigid after applying* $CNOT_{1,3}CNOT_{2,4}$ *such that*

$$|\langle 0_z,\chi|\psi\rangle|^2 \geq 1-\kappa d_{\mathsf{D}} - (\kappa+1)\sqrt{(\kappa+1)d_{\mathsf{M}}}\,. \tag{5.106}$$

Proof. By Lemma 173, there exists a state $|\phi\rangle = \sum_{z,v} \alpha_{z,v} |z,v\rangle \left|c_{z,v}, v\right\rangle \left|c_{z,v}\right\rangle$ such that

$$|\langle\psi|\phi\rangle|^2 \geq 1 - (\kappa+1)d_{\mathsf{M}}\,. \tag{5.107}$$

By Fact 147, for any unit vector $|\mu\rangle$, we have $\left||\langle\mu|\psi\rangle|^2 - |\langle\mu|\phi\rangle|^2\right| \leq \sqrt{(\kappa+1)d_{\mathsf{M}}}$. We use this in two places. Let $\left|+'\right\rangle = |0_z\rangle \otimes \big(\mathrm{CNOT}_{1,3}\mathrm{CNOT}_{2,4} |+\rangle |0\rangle |0\rangle\big)$. First, since $|\left\langle+'\middle|\psi\right\rangle|^2 \geq \frac{1}{\kappa} - d_{\mathsf{D}}$, applying Fact 147 with $|\mu\rangle = \left|+'\right\rangle$ implies $|\left\langle+'\middle|\phi\right\rangle|^2 \geq \frac{1}{\kappa} - d_{\mathsf{D}} - \sqrt{(\kappa+1)d_{\mathsf{M}}}$. Now let $|\chi\rangle = \frac{1}{\sqrt{R}} \sum_v |v\rangle \left|c_{0_z,v}, v\right\rangle \left|c_{0_z,v}\right\rangle$ be such that $|0_z\rangle|\chi\rangle$ has exactly the shared computational basis elements of $\left|+'\right\rangle$ and $|\phi\rangle$. Then

$$|\langle 0_z,\chi|\phi\rangle|^2 = \left|\frac{1}{\sqrt{R}} \sum_{v\in[R]} \alpha_{0_z,v}\right|^2 = \kappa \left|\frac{1}{\sqrt{\kappa R}} \sum_{v\in[R]}^{R} \alpha_{0_z,v}\right|^2 \tag{5.108}$$

$$= \kappa |\left\langle+'\middle|\phi\right\rangle|^2 \geq 1 - \kappa\left(d_{\mathsf{D}} + \sqrt{(\kappa+1)d_{\mathsf{M}}}\right)\,. \tag{5.109}$$

Applying Fact 147 using $|\mu\rangle = |0_z\rangle|\chi\rangle$ then implies

$$|\langle 0_z,\chi|\psi\rangle|^2 \geq 1 - \kappa\left(d_{\mathsf{D}} + \sqrt{(\kappa+1)d_{\mathsf{M}}}\right) - \sqrt{(\kappa+1)d_{\mathsf{M}}} = 1 - \kappa d_{\mathsf{D}} - (\kappa+1)\sqrt{(\kappa+1)d_{\mathsf{M}}}\,.$$

□

Lemma 175 (Padded version of Lemma 149)**.** *Suppose* Density *and* MatchCheck *succeed on state* $|\psi\rangle$ *with probability* $w_{\mathsf{D}} \geq \frac{1}{\kappa}$ *and* w_{M}*, respectively. Then* $(w_{\mathsf{D}} - \frac{1}{\kappa})^2 + (\kappa+1)w_{\mathsf{M}} \leq \kappa+1$.

Proof. By Lemma 173, there exists a $|\phi\rangle = \sum_{z,v} \alpha_{z,v} |z,v\rangle \left|c_{z,v}, v\right\rangle \left|c_{z,v}\right\rangle$ such that

$$|\langle\psi|\phi\rangle|^2 \geq 1 - (\kappa+1)(1 - w_{\mathsf{M}})\,. \tag{5.110}$$

Using Fact 147, we see that $\left||\left\langle+'\middle|\psi\right\rangle|^2 - |\left\langle+'\middle|\phi\right\rangle|^2\right|$ is at most $\sqrt{(\kappa+1)(1-w_{\mathsf{M}})}$. Note

that by assumption, $|\langle +'|\psi\rangle|^2 = w_{\mathsf{D}} \geq \frac{1}{\kappa}$, and by Cauchy-Schwarz,

$$|\langle +'|\phi\rangle|^2 = \frac{1}{R\cdot\kappa}\left|\sum_{v\in[R]}\alpha_{0_z,v}\right|^2 \leq \frac{1}{\kappa}\sum_{v\in[R]}|\alpha_{0_z,v}|^2 \leq \frac{1}{\kappa}.$$

Combining these claims, we have

$$w_{\mathsf{D}} - \frac{1}{\kappa} \leq |\langle +'|\psi\rangle|^2 - |\langle +'|\phi\rangle|^2 \leq \sqrt{(\kappa+1)(1-w_{\mathsf{M}})}\,. \tag{5.111}$$

The lemma follows after rearranging the terms. □

REFERENCES

Scott Aaronson. Quantum lower bound for the collision problem, 2001. URL https://arxiv.org/abs/quant-ph/0111102.

