

A quantum oracle separation between $\text{QMA}(2)$ and QMA

John Bostanci* Sabee Grewal[†] Jonas Haferkamp[‡] Andrew Huang[§]
Yeongwoo Hwang^Δ Anand Natarajan^{||} Chinmay Nirkhe^{**}

Abstract

We find a quantum oracle relative to which $\text{QMA} \neq \text{QMA}(2)$. As a consequence, we resolve the no-disentanglers conjecture of Watrous: for every $\varepsilon + \delta < 1$, any (ε, δ) -disentangler requires input size exponential in the number of output qubits. Our proof combines the unitarily invariant polynomial method of She and Yuen (ITCS '23) with a new construction based on the symmetric and antisymmetric subspace projectors, reducing the QMA lower bound to the approximate degree of OR .

1 Introduction

Quantum information can fundamentally alter the structure and power of proof systems. Two major results illustrate this especially clearly. First, quantum interactive proofs can be parallelized to three messages, $\text{QIP} = \text{QIP}(3)$ [KW00], whereas an analogous parallelization for classical interactive proofs would imply $\text{PSPACE} = \text{AM}$, collapsing PSPACE to the second level of the polynomial hierarchy. Second, and perhaps even more dramatically, $\text{MIP} = \text{NEXP}$ while $\text{QMIP} = \text{MIP}^* = \text{RE}$, so allowing entanglement between provers increases the power of multiprover proofs all the way to undecidable problems [BFL91; JNVWY20].

In this work, we study a more mysterious quantum resource: *unentanglement*. The class $\text{QMA}(2)$ consists of problems that can be verified efficiently given two quantum proofs that are promised to be unentangled with one another [KMY01]. Whereas entanglement is usually viewed as a source of quantum power, $\text{QMA}(2)$ asks whether the *absence* of entanglement can improve verification for untrusted quantum systems.

Understanding the power of $\text{QMA}(2)$ is a major open problem in quantum complexity theory and is closely connected to fundamental questions about entanglement, separability, and optimization; see the recent survey of Jeronimo, Wu, and Leigh [JWL26]. Despite 25 years of study, at the level of a general complexity-class containment, essentially all that is known is $\text{QMA} \subseteq \text{QMA}(2) \subseteq \text{NEXP}$, and there is surprisingly little evidence to suggest whether $\text{QMA}(2)$ is “closer” in power to QMA or to NEXP .

A basic test of whether a promise of unentanglement improves verification is whether it can yield an advantage in the black-box setting. Can a $\text{QMA}(2)$ verifier solve a quantum oracle problem efficiently that no QMA verifier can? Such a separation would isolate unentanglement as a genuine source of computational power in proof systems: the two proof systems are given identical black-box access to the problem and differ only in the structure imposed on their proofs. This question has remained open for nearly two decades [AK07; ABDFS09], and Aaronson [Aar21] highlighted it as a fundamental question in quantum

*Simons Institute & UC Berkeley. cboostanc@berkeley.edu

[†]IBM Research, Columbia University. sabee@ibm.com

[‡]Ruhr-University Bochum. jonas.haferkamp@rub.de

[§]University of California, Berkeley. a_huang@berkeley.edu

^ΔHarvard University. yeongwoohwang@g.harvard.edu

^{||}MIT. anandn@mit.edu

^{**}University of Washington. nirkhe@cs.washington.edu

query complexity. We resolve one version¹ of this question by giving the first unitary-oracle separation between QMA and QMA(2).

Theorem 1.1 (Informal). *There exists a unitary oracle U such that*

$$\text{QMA}^U \neq \text{QMA}(2)^U.$$

In particular, we introduce a family of black-box problems that a QMA(2) verifier can solve using a single query and linear-size proofs, whereas any QMA verifier must use either exponentially many queries or an exponentially large proof.

The same QMA query lower bound also has consequences beyond the oracle setting. Specifically, it gives strong lower bounds on approximate *disentanglers*: quantum channels whose outputs are always close to separable states and whose images approximately contain every separable state. Watrous observed (as reported in [ABDFS09]) that in the query model a disentangler with only polynomially many input qubits would allow QMA(2) to be simulated by QMA. Our query separation thus implies that no such disentanglers exist, as Watrous conjectured.²

Theorem 1.2 (Informal). *For every fixed $\varepsilon, \delta \geq 0$ with $\varepsilon + \delta < 1$, an (ε, δ) -disentangler requires exponentially many input qubits in the number of output qubits.*

Before our work, unconditional lower bounds for general approximate disentanglers were known only in much more restricted parameter regimes. Aaronson et al. [ABDFS09] ruled out perfect $(0, 0)$ -disentanglers in every finite dimension, while Harrow, Natarajan, and Wu [HNW19] obtained a quasipolynomial lower bound on the input dimension when $\varepsilon + \delta$ is exponentially small in the number of output qubits. For constant error, Akibue, Kato, and Tani [AKT25] proved an exponential lower bound for the more restrictive class of *strong* disentanglers.

Acknowledgments. SG is supported by the Herman Goldstine Memorial Postdoctoral Fellowship. AN is supported by a Sloan Fellowship and NSF CAREER award 2339948. CN is supported by NSF CAREER award IIS2541127. YH is supported by NSF Award Nos. 2238836, 2430375, and an IBM PhD Fellowship. This work was initiated [in part] while the author(s) were visiting the Simons Institute for the Theory of Computing, supported by NSF QLCI Grant No. 2016245.

Tool and computational resource disclosure. A research manuscript should record not only its findings, but also how those findings were obtained. For AI-assisted research, we believe such disclosure gives reviewers relevant context for evaluating the work, provides transparency about the research process as the community develops norms around the use of these tools, and helps establish an accurate understanding of their capabilities and limitations.

Accordingly, we disclose that the proof idea underlying the main theorem was generated using ChatGPT 5.6 Sol. Our initial prompts directed the model to the work of She and Yuen [SY23], and, through subsequent prompting with minimal additional guidance, the model proposed the proof idea presented in this paper. The authors subsequently verified, simplified, and developed the argument presented here, and take full responsibility for its correctness, and exposition.

¹A classical oracle separation still eludes us!

²Aaronson et al. [ABDFS09] formulated a stronger quantitative conjecture, suggesting a lower bound of $\dim(\mathcal{H}) = 2^{\Omega(\dim(\mathcal{K}))}$ for a channel from $\mathcal{H}$ to $\mathcal{K} \otimes \mathcal{K}$. Recently, Jeronimo, Wu, and Xu [JWX26] refuted this quantitative conjecture by constructing, for every fixed $\varepsilon \in (0, 1)$, an $(\varepsilon, 0)$ -disentangler satisfying

$$\dim(\mathcal{H}) = \exp\left(O_\varepsilon\left(\sqrt{\dim(\mathcal{K}) \log \dim(\mathcal{K})}\right)\right).$$

2 Overview

Any computational power of $\text{QMA}(2)$ over QMA must be derived from its soundness guarantee: a $\text{QMA}(2)$ verifier may assume that even a dishonest proof is a product state. This suggests a natural unitary query problem for separating the two classes. Consider a black-box unitary that marks a subspace, with the promise that either the marked subspace contains a product state, or every product state has small overlap with it. A $\text{QMA}(2)$ verifier can solve such a problem trivially: any witness is guaranteed to be product, so it need only test whether the received state lies in the marked subspace. For a QMA verifier, by contrast, this test need not be sound, since the witness may be an arbitrary entangled state.

The main challenge in implementing this idea is to construct a family of marked subspaces for which every query-efficient verifier fails on some instance. Our approach combines the unitary polynomial method of She and Yuen [SY23]—an extension of the standard polynomial method of Beals et al. [BBCMD01]—with a highly symmetric family of marked subspaces. We view the identification of this family, together with the “miraculous” symmetry properties that drastically simplify the polynomials arising in She and Yuen’s method, as the main technical innovation of this work.