Scott Aaronson. Quantum computing, postselection, and probabilistic polynomial-time, 2004. URL https://arxiv.org/abs/quant-ph/0412187.

Scott Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1(1):1–28, 2005a. ISSN 1557-2862. doi:10.4086/toc.2005.v001a001. URL https://arxiv.org/abs/quant-ph/0402095.

Scott Aaronson. Quantum computing and hidden variables. *Phys. Rev. A*, 71:032325, March 2005b. doi:10.1103/PhysRevA.71.032325. URL https://doi.org/10.1103/PhysRevA.71.032325.

Scott Aaronson. Pdqp/qpoly = all, 2018. URL https://arxiv.org/abs/1805.08577.

Scott Aaronson. Open problems related to quantum query complexity, 2021. URL https://arxiv.org/abs/2109.06917.

Scott Aaronson. Quantum miscellany — shtetl-optimized, 2023. URL https://scottaaronson.blog/?p=7516. [Online; accessed 10-February-2024].

Scott Aaronson and Andrew Drucker. A full characterization of quantum advice, 2013. URL https://arxiv.org/abs/1004.0377.

Scott Aaronson and Greg Kuperberg. Quantum versus classical proofs and advice. *Theory Comput.*, 3(1):129–157, 2007. URL https://arxiv.org/abs/quant-ph/0604056.

Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. The power of unentanglement, 2008. URL https://arxiv.org/abs/0804.0802.

Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. The power of unentanglement. *Theory of Computing*, 5(1):1–42, 2009. doi:10.4086/toc.2009.v005a001.

Scott Aaronson, Adam Bouland, Joseph Fitzsimons, and Mitchell Lee. The space "just above" bqp, 2014. URL https://arxiv.org/abs/1412.6507.

Scott Aaronson, Robin Kothari, William Kretschmer, and Justin Thaler. Quantum lower bounds for approximate counting via laurent polynomials, 2020. URL https://drops.dagstuhl.de/opus/volltexte/2020/12559/.

Scott Aaronson, Adam Bouland, Bill Fefferman, Soumik Ghosh, Umesh Vazirani, Chenyi Zhang, and Zixin Zhou. Quantum pseudoentanglement. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, 2024a. doi:10.4230/LIPIcs.ITCS.2024.2.

Scott Aaronson, Sabee Grewal, Vishnu Iyer, Simon C. Marshall, and Ronak Ramachandran. Pdqma = dqma = nexp: Qma with hidden variables and non-collapsing measurements, 2024b. URL TBA.

Dorit Aharonov and Tomer Naveh. Quantum np - a survey, 2002a. URL https://arxiv.org/abs/quant-ph/0210077.

Dorit Aharonov and Tomer Naveh. Quantum np - a survey, 2002b. URL https://arxiv.org/abs/quant-ph/0210077.

Dorit Aharonov and Oded Regev. A lattice problem in quantum np, 2003a. URL https://arxiv.org/abs/quant-ph/0307220.

Dorit Aharonov and Oded Regev. A lattice problem in quantum np, 2003b. URL https://arxiv.org/abs/quant-ph/0307220.

Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum circuits with mixed states, 1998. URL https://arxiv.org/abs/quant-ph/9806029.

Dorit Aharonov, Michael Ben-Or, Fernando G. S. L. Brandão, and Or Sattath. The pursuit of uniqueness: Extending valiant-vazirani theorem to the probabilistic and quantum settings. *Quantum*, 6:668, 2022. doi:10.22331/Q-2022-03-17-668. URL https://doi.org/10.22331/q-2022-03-17-668.

Seiseki Akibue, Go Kato, and Seiichiro Tani. On the hardness of conversion from entangled proof into separable one, 2024. URL https://arxiv.org/abs/2402.08981.

Noga Alon. On the edge-expansion of graphs. *Combinatorics, Probability and Computing*, 6 (2):145–152, 1997.

Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, pages 1–17, 2021. URL https://arxiv.org/abs/2003.11673.

Noga Alon and Shmuel Friedland. The maximum number of perfect matchings in graphs with a given degree sequence, 2008. URL https://arxiv.org/abs/0803.2578.

Andris Ambainis. Quantum lower bounds by quantum arguments. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing - STOC '00*. ACM Press, 2000. URL https://arxiv.org/abs/quant-ph/0002066.

Andris Ambainis. On physical problems that are slightly more difficult than qma. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 32–43. IEEE, 2014.

Andris Ambainis, Andrew M. Childs, and Yi-Kai Liu. Quantum property testing for bounded-degree graphs. *Lecture Notes in Computer Science*, page 365–376, 2011a. ISSN 1611-3349. doi:10.1007/978-3-642-22935-0_31. URL https://arxiv.org/abs/1012.3174.

Andris Ambainis, Andrew M. Childs, and Yi-Kai Liu. Quantum property testing for bounded-degree graphs. *Lecture Notes in Computer Science*, page 365–376, 2011b. ISSN 1611-3349. doi:10.1007/978-3-642-22935-0_31. URL http://dx.doi.org/10.1007/978-3-642-22935-0_31.

Atul Singh Arora, Alexandru Gheorghiu, and Uttam Singh. Oracle separations of hybrid quantum-classical circuits, 2022. URL https://arxiv.org/abs/2201.01904.

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

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

Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998b. URL https://doi.org/10.1145/1236457.1236459.

Sanjeev Arora, David Karger, and Marek Karpinski. Polynomial time approximation schemes for dense instances of np-hard problems. *Journal of Computer and System Sciences*, 58 (1):193–210, 1999. doi:10.1006/jcss.1998.1605. URL https://theory.cs.uni-bonn.de/ftp/reports/cs-reports/1994/85119-CS.pdf.

L. Babai, L. Fortnow, and C. Lund. Nondeterministic exponential time has two-prover interactive protocols. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, SFCS '90, page 16–25 vol.1, USA, 1990. IEEE Computer Society. ISBN 081862082X. doi:10.1109/FSCS.1990.89520. URL https://doi.org/10.1109/FSCS.1990.89520.

Boaz Barak, Jonathan A. Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, 2014. doi:10.1145/2591796.2591886.

Boaz Barak, Pravesh K. Kothari, and David Steurer. Quantum entanglement, sum of squares, and the log rank conjecture. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, 2017. doi:10.1145/3055399.3055488.

Roozbeh Bassirian and Kunal Marwaha. Superposition detection and QMA with non-collapsing measurements. *CoRR*, abs/2403.02532, 2024. doi:10.48550/ARXIV.2403.02532. URL https://doi.org/10.48550/arXiv.2403.02532.

Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. On the power of nonstandard quantum oracles. In *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*, 2023a. doi:10.4230/LIPIcs.TQC.2023.11. URL https://arxiv.org/abs/2212.00098.

Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. Quantum merlin-arthur and proofs without relative phase, 2023b. URL https://arxiv.org/abs/2306.13247.

Roozbeh Bassirian, Bill Fefferman, Itai Leigh, Kunal Marwaha, and Pei Wu. Quantum merlin-arthur with an internally separable proof. *CoRR*, abs/2410.19152, 2024a. doi:10.48550/ARXIV.2410.19152. URL https://doi.org/10.48550/arXiv.2410.19152.

Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. Quantum merlin-arthur and proofs without relative phase. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference, ITCS 2024, January 30 to February 2, 2024, Berkeley, CA, USA*, volume 287 of *LIPIcs*, pages 9:1–9:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024b. doi:10.4230/LIPICS.ITCS.2024.9. URL https://doi.org/10.4230/LIPIcs.ITCS.2024.9.

Richard Beigel. Bounded queries to sat and the boolean hierarchy. *Theoretical Computer Science*, 84, 1991.

Salman Beigi. Np vs qma_log(2), 2008.

Edward A Bender and E Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, 1978.

Charles H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 17:525–532, 1973. URL https://www.math.ucsd.edu/~sbuss/CourseWeb/Math268_2013W/Bennett_Reversibiity.pdf.

Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, October 1997. ISSN 1095-7111. doi:10.1137/s0097539796300933. URL https://arxiv.org/abs/quant-ph/9701001.

Charles H. Bennett, Sandu Popescu, Daniel Rohrlich, John A. Smolin, and Ashish V. Thapliyal. Exact and asymptotic measures of multipartite pure-state entanglement. *Phys. Rev. A*, 63, 12 2000. doi:10.1103/PhysRevA.63.012307.

Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556, 6 2004. doi:10.1145/1008731.1008733. URL https://people.orie.cornell.edu/miketodd/bertvemp.pdf.

Hugue Blier and Alain Tapp. A quantum characterization of np, 2010. URL https://arxiv.org/abs/0709.0738.

Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On the power of statistical zero knowledge, 2017. URL https://arxiv.org/abs/1609.02888.

Fernando G.S.L. Brandão and Matthias Christandl. Detection of multiparticle entanglement: Quantifying the search for symmetric extensions. *Physical Review Letters*, 10 2012. doi:10.1103/physrevlett.109.160502.

Fernando G.S.L. Brandão and Aram W. Harrow. Quantum de finetti theorems under local measurements with applications. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, 2013. doi:10.1145/2488608.2488718.

Fernando G.S.L. Brandão and Aram W. Harrow. Estimating operator norms using covering nets, 2015.

Fernando G.S.L. Brandão, Matthias Christandl, and Jon Yard. Faithful squashed entanglement. *Communications in Mathematical Physics*, 306(3):805–830, August 2011a. ISSN 1432-0916. doi:10.1007/s00220-011-1302-1. URL https://arxiv.org/abs/1010.1750.

Fernando G.S.L. Brandão, Matthias Christandl, and Jon Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, June 2011b. doi:10.1145/1993636.1993683.