2.1 A starting point: guessing the witness

A useful starting point is to eliminate the witness in QMA by the standard technique of amplifying and guessing the witness. Suppose a query problem has a T -query QMA algorithm with completeness $2/3$ and soundness $1/3$, with a witness consisting of m qubits. By applying the in-place amplification procedure of Marriott-Watrous, this implies an $O(Tm)$ -query QMA algorithm with completeness $1 - 2^{-2m}$ and soundness 2^{-2m} , that still takes a witness of m qubits.

We can now eliminate the witness by running this verifier on a *uniformly random* witness. In the YES case, the acceptance probability is at least $2^{-m}(1 - 2^{-2m}) \geq 2^{-(m+1)}$, and in the NO case, it is at most 2^{-2m} . Although both acceptance-probability thresholds are exponentially close, they are still far apart in a *multiplicative* sense: their ratio is exponentially large. What we will show is that there is no witness-free quantum algorithm making polynomially many queries whose acceptance probability satisfies this multiplicative-error condition. (Technically, what we are doing is using the containment $\text{QMA} \subseteq \text{SBQP}$, and showing a lower bound against SBQP .) We will show this by applying a version of the polynomial method, as we will describe in the next section.

Before going on, it is worth noting that the technique of guessing the witness does *not* apply to $\text{QMA}(2)$. In fact, the first step of in-place amplification breaks down: it is known [ABDFS09] that any in-place amplification procedure for $\text{QMA}(2)$ would imply that $\text{QMA}(2) \subseteq \text{QMA}$, and this argument holds even in the oracle setting. Thus, our oracle separation implies that any in-place amplification for $\text{QMA}(2)$ in the plain model, if it exists, must be non-relativizing.

2.2 The polynomial method

The polynomial method of Beals et al. [BBCMD01] starts from the observation that the acceptance probability of a quantum algorithm making T queries to a classical oracle $(x_1, \dots, x_N)$ is a degree- $2T$ multivariate polynomial $p(x_1, \dots, x_N)$ in the oracle bits. Thus, to prove a query lower bound, it suffices to show that no low-degree polynomial has the acceptance behavior of the quantum algorithm. Since multivariate polynomials are hard to reason about, it is often crucial in these arguments to find symmetries in the problem that let us reduce the number of variables as much as possible—ideally to one.

The prototypical example of this method in action is the BQP query lower bound for the OR problem (also known as the unstructured search problem). The problem is to distinguish the NO case, in which every entry of the oracle is 0, from the YES case, in which at least one entry is 1. This problem is invariant

under permutations of the oracle bits: whether the input $(x_1, \dots, x_N)$ is a YES or NO instance depends only on its *Hamming weight* $x_1 + \dots + x_N$. To apply this symmetry, let us consider the *averaged* acceptance probability of the algorithm over all instances of Hamming weight r :

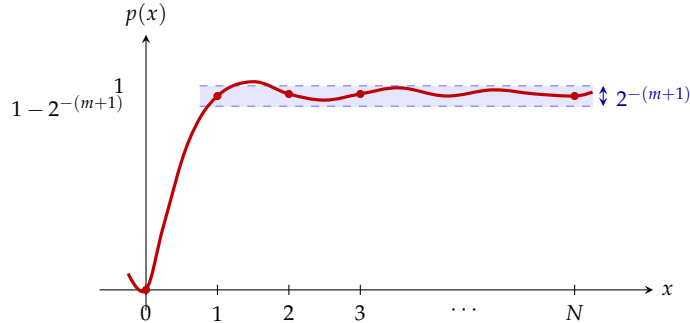
$$f(r) := \mathbb{E}_{(x_1, \dots, x_N) \in \{0,1\}^N: \sum_i x_i = r} p(x_1, \dots, x_N). \quad (1)$$

Basic results in invariant theory imply that f itself must be a degree- $2T$ univariate polynomial in r .

To obtain a query lower bound, observe that the success conditions require $f(0)$ to be close to 0, while $f(1), \dots, f(N)$ must all be close to 1. This means the graph of f must make a sharp “jump” near $r = 0$, followed by a long flat “runway.” A low-degree polynomial cannot have this shape, and standard bounds imply that in fact f must have degree at least $\Omega(\sqrt{N})$.

The same picture can be adapted to QMA using the amplify-and-guess argument mentioned above. Consider, for instance, the complement of OR, where $r = 0$ is the YES case and every $r \geq 1$ is the NO case.³ Let $f(r)$ now denote the averaged acceptance probability of the corresponding witness-guessing algorithm — i.e., the witness-free algorithm that accepts YES instances with probability $\geq 2^{-m}(1 - 2^{-2m})$ and accepts NO instances with probability $\leq 2^{-2m}$. Consider the rescaled polynomial $g := 2^m f$. Then $g(0) \geq 2/3$ (and in fact possibly as large as 2^m), while $g(r) \in [0, 2^{-m}]$ for $r = 1, 2, \dots, N$. Thus the long NO runway remains exponentially close to zero after rescaling, while the isolated YES point remains bounded away from zero. This again forces a large polynomial degree.

Figure 1: The polynomial degree lower bound we use. Consider a polynomial $p : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $p(0) = 0$ and $1 - 2^{-(m+1)} \leq p(j) \leq 1$ for every $j \in \{1, \dots, N\}$. Standard approximate-degree bounds for OR imply that such a polynomial has degree $\Omega(\min\{N, \sqrt{Nm}\})$.



The idea of applying the polynomial method template to *unitary* query lower bounds, and to the QMA vs QMA(2) problem in particular, was introduced by She and Yuen [SY23]. For ease of exposition, we explain their method in the case where the unitary oracle is Hermitian, which is the only case we will need. She and Yuen started by observing that for any quantum algorithm making T queries to a *unitary* oracle U , its acceptance probability is given by a polynomial $p(U)$ of total degree $2T$ in the entries of U . They then identified a symmetry that is natural in the QMA(2) setting: suppose the oracle U acts on a *bipartite* Hilbert space $\mathbb{C}^d \otimes \mathbb{C}^d$. Then it is natural to ask the QMA(2) algorithm to decide properties of U that are *invariant under conjugation by local unitaries*:

$$U \text{ is a YES instance} \iff (V \otimes W)U(V^\dagger \otimes W^\dagger) \text{ is a YES instance.}$$

³Note that the argument we are about to discuss does not apply to OR, which is actually in QMA! The reason is that, after rescaling, the “runway” for the OR problem need not remain flat: the rescaled polynomial is allowed to take any value from around 1 up to 2^m on these points.

For instance, suppose U is a reflection that marks a subspace, and consider the problem of distinguishing whether the marked subspace contains a product state or has small overlap with every product state. This property has the desired local-unitary symmetry and is also easily decidable by a QMA(2) proof system: simply ask for a product state in the marked subspace as the witness. This is the *Entangled Subspace problem* studied by She and Yuen.

She and Yuen identified an analog of Eq. (1) for local unitary invariance. For any oracle U , let $g(U)$ be the distribution over oracles obtained by conjugating U by independent Haar-random local unitaries. Then the acceptance probability of a quantum algorithm, averaged over oracles drawn from $g(U)$, can be written as a linear combination of certain low-degree polynomials in U :

$$\mathbb{E}_{\tilde{U} \sim g(U)} p(\tilde{U}) = \sum_{t, \sigma, \tau} \alpha_{t, \sigma, \tau} \text{tr}[(R_\sigma \otimes R_\tau) U^{\otimes t}]. \quad (2)$$

Here, the argument of each trace is an operator on t copies of $\mathbb{C}^d \otimes \mathbb{C}^d$, viewed as a tensor product of t copies of the “left” system and t copies of the “right” system. The operators R_σ and R_τ are permutations of the t left and right registers, respectively.