Gilles Brassard and Tal Mor. Multi-particle entanglement via two-party entanglement. *Journal of Physics A: Mathematical and General*, 34(35), 8 2001. doi:10.1088/0305-4470/34/35/306. URL https://csaws.cs.technion.ac.il/~talmo/CV/my-papers/BM01-multipartyEnt.pdf.

Anne Broadbent and Eric Culf. Rigidity for monogamy-of-entanglement games, 2023. URL https://drops.dagstuhl.de/opus/volltexte/2023/17531/.

Anne Broadbent and Alex B. Grilo. Qma-hardness of consistency of local density matrices with applications to quantum zero-knowledge. *SIAM Journal on Computing*, 51(4), August 2022. doi:10.1137/21m140729x.

Andrei Z. Broder, Alan M. Frieze, Stephen Suen, and Eli Upfal. Optimal construction of edge-disjoint paths in random graphs. *SIAM Journal on Computing*, 28(2):541–573, 1998. doi:10.1137/S0097539795290805. URL https://doi.org/10.1137/S0097539795290805.

Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87:167902, September 2001. doi:10.1103/PhysRevLett.87.167902. URL https://arxiv.org/abs/quant-ph/0102001.

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

Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. Unknown quantum states: The quantum de finetti representation. *Journal of Mathematical Physics*, 43(9), September 2002. doi:10.1063/1.1494475.

Andre Chailloux and Or Sattath. The complexity of the separable hamiltonian problem. In *2012 IEEE 27th Conference on Computational Complexity*. IEEE, 6 2012. doi:10.1109/ccc.2012.42.

Jing Chen and Andrew Drucker. Short multi-prover quantum proofs for sat without entangled measurements. *arXiv preprint arXiv:1011.0716*, 2010. URL https://arxiv.org/abs/1011.0716.

Yuanyou Furui Cheng. Explicit estimate on primes between consecutive cubes, 2013. URL https://arxiv.org/abs/0810.2113.

Amanda G Chetwynd and Anthony JW Hilton. Regular graphs of high degree are 1-factorizable. *Proceedings of the London Mathematical Society*, 3(2):193–206, 1985.

Alessandro Chiesa and Michael A Forbes. Improved soundness for qma with multiple provers. *arXiv preprint arXiv:1108.2098*, 2011. URL https://arxiv.org/abs/1108.2098.

Andrew Childs. Lecture Notes on Quantum Algorithms, 2022. URL https://www.cs.umd.edu/~amchilds/qa/qa.pdf.

Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner. One-and-a-half quantum de finetti theorems. *Communications in Mathematical Physics*, 273(2):473–498, March 2007. doi:10.1007/s00220-007-0189-3.

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

Nicholas Cook, Larry Goldstein, and Tobias Johnson. Size biased couplings and the spectral gap for random regular graphs. *The Annals of Probability*, 46(1):72 – 125, 2018. doi:10.1214/17-AOP1180. URL https://doi.org/10.1214/17-AOP1180.

Steven A Cuccaro, Thomas G Draper, Samuel A Kutin, and David Petrie Moulton. A new quantum ripple-carry addition circuit. *arXiv preprint*, 2004. URL https://arxiv.org/abs/quant-ph/0410184.

David Cui, Arthur Mehta, Hamoon Mousavi, and Seyed Sajjad Nezhadi. A generalization of CHSH and the algebraic structure of optimal strategies. *Quantum*, 4:346, October 2020. doi:10.22331/q-2020-10-21-346. URL https://arxiv.org/abs/1911.01593.

Abhinav Deshpande, Alexey V. Gorshkov, and Bill Fefferman. The importance of the spectral gap in estimating ground-state energies, 2020. URL https://arxiv.org/abs/2007.11582.

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

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

Irit Dinur and Prahladh Harsha. Composition of low-error 2-query pcps using decodable pcps. *SIAM Journal on Computing*, 42(6):2452–2486, 2013. URL https://ieeexplore.ieee.org/document/5438607.

Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Distinguishing separable and entangled states. *Phys. Rev. Lett.*, 88, 4 2002. doi:10.1103/PhysRevLett.88.187904.

Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2), February 2004. doi:10.1103/physreva.69.022308.

W. Dür, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62, 11 2000. doi:10.1103/PhysRevA.62.062314.

Matthias Englbrecht, Tristan Kraft, Christoph Dittel, Andreas Buchleitner, Geza Giedke, and Barbara Kraus. Indistinguishability of identical bosons from a quantum information theory perspective. *Physical Review Letters*, 132(5), 1 2024. ISSN 1079-7114. doi:10.1103/physrevlett.132.050201. URL `http://dx.doi.org/10.1103/PhysRevLett.132.050201`.

Paul Erdos. On the central limit theorem for samples from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 4:49–61, 1959. URL `https://old.renyi.hu/~p_erdos/1959-13.pdf`.

Bill Fefferman and Shelby Kimmel. Quantum vs classical proofs and subset verification, 2018. URL `https://arxiv.org/abs/1510.06750`.

Bill Fefferman and Cedric Lin. Quantum merlin arthur with exponentially small gap, 2016a. URL `https://arxiv.org/abs/1601.01975`.