While Eq. (2) is a significant improvement over working with all the entries of U , it is still much less clean than the analogous result for the OR problem in Eq. (1), which yields a univariate polynomial. The main technical contribution of this work is to identify a highly symmetric family of marked subspaces for which all of the trace expressions in Eq. (2) are captured by low-degree univariate polynomials in a single parameter, the dimension d of the subspace. Even more importantly, the same polynomial will simultaneously describe both the YES and NO instances of our promise problem. This will make our problem formally analogous to the OR problem, almost immediately yielding a QMA query lower bound.

2.3 Our oracle problem and the QMA lower bound

Our oracle family is a restriction of the Entangled Subspace problem of She and Yuen to a particular set of subspaces. We take the oracle U to be a reflection $I - 2\Pi$ about a subspace defined by a projector Π , with YES instances corresponding to rank-one product subspaces and NO instances corresponding to subspaces with low overlap with every product state. She and Yuen suggested that a family of hard-to-distinguish instances could be constructed by taking random subspaces as NO instances and planting a product state to obtain YES instances. Our innovation is to consider a more structured family of instances that lets us make better use of local-unitary invariance.

Let $D = 2^{\text{poly}(n)}$ be the local ambient dimension, so that the marked subspace lies in $\mathbb{C}^D \otimes \mathbb{C}^D$. We will construct $\Theta(\sqrt{D})$ different NO instances, together with a family of YES instances, with the following two properties:

- Every YES instance is a rank-one product projector, whereas every NO instance has overlap at most $1/3$ with every product state.
- For every local-unitary invariant polynomial arising from Eq. (2), its values on all of these YES and NO instances are simultaneously captured by a single low-degree univariate polynomial.

The construction underlying these two properties will be described in the next subsection.

These two facts, together with Eq. (2), mean that we can apply the polynomial method. Suppose there is a T -query QMA verifier using an m -qubit witness. By the amplify-and-guess argument above, and then averaging over local changes of basis, we obtain a local-unitary invariant polynomial of degree $O(Tm)$ —where the factor m comes from in-place amplification—whose value is noticeably large on YES instances and exponentially smaller on NO instances.

By the second property above, this invariant polynomial becomes a univariate polynomial. After a simple change of variables, normalization, and flipping the polynomial, we obtain a polynomial p such that $p(0) = 0$, while $1 - 2^{-(m+1)} \leq p(j) \leq 1$ at $\Theta(\sqrt{D})$ consecutive positive integer values of j . This is exactly the shape shown in Fig. 1. The polynomial degree lower bound therefore gives

$$Tm \geq \Omega\left(\min\{\sqrt{D}, D^{1/4}\sqrt{m}\}\right) \implies T^2m = \Omega(\sqrt{D}).$$

Thus, when D is exponential in the input length, any QMA verifier must use either superpolynomially many queries or a superpolynomial-size witness. By a standard diagonalization argument, we conclude that there exists a unitary oracle U such that $\text{QMA}^U \neq \text{QMA}(2)^U$.

2.4 Negative dimensions

In constructing our families of YES and NO instances, the main ingredients we use are projectors onto the symmetric and antisymmetric subspaces of bipartite Hilbert spaces of the form $\mathbb{C}^d \otimes \mathbb{C}^d$, where we will refer to the two tensor factors as the “left” and “right” systems. Since these two subspaces are the eigenspaces of the SWAP operators with eigenvalue $+1$ and -1 , respectively, in this section we use the notation $P_d^\pm$ to denote the symmetric and antisymmetric subspace projectors. These are defined as

$$P_d^\pm = \frac{I \pm \text{SWAP}}{2}.$$

These fulfill the required properties of YES and NO instances; the symmetric subspace contains many product states, whereas any product state has bounded overlap with the antisymmetric subspace.

By the symmetrization argument sketched in the previous section, the acceptance probability of a T -query quantum algorithm, averaged over the local changes of basis, is a linear combination of terms of the form

$$\text{tr}[(R_\sigma \otimes R_\tau)(P_d^\pm)^{\otimes t}],$$

where t ranges from 0 to $2T$, and R_σ and R_τ are separate permutations of the “left” and “right” systems. To carry out the polynomial method analysis from the previous section, we would like following “dream identity” to hold: for every σ, τ, t ,

$$\text{tr}[(R_\sigma \otimes R_\tau)(P_d^\pm)^{\otimes t}] \stackrel{?}{=} f_{\sigma, \tau, t}(\pm d), \quad (3)$$

where $f_{\sigma, \tau, t}$ is a polynomial of degree t , depending on σ, τ , and t . Crucially, we want the symmetric and antisymmetric projectors to be captured by the *same* polynomial, evaluated at positive and negative values of the argument, respectively. Or, more pithily, we want the antisymmetric projector to behave like a symmetric projector in a “negative dimensional space.” It turns out the dream identity is not always true, but it will hold for a class of σ, τ that will turn out to be sufficient.

To get a very basic intuition for why the dream identity might be expected to hold, we can look at the simplest such identity, the trace: $\text{tr}[P_d^\pm]$. To compute the trace, we expand the projector in terms of the identity and SWAP operators:

$$\text{tr}[P_d^\pm] = \frac{1}{2}(\text{tr}[I] \pm \text{tr}[\text{SWAP}]) = \frac{1}{2}(d^2 \pm d) = \frac{1}{2}((\pm d)^2 + (\pm d)).$$

We see that this expression has exactly the desired form: it is a polynomial that, when evaluated at $+d$, yields the value for P_d^+ , and when evaluated at $-d$, yields the value for P_d^- .

Essentially the same strategy will serve us in the general case: simply expand the tensor product of

projectors in terms of identity and swap operators, to get

$$\text{tr}[(R_\sigma \otimes R_\tau)(P_d^\pm)^{\otimes t}] = \frac{1}{2^t} \sum_{S \subseteq [t]} (\pm 1)^{|S|} \text{Tr} \left[\underbrace{(R_\sigma \otimes R_\tau) \text{SWAP}^S}_{=\text{some permutation } R_\xi \text{ of the subsystems}} \right],$$

where we have introduced the shorthand notation SWAP^S to refer to the operator that acts as a tensor product of SWAPs on the pairs of left and right systems indexed by elements of the set S . Each of the individual traces in this sum is a trace of some permutation R_ξ acting on the $2t$ registers of the quantum system. The value of such a trace is determined entirely by the *cycle decomposition* of ξ , and the dimension d : in fact,

$$\text{tr}[R_\xi] = d^{C(\xi)},$$

where $C(\xi)$ is the number of cycles in the cycle decomposition of ξ . A very intuitive way to see this is with Penrose's diagrammatic notation for tensors, as illustrated in Figure 2. In the diagram, every cycle of π is converted into a closed loop upon taking the trace, which yields a factor of the dimension. Plugging this back into the above, we get

$$\text{tr}[(R_\sigma \otimes R_\tau)(P_d^\pm)^{\otimes t}] = \frac{1}{2^t} \sum_{S \subseteq [t]} (\pm 1)^{|S|} d^{C(\xi)},$$

where we have notationally suppressed the dependence of ξ on S, σ, τ to avoid clutter.

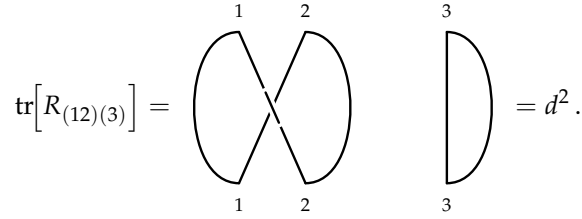


Figure 2: A tensor network diagram illustrating $\text{tr}[R_{(12)(3)}] = d^2$.