Bill Fefferman and Cedric Lin. Quantum merlin arthur with exponentially small gap, 2016b.

Bill Fefferman and Cedric Yen-Yu Lin. A complete characterization of unitary quantum space, 2016c. URL `https://arxiv.org/abs/1604.01384`.

Bill Fefferman and Zachary Remscrim. Eliminating intermediate measurements in space-bounded quantum computation. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 6 2021. doi:10.1145/3406325.3451051. URL `https://doi.org/10.1145/3406325.3451051`.

Stephen A. Fenner, Lance J. Fortnow, and Stuart A. Kurtz. Gap-definable counting classes. *Journal of Computer and System Sciences*, 48(1):116–148, 1994. ISSN 0022-0000. doi:https://doi.org/10.1016/S0022-0000(05)80024-8. URL `https://doi.org/10.1016/S0022-0000(05)80024-8`.

Asaf Ferber and Vishesh Jain. 1-factorizations of pseudorandom graphs. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 698–708. IEEE, 2018.

Lance Fortnow and John D. Rogers. Complexity limitations on quantum computation, 1998. URL `https://arxiv.org/abs/cs/9811023`.

Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *CoRR*, 2004. URL `https://arxiv.org/abs/cs/0405020`.

Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum mechanical states, 1998. URL https://arxiv.org/abs/quant-ph/9712042.

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

Pu Gao. Triangles and subgraph probabilities in random regular graphs, 2021.

Pu Gao and Yuval Ohapkin. Subgraph probability of random graphs with specified degrees and applications to chromatic number and connectivity, 2020.

Sevag Gharibian. Strong np-hardness of the quantum separability problem, 2009. URL https://arxiv.org/abs/0810.4507.

Sevag Gharibian. Guest column: The 7 faces of quantum NP. *SIGACT News*, 54(4):54–91, 2023a. doi:10.1145/3639528.3639535.

Sevag Gharibian. Guest column: The 7 faces of quantum np. *ACM SIGACT News*, 54(4): 54–91, December 2023b. ISSN 0163-5700. doi:10.1145/3639528.3639535. URL https://arxiv.org/abs/2310.18010.

Sevag Gharibian. The 7 faces of quantum np, 2023c. URL https://arxiv.org/abs/2310.18010.

Sevag Gharibian and Justin Yirka. The complexity of simulating local measurements on quantum systems. *Quantum*, 3:189, September 2019. ISSN 2521-327X. doi:10.22331/q-2019-09-30-189. URL https://arxiv.org/abs/1606.05626.

Sevag Gharibian, Jamie Sikora, and Sarvagya Upadhyay. Qma variants with polynomially many provers. *arXiv preprint 1108.0617*, 2011. doi:10.48550/ARXIV.1108.0617. URL https://arxiv.org/abs/1108.0617.

Sevag Gharibian, Jamie Sikora, and Sarvagya Upadhyay. Qma variants with polynomially many provers, 2012. URL https://arxiv.org/abs/1108.0617.

Sevag Gharibian, Miklos Santha, Jamie Sikora, Aarthi Sundaram, and Justin Yirka. Quantum generalizations of the polynomial hierarchy with applications to qma(2), 2018. URL http://drops.dagstuhl.de/opus/volltexte/2018/9640/.

Sevag Gharibian, Stephen Piddock, and Justin Yirka. Oracle complexity classes and local measurements on physical hamiltonians. In Christophe Paul and Markus Bläser, editors, *37th International Symposium on Theoretical Aspects of Computer Science, STACS 2020, March 10-13, 2020, Montpellier, France*, volume 154 of *LIPIcs*, pages 20:1–20:37. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPICS.STACS.2020.20. URL https://doi.org/10.4230/LIPIcs.STACS.2020.20.

Chris Godsil and Hanmeng Zhan. Discrete-time quantum walks and graph structures. *Journal of Combinatorial Theory, Series A*, 167:181–212, 2019. ISSN 0097-3165. doi:https://doi.org/10.1016/j.jcta.2019.05.003. URL https://arxiv.org/abs/1701.04474.

Dardo Goyeneche, Daniel Alsina, José I Latorre, Arnau Riera, and Karol Życzkowski. Absolutely maximally entangled states, combinatorial designs, and multiunitary matrices. *Physical Review A*, 92(3), 2015.

Alex B. Grilo, Iordanis Kerenidis, and Jamie Sikora. Qma with subset state witnesses, 2014. URL https://arxiv.org/abs/1410.2882.

Alex Bredariol Grilo, Iordanis Kerenidis, and Jamie Sikora. QMA with subset state witnesses. In *Mathematical Foundations of Computer Science 2015 - 40th International Symposium*. Springer, 2015. doi:10.1007/978-3-662-48054-0_14. URL https://arxiv.org/abs/1410.2882.

Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 1988. doi:10.1007/978-3-642-97881-4.

Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, 2003. doi:10.1145/780542.780545.

Leonid Gurvits and Howard Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physical Review A*, 66(6), December 2002. ISSN 1094-1622. doi:10.1103/physreva.66.062311.