Now, at this point we would have shown the “dream identity” if the sign factor $(\pm 1)^{|S|}$ were equal to $(\pm 1)^{C(\xi)}$. This is unfortunately not true in general, but *is* true when σ, τ are both *even* permutations. Under this assumption, we have

$$\text{tr}[(R_\sigma \otimes R_\tau)(P_d^\pm)^{\otimes t}] = \frac{1}{2^t} \sum_{S \subseteq [t]} (\pm 1)^{|S|} d^{C(\xi)} = \frac{1}{2^t} \sum_{S \subseteq [t]} (\pm 1)^{C(\xi)} d^{C(\xi)} = \frac{1}{2^t} \sum_{S \subseteq [t]} (\pm d)^{C(\xi)},$$

as desired.

It turns out one final trick lets us entirely remove the restriction on σ and τ and derive the dream identity for a different set of projectors: define $\Pi_d^\pm := (P_d^\pm)^{\otimes 2}$ to be projectors on two copies of $\mathbb{C}^d \otimes \mathbb{C}^d$, viewed as $\mathbb{C}^{d^2} \otimes \mathbb{C}^{d^2}$ by “merging” the two left-hand systems and the two right-hand systems. Now, any permutation of t merged d^2 -dimensional registers automatically becomes an *even* permutation of the $2t$ underlying d -dimensional registers and

$$\text{tr}[(R_\sigma \otimes R_\tau)(\Pi_d^\pm)^{\otimes t}] = \text{tr} \left[\underbrace{(R_{\sigma'} \otimes R_{\tau'})}_{\text{even permutations}} (P_d^\pm)^{\otimes 2t} \right] = f_{\sigma', \tau', 2t}(\pm d).$$

This doubled construction already captures the key intuition we need: by grouping registers into pairs, the relevant permutations become even, and the symmetric and antisymmetric projectors can be described by the same polynomial after formally negating the dimension. It would therefore already suffice to derive the qualitative separation $\text{QMA}^U \subsetneq \text{QMA}(2)^U$. However, in the main body of the paper, we use a slightly different construction that gives stronger quantitative bounds and makes the final promise problem cleaner. Specifically, in Section 4 we will instead take $\Pi_d^- = \Pi_{\text{anti}}^{d,A}$, the projector onto the 4-copy antisymmetric subspace, viewed as a bipartite operator by grouping copies 1,2 and 3,4. This solves the sign issue in essentially the same way as the doubled projectors, since every permutation of the merged registers becomes an even permutation on the underlying systems. Moreover, this choice is “stronger” in the sense that $\Pi_{\text{anti}}^{d,A} \preceq P_d^- \otimes P_d^-$, yielding an improved soundness guarantee and, ultimately, better parameters for our disentangler result.

For either construction, each projector $\Pi_d^\pm$ acts on $\mathbb{C}^{d^2} \otimes \mathbb{C}^{d^2}$, so we can embed it into a common ambient space $\mathbb{C}^D \otimes \mathbb{C}^D$ whenever $d^2 \leq D$. For the construction used in the main body, the formal point $d = -1$ corresponds to the symmetric projector Π_1^+ , which is a rank-one product projector and hence gives our YES instances. For the NO instances, we take the antisymmetric projector at $d = 4, 9, 14, \dots$ up to $\sqrt{D}$, giving $\Theta(\sqrt{D})$ different NO instances. Under the simple change of variables $d = 5j - 1$, the formal YES point $d = -1$ maps to $j = 0$, while the NO instances map to the consecutive positive integers $j = 1, 2, \dots$. These are exactly the points forming the long “runway” in Fig. 1.

2.5 Disentanglers

Our work has immediate consequences for the existence of *approximate disentanglers*.

Definition 2.1 (Approximate disentangler). Let $\Lambda : \mathcal{B}(\mathbb{C}^{2^m}) \rightarrow \mathcal{B}(\mathbb{C}^D \otimes \mathbb{C}^D)$ be a quantum channel. We say that Λ is an (ε, δ) -disentangler if:

1. for every input state ρ , there is a separable state σ such that $\frac{1}{2} \|\Lambda(\rho) - \sigma\|_1 \leq \varepsilon$ and,
2. for every separable state σ , there is an input state ρ such that $\frac{1}{2} \|\Lambda(\rho) - \sigma\|_1 \leq \delta$.

Indeed, suppose we had an (ε, δ) -disentangler Λ , with $m = m(D)$. Then, in the query model⁴, any $\text{QMA}(2)$ protocol whose completeness-soundness gap is larger than $\varepsilon + \delta$ can be simulated by a QMA protocol. To see this, let us take a $\text{QMA}(2)$ protocol with completeness c and soundness s , where the verifier expects two unentangled proofs of w qubits each. A QMA verifier for the same problem is to apply the disentangler Λ on a witness of $m(2^w)$ qubits, then simulate the $\text{QMA}(2)$ verifier on the output of the disentangler. This yields a QMA verifier with completeness at least $c - \varepsilon$ (the prover sends a state whose image, under the disentangler, is ε -close to the optimal QMA witness), and soundness at most $s + \delta$ (any state the prover sends gets mapped to a state that is δ -close to separable). Moreover, this QMA verifier makes the same number of queries as the original $\text{QMA}(2)$ verifier.

By taking the contrapositive of this simulation argument, our separation between QMA and $\text{QMA}(2)$ will rule out the existence of disentanglers where $m(D)$ is small. To get quantitative results, we start with a more precise version of our separation: we show that for our version of the Entangled Subspace problem, for every D' that is a perfect square, the following statements hold.

1. A single-query $\text{QMA}(2)$ algorithm solves the problem on oracle instances acting on the space $\mathbb{C}^{D'} \otimes \mathbb{C}^{D'}$, with a witness consisting a product state in this space, completeness $c = 1$, and constant soundness $s < 1$.

⁴Since we are working in the query model, we ignore the computational efficiency of the disentangler: all that matters is that it takes no queries to implement it.

2. Any QMA algorithm with m -qubit witnesses on these instances requires $\Omega(\sqrt{(D')^{1/2}/m})$ queries.

Putting these together, suppose we have an (ε, δ) -disentangler with $\varepsilon + \delta < c - s$. Then the simulation yields a QMA algorithm making 1 query and taking a witness of size $m(D')$. The QMA lower bound therefore implies that

$$m(D') \geq \Omega((D')^{1/2}).$$

In the regime $\varepsilon + \delta \in [0, c - s)$, this lower bound matches the upper bound obtained by [JWX26], up to a $\log(D)$ factor. With a little bit of additional work, we can extend the result to hold for $\varepsilon + \delta > c - s$, with a weaker quantitative lower bound.

3 Preliminaries

We write $[n] := \{1, \dots, n\}$. For a permutation π , we write $C(\pi)$ to denote the number of cycles in its cycle decomposition. Let $U(d)$ denote the unitary group on $\mathbb{C}^d$. Whenever we average over $U(d)$, it is with respect to the Haar measure. We use standard definitions of QMA and QMA(2); see, e.g., Watrous [Wat09]. We also assume familiarity with the basics of quantum computing and quantum information; see, e.g., [Wat18].

Next, we define the product value of a bipartite operator, which measures its maximum overlap with a product state.

Definition 3.1 (Product value). For a positive semidefinite operator P on a bipartite Hilbert space $A \otimes B$, define its *product value* by

$$h_{\text{Sep}}(P) := \max_{\substack{|\alpha\rangle \in A, |\beta\rangle \in B \\ \|\alpha\| = \|\beta\| = 1}} \langle \alpha | \langle \beta | P | \alpha \rangle | \beta \rangle.$$

We also use the following well-known lower bound on the degree of polynomial approximations to OR.