Gus Gutoski, Patrick Hayden, Kevin Milner, and Mark M. Wilde. Quantum interactive proofs and the complexity of separability testing. *Theory of Computing*, 11(1):59–103, 2015. doi:10.4086/toc.2015.v011a003.

Otfried Gühne and Géza Tóth. Entanglement detection. *Physics Reports*, 474(1–6):1–75, April 2009. doi:10.1016/j.physrep.2009.02.004.

Aram W. Harrow. The church of the symmetric subspace, 2013a. URL https://arxiv.org/abs/1308.6595.

Aram W. Harrow. The church of the symmetric subspace, 2013b. URL https://arxiv.org/abs/1308.6595.

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

Aram W. Harrow and David J. Rosenbaum. Uselessness for an oracle model with internal randomness, 2011. URL https://arxiv.org/abs/1111.1462.

Aram W. Harrow, Anand Natarajan, and Xiaodi Wu. An improved semidefinite programming hierarchy for testing entanglement. *Communications in Mathematical Physics*, 352 (3):881–904, March 2017. doi:10.1007/s00220-017-2859-0.

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

Prahladh Harsha, Subhash Khot, Euiwoong Lee, and Devanathan Thiruvenkatachari. Improved 3lin hardness via linear label cover. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019. URL https://drops.dagstuhl.de/opus/volltexte/2019/11224/.

Patrick Hayden, Kevin Milner, and Mark M. Wilde. Two-message quantum interactive proofs and the quantum separability problem. In *2013 IEEE Conference on Computational Complexity*, June 2013. doi:10.1109/ccc.2013.24. URL 1211.6120.

Matthew Joel Hayward. Lower query bounds in the quantum oracle model. *Citeseer*, 2002. URL https://quantum-algorithms.herokuapp.com/thesis/thesis-ex.pdf.

Ryo Hiromasa, Akihiro Mizutani, Yuki Takeuchi, and Seiichiro Tani. Rewindable quantum computation and its equivalence to cloning and adaptive postselection, 2023. URL https://arxiv.org/abs/2206.05434.

Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006. URL https://www.ams.org/journals/bull/2006-43-04/S0273-0979-06-01126-8/.

Pawel Horodecki, Łukasz Rudnicki, and Karol Życzkowski. Multipartite entanglement, 2024.

Ryszard Horodecki, Pawel Horodecki, Michal Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 6 2009. doi:10.1103/RevModPhys.81.865.

Lawrence M. Ioannou. Computational complexity of the quantum separability problem. *Quantum Info. Comput.*, 7(4):335–370, May 2007.

Tsuyoshi Ito, Hirotada Kobayashi, and John Watrous. Quantum interactive proofs with weak error bounds, 2011. URL https://arxiv.org/abs/1012.4427.

Rahul Jain, Sarvagya Upadhyay, and John Watrous. Two-message quantum interactive proofs are in pspace, 2009. URL https://arxiv.org/abs/0905.1300.

Fernando G. Jeronimo and Pei Wu. Dimension independent disentanglers from unentanglement and applications, 2024. URL https://arxiv.org/abs/2402.15282.

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

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

Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Mip*=re. *Communications of the ACM*, 64(11):131–138, 2021a. URL https://arxiv.org/abs/2001.04383.

Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Mip*= re. *Communications of the ACM*, 64(11):131–138, 2021b. URL https://arxiv.org/abs/2001.04383.

Jonas Kamminga and Dorian Rudolph. On the complexity of pure-state consistency of local density matrices, 2024.

Elham Kashefi, Adrian Kent, Vlatko Vedral, and Konrad Banaszek. Comparison of quantum oracles. *Physical Review A*, 65(5), May 2002. doi:10.1103/physreva.65.050304. URL https://arxiv.org/abs/quant-ph/0109104.

Jonathan Katz and Yehuda Lindell. *Introduction to modern cryptography*. CRC press, 2020. URL http://www.cs.umd.edu/~jkatz/imc.html.

Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem, 2004. URL https://arxiv.org/abs/quant-ph/0406180.

Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 592–601. IEEE, 2018. URL https://eccc.weizmann.ac.il/report/2018/006/.

Jeong Han Kim and Nicholas C. Wormald. Random matchings which induce hamilton cycles and hamiltonian decompositions of random regular graphs. *Journal of Combinatorial Theory, Series B*, 81(1):20–44, 2001. ISSN 0095-8956. doi:https://doi.org/10.1006/jctb.2000.1991. URL https://www.sciencedirect.com/science/article/pii/S0095895600919919.

Jeong Han Kim, Benny Sudakov, and Van Vu. Small subgraphs of random regular graphs. *Discrete Mathematics*, 307(15):1961–1967, 2007. URL https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.9553&rep=rep1&type=pdf.

Gen Kimura. The bloch vector for n-level systems. *Physics Letters A*, 314(5–6):339–349, August 2003. doi:10.1016/s0375-9601(03)00941-1.

Yusuke Kinoshita. Qma(2) with postselection equals to nexp, 2018. URL https://arxiv.org/abs/1806.09732.