Theorem 3.2 (Approximate Degree Bound for OR, [BCDZ99]). *Let $q : \mathbb{R} \rightarrow \mathbb{R}$ be a real polynomial of degree d . If $q(0) = 0$ and $1 - \varepsilon \leq q(i) \leq 1$ for all $1 \leq i \leq N$, then $d \geq \Omega(\min\{N, \sqrt{N \log(1/\varepsilon)}\})$.*

3.1 The antisymmetric and symmetric subspaces

Let S_k denote the symmetric group on k elements. For $\pi \in S_k$, let R_π be the unitary operator on $(\mathbb{C}^d)^{\otimes k}$ that permutes the k subsystems according to π :

$$R_\pi(v_1 \otimes \dots \otimes v_k) = v_{\pi^{-1}(1)} \otimes \dots \otimes v_{\pi^{-1}(k)}.$$

Definition 3.3 (Antisymmetric subspace). The *antisymmetric subspace* of $(\mathbb{C}^d)^{\otimes k}$ is

$$\wedge^k \mathbb{C}^d := \left\{ |\psi\rangle \in (\mathbb{C}^d)^{\otimes k} : R_\pi |\psi\rangle = \text{sgn}(\pi) |\psi\rangle \text{ for every } \pi \in S_k \right\}.$$

The orthogonal projector onto the antisymmetric subspace is defined as

$$\Pi_{\text{anti}}^{d,k} := \frac{1}{k!} \sum_{\pi \in S_k} \text{sgn}(\pi) R_\pi. \quad (4)$$

We refer to $\Pi_{\text{anti}}^{d,k}$ as the *antisymmetric projector*.

Definition 3.4 (Symmetric subspace). The *symmetric subspace* of $(\mathbb{C}^d)^{\otimes k}$ is

$$\vee^k \mathbb{C}^d := \left\{ |\psi\rangle \in (\mathbb{C}^d)^{\otimes k} : R_\pi |\psi\rangle = |\psi\rangle \text{ for every } \pi \in S_k \right\}.$$

The orthogonal projector onto the symmetric subspace is defined as

$$\Pi_{\text{sym}}^{d,k} := \frac{1}{k!} \sum_{\pi \in S_k} R_\pi. \quad (5)$$

We refer to $\Pi_{\text{sym}}^{d,k}$ as the *symmetric projector*.

3.2 Local-unitary invariant polynomials

Let $A = \mathbb{C}^{d_A}$ and $B = \mathbb{C}^{d_B}$. A polynomial f in the entries of an operator X on $A \otimes B$ is *local-unitary invariant* if

$$f\left((U \otimes V)X(U \otimes V)^\dagger\right) = f(X)$$

for every $U \in U(d_A)$ and $V \in U(d_B)$.

We will use the following standard characterization of these invariant polynomials [Bra37; Pro76; SY23].

Theorem 3.5. For $t \geq 1$ and $\pi, \sigma \in S_t$, let R_π^A and R_σ^B denote the operators that permute the t copies of A and B , respectively. Then every homogeneous degree- t local-unitary invariant polynomial is a linear combination of the polynomials

$$X \mapsto \text{tr}\left[X^{\otimes t} \left(R_\pi^A \otimes R_\sigma^B\right)\right], \quad \pi, \sigma \in S_t.$$

4 On the antisymmetric and symmetric projectors

We study the antisymmetric and symmetric projectors acting on four systems under the bipartition $(12 \mid 34)$. For convenience, we write $\Pi_d^- := \Pi_{\text{anti}}^{d,4}$ and $\Pi_d^+ := \Pi_{\text{sym}}^{d,4}$.

We prove two properties that drive the oracle separation. First, we show that the antisymmetric projector is a good candidate for NO instances of the entangled subspace problem, i.e., it has small overlap with every product state:

$$h_{\text{Sep}}(\Pi_d^-) \leq \frac{1}{3}.$$

Second, for every fixed local-unitary invariant polynomial G , we show that there is a single univariate polynomial g such that

$$g(d) = G(\Pi_d^-), \quad g(-d) = G(\Pi_d^+).$$

Intuitively, this means we will be able to apply the polynomial method to quantum algorithms that switch between querying the symmetric and antisymmetric subspaces simultaneously. Our first property shows that switching the sign of the subspace corresponds to changing between YES and NO instances of the entangled subspace problem, which will give us our degree lower bound.

4.1 Small overlap with product states

We compute the product value of the antisymmetric projector across the bipartition $(12 \mid 34)$.

Lemma 4.1. For every $d \geq 4$, $h_{\text{Sep}}(\Pi_d^-) \leq \frac{1}{3}$.

Proof. Since $\Pi_d^- = \Pi_d^-(\Pi_{\text{anti}}^{d,2} \otimes \Pi_{\text{anti}}^{d,2})$, it suffices to consider pairs of unit vectors in the antisymmetric subspace, $|\alpha\rangle, |\beta\rangle \in \wedge^2 \mathbb{C}^d$. Set $|\psi\rangle := |\alpha\rangle_{12} |\beta\rangle_{34}$.

By definition, the antisymmetric subspace projector can be written as $\sum_{\tau} \text{sgn}(\tau) R_{\tau}$. However, consider the subgroup $H = \{e, (12), (34), (12)(34)\}$, i.e., permutations that only swap within the two parts of $|\psi\rangle = |\alpha\rangle_{12} |\beta\rangle_{34}$. Since every $h \in H$ applies a permutation to a state in the antisymmetric subspace, by definition we have $R_h |\psi\rangle = \text{sgn}(h) |\psi\rangle$. Therefore, for any permutation $r \in S_4$,

$$\text{sgn}(rh) R_{rh} |\psi\rangle = \text{sgn}(r) R_r |\psi\rangle.$$

This means that for any pair of permutations π, π' such that $\pi = \pi' \circ \sigma$ for $\sigma \in H$, the terms $\text{sgn}(\pi) R_{\pi}$ and $\text{sgn}(\pi') R_{\pi'}$ act the same on $|\psi\rangle$. There are $|S_4|/|H| = 6$ cosets of H , and we choose representatives $e, (13), (14), (23), (24), (13)(24)$ from each coset to obtain our expression. Thus, when restricted to $\wedge^2 \mathbb{C}^d \otimes \wedge^2 \mathbb{C}^d$, the four-system antisymmetric projector has the following form:

$$\Pi_d^- = \frac{1}{6} (I - R_{13} - R_{14} - R_{23} - R_{24} + R_{13}R_{24}). \quad (6)$$

Here, we use R_{ij} denote the swap of registers i and j .

We can now evaluate $\langle \psi | \Pi_d^- | \psi \rangle$ given this reduced form. Let ρ_{α} and ρ_{β} denote the one-register reduced states of $|\alpha\rangle$ and $|\beta\rangle$, respectively. By antisymmetry, the two one-register marginals of each state are equal, so the choice of register traced out is irrelevant.

For every $i \in \{1, 2\}$ and $j \in \{3, 4\}$, the swap identity $\text{tr}(R(A \otimes B)) = \text{tr}(AB)$ gives

$$\langle \psi | R_{ij} | \psi \rangle = \text{tr}[R_{ij}(\rho_{\alpha} \otimes \rho_{\beta})] = \text{tr}(\rho_{\alpha} \rho_{\beta}).$$

Because ρ_{α} and ρ_{β} are positive semi-definite, $\text{tr}(\rho_{\alpha} \rho_{\beta}) \geq 0$, and all four single-swap terms in Equation (6) are nonnegative. The remaining term is also simple to evaluate. The operator $R_{13}R_{24}$ exchanges the two pairs of registers, so $R_{13}R_{24} |\alpha\rangle_{12} |\beta\rangle_{34} = |\beta\rangle_{12} |\alpha\rangle_{34}$. Therefore,

$$\langle \psi | R_{13}R_{24} | \psi \rangle = |\langle \alpha | \beta \rangle|^2 \leq 1.$$

Substituting these identities into Equation (6) gives

$$6 \langle \psi | \Pi_d^- | \psi \rangle = 1 - 4 \text{tr}(\rho_{\alpha} \rho_{\beta}) + |\langle \alpha | \beta \rangle|^2 \leq 2.$$

Hence, we have that for all product states $|\psi\rangle$,

$$\langle \psi | \Pi_d^- | \psi \rangle \leq \frac{1}{3}.$$

Taking the maximum over product states proves the claim. $\square$

4.2 Local-unitary invariants of the symmetric and antisymmetric projectors

We next study the behavior of Π_d^- and Π_d^+ under local-unitary invariant polynomials. In particular, we show that for every fixed local-unitary invariant polynomial G , there is a univariate polynomial g such that $g(d) = G(\Pi_d^-)$ and $g(-d) = G(\Pi_d^+)$.

To show this, it suffices by Theorem 3.5 to consider the spanning family of invariants defined as follows. For the bipartition $(12 | 34)$, let $A = B = (\mathbb{C}^d)^{\otimes 2}$. For an operator P on a bipartite Hilbert space $A \otimes B$, and

for $t \geq 1$ and $\pi, \sigma \in S_t$, define

$$J_{\pi, \sigma}(P) := \text{tr} \left[P^{\otimes t} \left(R_{\pi}^A \otimes R_{\sigma}^B \right) \right],$$

where R_{π}^A and R_{σ}^B permute the t copies of A and B , respectively.

Lemma 4.2. Fix $t \geq 1$ and $\pi, \sigma \in S_t$. There is a polynomial $q_{\pi, \sigma} \in \mathbb{R}[x]$ of degree at most $4t$ such that, for every integer $d \geq 4$,

$$q_{\pi, \sigma}(d) = J_{\pi, \sigma}(\Pi_d^-),$$

and, for every integer $d \geq 1$,

$$q_{\pi, \sigma}(-d) = J_{\pi, \sigma}(\Pi_d^+).$$

Proof. Let $\xi_{\pi, \sigma} \in S_{4t}$ be the permutation of the $4t$ underlying systems induced by $R_{\pi}^A \otimes R_{\sigma}^B$. Since A and B each consist of two systems and π and σ permute two-system blocks, $\xi_{\pi, \sigma}$ is even. For $\tau = (\tau_1, \dots, \tau_t) \in S_4^t$, write $\bar{\tau} := \tau_1 \oplus \dots \oplus \tau_t \in S_{4t}$.

We expand $(\Pi_d^-)^{\otimes t}$ as $\sum_{\tau \in S_4^t} \text{sgn}(\bar{\tau}) R_{\bar{\tau}}$ and apply this identity to get

$$J_{\pi, \sigma}(\Pi_d^-) = \text{tr}[(\Pi_d^-)^{\otimes t} R_{\xi_{\pi, \sigma}}] = \frac{1}{(4!)^t} \sum_{\tau \in S_4^t} \text{sgn}(\bar{\tau}) \text{tr}(R_{\bar{\tau} \xi_{\pi, \sigma}}).$$

Observe that $\text{tr}(R_{\rho}) = d^{C(\rho)}$. To see this, first note that the trace of R_{ρ} is the number of standard basis vectors that it fixes. A standard basis vector is fixed by R_{ρ} exactly when its indices are constant on each cycle of ρ , giving $d^{C(\rho)}$ such basis vectors. Continuing the above derivation, this yields

$$J_{\pi, \sigma}(\Pi_d^-) = \frac{1}{(4!)^t} \sum_{\tau \in S_4^t} \text{sgn}(\bar{\tau}) \text{tr}(R_{\bar{\tau} \xi_{\pi, \sigma}}) = \frac{1}{(4!)^t} \sum_{\tau \in S_4^t} \text{sgn}(\bar{\tau}) d^{C(\bar{\tau} \xi_{\pi, \sigma})} =: q_{\pi, \sigma}(d).$$

Since $C(\bar{\tau} \xi_{\pi, \sigma}) \leq 4t$, we have $\deg(q_{\pi, \sigma}) \leq 4t$.

Until now, we defined the polynomial $q_{\pi, \sigma}$ using only the antisymmetric subspace projector, but it has an expression that only depends on d , the input. Surprisingly, we will see that when we plug in a negative number for d , we get the same expression as if we had defined $q_{\pi, \sigma}$ using the symmetric subspace projector. Notice that for every $\rho \in S_{4t}$, we have $(-d)^{C(\rho)} = \text{sgn}(\rho) d^{C(\rho)}$. Indeed, suppose the disjoint-cycle decomposition of ρ has cycle lengths $\ell_1, \dots, \ell_{C(\rho)}$. Then each cycle of length ℓ_i has sign $(-1)^{\ell_i-1}$. Since $\sum_{i=1}^{C(\rho)} \ell_i = 4t$, we obtain

$$\text{sgn}(\rho) = \prod_{i=1}^{C(\rho)} (-1)^{\ell_i-1} = (-1)^{4t-C(\rho)} = (-1)^{C(\rho)},$$

where the last equality uses that $4t$ is even.

Plugging this back into the definition of $q_{\pi, \sigma}$, we have

$$q_{\pi, \sigma}(-d) = \frac{1}{(4!)^t} \sum_{\tau \in S_4^t} \text{sgn}(\bar{\tau}) \text{sgn}(\bar{\tau} \xi_{\pi, \sigma}) d^{C(\bar{\tau} \xi_{\pi, \sigma})} = \frac{1}{(4!)^t} \sum_{\tau \in S_4^t} d^{C(\bar{\tau} \xi_{\pi, \sigma})},$$

where the second equality uses $\text{sgn}(\bar{\tau}) \text{sgn}(\bar{\tau} \xi_{\pi, \sigma}) = \text{sgn}(\xi_{\pi, \sigma}) = 1$.

Remarkably, the last expression is exactly the expansion obtained from $(\Pi_d^+)^{\otimes t}$, since the symmetric projector contains no permutation signs. Hence

$$q_{\pi, \sigma}(-d) = J_{\pi, \sigma}(\Pi_d^+). \quad \square$$

Lemma 4.2 shows that the basic local-unitary invariants of Π_d^- and Π_d^+ are captured by the same low-

degree polynomial, with the symmetric family appearing at negative values of the dimension. We now extend this correspondence to arbitrary local-unitary invariant polynomials by expanding them into the local-unitary basis polynomials.

In the following lemma, we consider algorithms that can query the projectors $\Pi_d^\pm$ for different values of d . When this happens, we imagine that the projectors act on the first d dimensions of a pair of $\sqrt{D}$ -dimensional subspace, and act as 0 elsewhere. Since we will eventually twirl by a pair of Haar random unitaries, the choice of which subspace the projectors act on will not matter.

Lemma 4.3. *Let G be a real, local-unitary invariant polynomial of degree at most L on operators acting on $\mathbb{C}^D \otimes \mathbb{C}^D$. Then there exists a polynomial $g \in \mathbb{R}[x]$ of degree at most $4L$ such that for every integer $4 \leq d \leq \sqrt{D}$,*

$$g(d) = G(\Pi_d^-),$$

and for every integer $1 \leq d \leq \sqrt{D}$,

$$g(-d) = G(\Pi_d^+).$$

In particular, for every rank-one product projector $\psi \otimes \phi$ on $\mathbb{C}^D \otimes \mathbb{C}^D$, $g(-1) = G(\psi \otimes \phi)$.

Proof. Write $G = \sum_{t=0}^L G_t$, where G_t is homogeneous of degree t . Since G is local-unitary invariant, each G_t is also local-unitary invariant.