Alexei Yu Kitaev, Alexander Shen, and Mikhail N. Vyalyi. *Classical and quantum computation*. American Mathematical Soc., 2002. URL https://doi.org/10.1090/gsm/047.

M. Kleinmann, H. Kampermann, T. Meyer, and D. Bruß. Physical purification of quantum states. *Physical Review A*, 73(6), 6 2006. ISSN 1094-1622. doi:10.1103/physreva.73.062309. URL http://dx.doi.org/10.1103/PhysRevA.73.062309.

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

Hirotada Kobayashi, François Le Gall, and Harumichi Nishimura. Generalized quantum arthur–merlin games. *SIAM Journal on Computing*, 48(3):865–902, January 2019. ISSN 1095-7111. doi:10.1137/17m1160173. URL https://arxiv.org/abs/1312.4673.

Michael Krivelevich. Finding and using expanders in locally sparse graphs, 2017. URL https://arxiv.org/abs/1704.00465.

Michael Krivelevich. Expanders - how to find them, and what to find in them, 2019. URL https://arxiv.org/abs/1812.11562.

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

Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. *SIAM J. Comput.*, 22(6):1331–1348, 1993. doi:10.1137/0222080. URL https://doi.org/10.1137/0222080.

Cécilia Lancien and Andreas Winter. Distinguishing multi-partite states by local measurements. *Communications in Mathematical Physics*, 323(2):555–573, August 2013. doi:10.1007/s00220-013-1779-x.

Ke Li and Graeme Smith. Quantum de finetti theorem under fully-one-way adaptive measurements. *Physical Review Letters*, 114(16), April 2015. doi:10.1103/physrevlett.114.160503.

Yulong Li. A simple proof of preciseqma = pspace, 2022. URL https://arxiv.org/abs/2206.09230.

Yi-Kai Liu. Consistency of local density matrices is qma-complete. In *APPROX and RANDOM*, 2006.

Yi-Kai Liu. *The Complexity of the Consistency and N-representability Problems for Quantum States*. PhD thesis, University of California, San Diego, 2007.

Yi-Kai Liu, Matthias Christandl, and F. Verstraete. N-representability is qma-complete. *Physical Review Letters*, 98(11), March 2007. doi:10.1103/physrevlett.98.110503.

Alexander Lubotzky. Finite simple groups of lie type as expanders, 2009. URL https://arxiv.org/abs/0904.3411.

Alexander Lubotzky. Expander graphs in pure and applied mathematics, 2011. URL https://arxiv.org/abs/1105.2389.

Andrew Lutomirski. Component mixers and a hardness result for counterfeiting quantum money, 2011. URL https://arxiv.org/abs/1107.0321.

Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Ramanujan graphs and the solution of the kadison-singer problem, 2014. URL https://arxiv.org/abs/1408.4421.

Chris Marriott and John Watrous. Quantum arthur-merlin games, 2005a. URL https://arxiv.org/abs/cs/0506068.

Chris Marriott and John Watrous. Quantum arthur-merlin games. *Comput. Complex.*, 2005b. doi:10.1007/s00037-005-0194-x.

Dominic Mayers and Andrew Yao. Self testing quantum apparatus, 2004. URL https://arxiv.org/abs/quant-ph/0307205.

Matthew McKague. On the power quantum computation over real hilbert spaces. *International Journal of Quantum Information*, 11(01):1350001, February 2013. doi:10.1142/s0219749913500019. URL https://arxiv.org/abs/1109.0795.

Brendan D McKay and Nicholas C Wormald. Asymptotic enumeration by degree sequence of graphs with degreeso (n 1/2). *Combinatorica*, 11(4):369–382, 1991.

Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. *Explicit Near-Ramanujan Graphs of Every Degree*, page 510–523. Association for Computing Machinery, New York, NY, USA, 2020. ISBN 9781450369794. doi:10.1145/3357713.3384231. URL https://arxiv.org/abs/1909.06988.

Tomoyuki Morimae and Harumichi Nishimura. Merlinization of complexity classes above bqp, 2017. URL https://arxiv.org/abs/1704.01514.

Daniel Nagaj, Pawel Wocjan, and Yong Zhang. Fast amplification of qma, 2009. URL https://arxiv.org/abs/0904.1549.

Anand Natarajan and Chinmay Nirkhe. A classical oracle separation between qma and qcma, 2022. URL https://arxiv.org/abs/2210.15380.

Anand Natarajan and Tina Zhang. Quantum free games. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 1603–1616. ACM, 2023. doi:10.1145/3564246.3585208. URL https://doi.org/10.1145/3564246.3585208.

Miguel Navascués, Masaki Owari, and Martin B. Plenio. Power of symmetric extensions for entanglement detection. *Physical Review A*, 80(5), 11 2009. doi:10.1103/physreva.80.052306.

Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. ISBN 978-1-107-00217-3. doi:10.1017/CBO9780511976667.