For $t \geq 1$, [Theorem 3.5](#) says we can express G_t as a linear combination of the basis polynomials

$$J_{\pi,\sigma}(P) = \text{tr} \left[P^{\otimes t} \left(R_\pi^A \otimes R_\sigma^B \right) \right], \quad \pi, \sigma \in S_t.$$

[Lemma 4.2](#) tells us that each of these basis polynomials can be written as a univariate polynomial in the dimension d . When we substitute those in, we find that there is a polynomial g_t in the dimension d , of degree at most $4t$, such that $g_t(d) = G_t(\Pi_d^-)$ for every $4 \leq d \leq \sqrt{D}$, and $g_t(-d) = G_t(\Pi_d^+)$ for every $1 \leq d \leq \sqrt{D}$. For $t = 0$, G_0 is constant, so we take $g_0 = G_0$.

Therefore, $g := \sum_{t=0}^L g_t$ has degree at most $4L$ and satisfies

$$g(d) = G(\Pi_d^-) \quad \text{and} \quad g(-d) = G(\Pi_d^+)$$

for every valid d .

Finally, note that every rank-one product projector on $\mathbb{C}^D \otimes \mathbb{C}^D$ is locally unitarily equivalent to Π_1^+ , so by the local-unitary invariance of G , $g(-1) = G(\psi \otimes \phi)$ for every rank-one product projector $\psi \otimes \phi$. $\square$

5 The oracle separation

We present the unitary-oracle separation between QMA and QMA(2). The separation is based on the antisymmetric-symmetric correspondence from [Section 4](#).

Theorem 5.1. *There exists a unitary oracle U such that*

$$\text{QMA}^U \neq \text{QMA}(2)^U.$$

We begin by defining the promise problem underlying the separation. Let

$$N = N(D) = \max\{j \in \mathbb{N} : (5j - 1)^2 \leq D\}. \quad (7)$$

We will consider $d_j := 5j - 1$ for each $j \in [N]$.⁵ Recall that Π_d^- acts on the bipartite space $(\mathbb{C}^d)^{\otimes 2} \otimes (\mathbb{C}^d)^{\otimes 2}$, whose local dimension is d^2 . By choice of N , $\Pi_{d_1}^-, \dots, \Pi_{d_N}^-$ can act on as projectors on $\mathbb{C}^D \otimes \mathbb{C}^D$. We fix such embeddings and, in a slight abuse of notation, continue to write $\Pi_{d_j}^-$ for the resulting projectors.

We define a family of yes and no instances for the entangled subspace problem, and we denote the problem of deciding between yes and no instances from this family the antisymmetric entangled subspace problem.

Definition 5.2 (Antisymmetric Entangled Subspace problem (AES)). Fix a rank-one product projector $\psi \otimes \phi$ on $\mathbb{C}^D \otimes \mathbb{C}^D$, and define

$$\begin{aligned}\mathcal{Y}_D &:= \left\{ (U \otimes V) \psi \otimes \phi (U \otimes V)^\dagger : U, V \in U(D) \right\}, \\ \mathcal{N}_D &:= \bigcup_{j=1}^{N(D)} \left\{ (U \otimes V) \Pi_{d_j}^- (U \otimes V)^\dagger : U, V \in U(D) \right\}.\end{aligned}$$

For a projector P on $\mathbb{C}^D \otimes \mathbb{C}^D$, define the unitary oracle

$$U_P := |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes (I - 2P). \quad (8)$$

The *Antisymmetric Entangled Subspace problem*, denoted $\text{AES}_D := (\mathcal{Y}_D, \mathcal{N}_D)$, is the promise problem of deciding, given oracle access to U_P , whether $P \in \mathcal{Y}_D$ or $P \in \mathcal{N}_D$.

By definition, YES instances are rank-one product projectors on $\mathbb{C}^D \otimes \mathbb{C}^D$, while the NO instances are locally unitarily equivalent to one of $\Pi_{d_1}^-, \dots, \Pi_{d_N}^-$. Since local unitaries and zero-padding preserve product value, our YES instances have $h_{\text{Sep}}(P) = 1$, and Lemma 4.1 implies that every NO instance satisfies $h_{\text{Sep}}(P) \leq \frac{1}{3}$.

The promise problem AES_D is a restriction of the Entangled Subspace problem of She and Yuen [SY23]. Recall that in the (a, b) -Entangled Subspace problem, a YES subspace contains a state that is a -close in trace distance to a product state, whereas in a NO subspace, every state is at least b -far from every product state. The problem AES_D is a special case of the $(0, 2/3)$ -Entangled Subspace problem. An interesting distinction between the general entangled subspace problem and the antisymmetric subspace problem is that our YES instances only have to consist of projectors onto rank one subspaces. On the other hand, every NO instance is at least $2/3$ -far from every product state, as follows from Lemma 4.1.

5.1 The QMA(2) upper bound

The QMA(2) upper bound for the Entangled Subspace problem of She and Yuen [SY23] already applies to our promise problem. For AES_D , however, the verification is particularly simple, so we include it for completeness.

Proposition 5.3. *AES_D has a one-query QMA(2) verifier with completeness 1 and soundness at most $1/3$.*

Proof. The QMA(2) verifier will receive a copy of a product state and measure the projector P . For completeness, we briefly remind the reader that the two-outcome measurement $\{I - P, P\}$ can be done with a single query to $\mathcal{O}_P$ via the Hadamard test: On input $|\psi\rangle$, prepare the control qubit in $|+\rangle$, query U_P , and applying a Hadamard to the control register to get the state

$$|0\rangle \otimes (I - P) |\psi\rangle + |1\rangle \otimes P |\psi\rangle.$$

⁵The choice of d_j is so that the affine map $d = 5x - 1$ sends $x = 0$ to the formal symmetric point $d = -1$, while sending $x = j$ to the antisymmetric dimension d_j . In particular, $d_1 = 4$ is the smallest dimension for which Π_d^- is nonzero.

Then, measuring the control qubit measures $\{P, I - P\}$.

If $P \in \mathcal{Y}_D$, then the range of P is spanned by a product state, which the two provers can send to make the verifier accept with probability 1.

If $P \in \mathcal{N}_D$, the projector P is locally unitarily equivalent to $\Pi_{d_j}^-$ for some $j \in [N(D)]$. Hence, by Lemma 4.1,

$$h_{\text{Sep}}(P) = h_{\text{Sep}}(\Pi_{d_j}^-) \leq \frac{1}{3},$$

where we also use that the zero-padding $\Pi_{d_j}^-$ into $\mathbb{C}^D \otimes \mathbb{C}^D$ does not change the product value. Therefore every product witness is accepted with probability at most $1/3$. $\square$

5.2 The QMA lower bound

We now prove that any QMA verifier for AES_D must either use many oracle queries or receive a large witness. Our proof has two main steps. On one hand, we apply the guessing lemma [Aar12] (or in-place amplification [MW05]) to find a low-degree polynomial that approximates the maximum acceptance probability of the verifier. Then, we use the symmetric–antisymmetric correspondence from Section 4.2 to lower bound the degree of this polynomial by the approximate degree of OR.

Our main theorem is as follows.

Theorem 5.4. *Fix constants $0 \leq s < c \leq 1$. Suppose a QMA verifier decides AES_D with completeness c and soundness s , using an m -qubit witness and q queries to U_P . Then*

$$q^2(m+1) = \Omega(\sqrt{D}).$$

We begin by converting the maximum acceptance probability of a QMA verifier into a low-degree polynomial. Fix a verifier V . For each projector P specifying an oracle U_P , let M_P denote its acceptance operator on the witness register, so that

$$\Pr[V^{U_P} \text{ accepts } \rho] = \text{tr}(M_P \rho)$$

for every witness state ρ . The maximum acceptance probability of V is thus

$$\lambda(P) := \lambda_{\max}(M_P).$$

Since U_P has degree at most one in the entries of P , the entries of M_P are low-degree polynomials in the entries of P . The obstacle is that taking the largest eigenvalue does not preserve this polynomial structure. We therefore use trace powers of M_P , which retain the polynomial dependence on P while still corresponding to its largest eigenvalue.