Silvia Noschese, Lionello Pasquini, and Lothar Reichel. Tridiagonal toeplitz matrices: properties and novel applications. *Numerical linear algebra with applications*, 20(2):302–326, 2013. URL http://www.math.kent.edu/~reichel/publications/toep3.pdf.

Christos H. Papadimitriou and Mihalis Yannakakis. A note on succinct representations of graphs. *Information and Control*, 71(3):181–185, 1986. doi:https://doi.org/10.1016/S0019-9958(86)80009-2.

Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8): 1413–1415, August 1996. doi:10.1103/physrevlett.77.1413. URL quant-ph/9604005.

Attila Pereszlényi. Multi-prover quantum merlin-arthur proof systems with small gap, 2012a.

Attila Pereszlényi. Multi-prover quantum merlin-arthur proof systems with small gap, 2012b. URL https://arxiv.org/abs/1205.2761.

Julius Petersen. Die Theorie der regulären graphs. *Acta Mathematica*, 15:193 – 220, 1900. doi:10.1007/BF02392606. URL https://doi.org/10.1007/BF02392606.

John E. Prussing. The principal minor test for semidefinite matrices. *Journal of Guidance, Control, and Dynamics*, 9(1):121–122, 1986. doi:10.2514/3.20077. URL https://doi.org/10.2514/3.20077.

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

Renato Renner. Security of quantum key distribution, 2006. URL https://arxiv.org/abs/quant-ph/0512258.

Jason D. M. Rennie. Relating the trace and frobenius matrix norms, August 2005. URL http://people.csail.mit.edu/jrennie/writing/traceFrobenius.pdf.

Milajiguli Rexiti, Laleh Memarzadeh, and Stefano Mancini. Discrimination of dephasing channels. *Journal of Physics A: Mathematical and Theoretical*, 55(24):245301, may 2022. doi:10.1088/1751-8121/ac6d2c. URL https://dx.doi.org/10.1088/1751-8121/ac6d2c.

Adi Shamir. IP=PSPACE. *Journal of the ACM (JACM)*, 39(4):869–877, 1992. URL https://doi.org/10.1145/146585.146609.

Adrian She and Henry Yuen. Unitary property testing lower bounds by polynomials, 2022. URL https://arxiv.org/abs/2210.05885.

Yaoyun Shi and Xiaodi Wu. Epsilon-net method for optimizations over separable states. *Theoretical Computer Science*, 598:51–63, 2015. doi:10.1016/j.tcs.2015.03.031.

Ashish V. Thapliyal. Multipartite pure-state entanglement. *Physical Review A*, 59(5): 3336–3342, 5 1999. doi:10.1103/physreva.59.3336.

Jessica K. Thompson, Ojas Parekh, and Kunal Marwaha. An explicit vector algorithm for high-girth MaxCut, 2021. URL https://arxiv.org/abs/2108.12477.

Géza Tóth and Otfried Gühne. Entanglement and permutational symmetry. *Phys. Rev. Lett.*, 102, 5 2009. doi:10.1103/PhysRevLett.102.170503.

Luca Trevisan. Max cut and the smallest eigenvalue, 2008. URL `https://arxiv.org/abs/0806.1978`.

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

Armin Uhlmann. Transition probability (fidelity) and its relatives. *Foundations of Physics*, 41(3):288–298, March 2011.

Péter Vrana and Matthias Christandl. Asymptotic entanglement transformation between W and GHZ states. *Journal of Mathematical Physics*, 56(2), February 2015.

John Watrous. Quantum simulations of classical random walks and undirected graph connectivity, 1998. URL `https://arxiv.org/abs/cs/9812012`.

John Watrous. Quantum computational complexity, 2008. URL `https://arxiv.org/abs/0804.3401`.

Tzu-Chieh Wei, Michele Mosca, and Ashwin Nayak. Interacting boson problems can be qma hard. *Phys. Rev. Lett.*, 104, 1 2010. doi:10.1103/PhysRevLett.104.040501.

Nicholas C Wormald et al. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.

Jianwei Xu. Quantifying the phase of quantum states, 2023. URL `https://arxiv.org/abs/2304.09028`.

Xiao-Dong Yu, Timo Simnacher, Nikolai Wyderka, H. Chau Nguyen, and Otfried Gühne. A complete hierarchy for the pure state marginal problem in quantum mechanics. *Nature Communications*, 12(1), February 2021. doi:10.1038/s41467-020-20799-5.

Xiao-Dong Yu, Timo Simnacher, H. Chau Nguyen, and Otfried Gühne. Quantum-inspired hierarchy for rank-constrained optimization. *PRX Quantum*, 3(1), March 2022. doi:10.1103/prxquantum.3.010340.

D.B. Yudin and A.S. Nemirovskii. Informational complexity and efficient methods for the solution of convex extremal problems. *Ekonomika i Matematicheskie Metody*, 12:357–369, 1976. URL `https://books.google.com/books?id=0kEzAAAAIAAJ`. English translation: Matekon 13 (3) pp.25-45 (1977).

Karol Życzkowski and Ingemar Bengtsson. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2 edition, 2017. doi:10.1017/9781139207010. The arXiv link is based on a book chapter.