Lemma 5.5 (Trace power method, or the guessing lemma [Aar12; MW05]). *For every positive integer ℓ , the function*

$$H_\ell(P) := \text{tr}(M_P^\ell)$$

is a real polynomial of degree at most $2q\ell$ in the entries of P , and satisfies

$$\lambda(P)^\ell \leq H_\ell(P) \leq 2^m \lambda(P)^\ell.$$

Proof. Since U_P has degree at most one in the entries of P and V makes q oracle queries, every entry of M_P is a polynomial of degree at most $2q$. Hence $H_\ell(P) = \text{tr}(M_P^\ell)$ has degree at most $2q\ell$. If $\lambda_1, \dots, \lambda_{2^m}$ are the eigenvalues of M_P , then $H_\ell(P) = \sum_{i=1}^{2^m} \lambda_i^\ell$. Since $\lambda(P) = \max_i \lambda_i$, the claimed bounds follow immediately. $\square$

Although the above lemma is phrased in terms of a function H_ℓ , we note that this function exactly implements the guessing lemma from [Aar12], in the sense that M_P^ℓ is the result of applying the in-place amplification algorithm of Marriot and Watrous [MW04], and taking the trace corresponds to inputting a uniformly random witness (up to scaling by 2^m). Thus, this function H_ℓ actually corresponds to the accepting probability of a $q\ell$ -query quantum algorithm making queries to P .

We are now ready to prove the main theorem.

Proof of Theorem 5.4. Let V denote the QMA verifier. Pick a sufficiently large $\ell = O(m+1)$ such that $2^m(s/c)^\ell \leq 2^{-(m+1)}$. Let $H := H_\ell$ be the polynomial from Lemma 5.5. If P is a YES instance, then $\lambda(P) \geq 2/3$, and hence

$$H(P) \geq c^\ell.$$

If P is a NO instance, then $\lambda(P) \leq 1/3$, and hence

$$0 \leq H(P) \leq 2^m s^\ell \leq 2^{-(m+1)} c^\ell.$$

Note also that $\deg(H) \leq 2q\ell$. However, H is not necessarily local-unitary invariant. We find a local-unitary invariant polynomial by averaging H over local unitaries. Define

$$\bar{H}(P) := \mathbb{E}_{U,V \in U(D)} H\left((U \otimes V)P(U \otimes V)^\dagger\right),$$

where the expectation is over the Haar measure. Then $\bar{H}$ is a real local-unitary invariant polynomial of degree at most $2q\ell$. To see this, first note that conjugation by $U \otimes V$ is linear in P , and therefore cannot increase the degree of H . The local-unitary invariance follows from the fact that Haar measure is invariant under left and right multiplication by a fixed unitary.

Because applying local unitaries maps yes instances to yes instances and no instances to no instances, the bounds we found for the value of H also apply to $\bar{H}$. In particular,

$$\bar{H}(E) \geq c^\ell,$$

while for every $j \in [N]$, with $N = N(D)$,

$$0 \leq \bar{H}(\Pi_{d_j}^-) \leq 2^{-(m+1)} c^\ell.$$

By Lemma 4.3, there is a univariate polynomial h of degree at most $8q\ell$ such that $h(-1) = \bar{H}(E)$ and $h(d_j) = \bar{H}(\Pi_{d_j}^-)$ for every $j \in [N]$. To apply Theorem 3.2, we need to normalize and flip the polynomial so that it goes between 0 and a value close to 1 (as opposed to c^ℓ and $2^{-(m+1)}c^\ell$). To that end, define

$$p(x) := 1 - \frac{h(5x-1)}{h(-1)}.$$

Since $h(-1) = \bar{H}(E) \geq c^\ell > 0$, this is well defined. Moreover, we have that $p(0) = 0$, and for every $j \in [N]$,

$$1 - 2^{-(m+1)} \leq p(j) \leq 1.$$

Thus, by Theorem 3.2, $\deg(p) = \Omega(\sqrt{N(m+1)})$. On the other hand, $\deg(p) \leq 8q\ell$. Since $\ell = O(m+1)$, we conclude that $q^2(m+1) = \Omega(N)$. Finally, recalling from Equation (7) that $N = \lfloor \frac{1}{5}(\sqrt{D}-1) \rfloor$, this yields

$$q^2(m+1) = \Omega(\sqrt{D}). \quad \square$$

The oracle separation [Theorem 5.1](#) now follows by relatively standard arguments. For completeness, we include a proof of the diagonalization argument in the appendix.

6 The no-disentangles theorem

We derive the no-disentangles theorem as a consequence of the QMA lower bound from [Theorem 5.4](#). We rule out polynomial-size disentanglers throughout the full regime $\varepsilon + \delta < 1$. In particular, if each output register contains n qubits, then any (ε, δ) -disentangler requires $2^{\Omega_{\varepsilon, \delta}(n)}$ input qubits.

At a high level, the idea is that a good disentangler allows us to construct a QMA algorithm for the AES_D problem by using the disentangler and then calling the QMA(2) algorithm. We will see that the parameters ε and δ eat into the completeness / soundness gap of the QMA(2) verifier, and thus there are slight complications when $\varepsilon + \delta$ become larger than $2/3$; we include a proof in the appendix which mitigates this issue via parallel amplification of the QMA(2) verifier.

Lemma 6.1. *Assume that there exists a QMA(2) verifier taking a pair of witnesses with dimension D' , with completeness c and soundness s , for the AES_D problem, making t queries to the oracle. If exists a (ε, δ) -disentangler, Λ , with output dimension D' and m input qubits, then there exists a QMA algorithm for the AES_D problem making t -queries, taking a witness of size m , with completeness $c - \delta$ and soundness $s + \varepsilon$.*

Proof. The QMA algorithm simply applies the disentangler to their m -qubit witness and then calls the QMA(2) verifier on the resulting state and its oracle. By the guarantee of the QMA(2) verifier, if $\mathcal{O}$ is a YES instance of the AES_D problem, there exists a pair of states $\sigma_1 \otimes \sigma_2$ that causes it to accept with probability c . By the definition of a (ε, δ) -disentangler, there is a state ρ on m -qubit such that $\Lambda(\rho)$ is δ -close to $\sigma_1 \otimes \sigma_2$ in trace distance. On this state, the QMA verifier accepts with probability $c - \delta$.

Similarly, if $\mathcal{O}$ is a NO instance of the AES_D problem, then for all states $\sigma_1 \otimes \sigma_2$, the QMA(2) verifier accepts with probability at most s . By the definition of a (ε, δ) -disentangler, for all ρ given to the QMA verifier, there exists $\sigma_1 \otimes \sigma_2$ that is ε -close in trace distance to $\Lambda(\rho)$. Therefore, the QMA verifier accepts with probability at most $s + \varepsilon$. $\square$

From this, we can immediately get an almost tight bound on disentangler that satisfy $\varepsilon + \delta < 2/3$.

Corollary 6.2. *Fix constants $\varepsilon, \delta \geq 0$ such that $\varepsilon + \delta < 2/3$. Then every (ε, δ) -disentangler with output dimension D requires*

$$m = \Omega_{\varepsilon, \delta}(\sqrt{D}).$$

Proof. Note that the QMA(2) verifier from [Proposition 5.3](#) has completeness 1 and soundness $1/3$, makes 1 query, and has input dimension D . Thus, the existence of an (ε, δ) -disentangler and [Lemma 6.1](#) implies the existence of a QMA verifier that also makes 1 query, takes as input a m -qubit witness, and has completeness $1 - \delta$ and soundness $1/3 + \varepsilon$. Since ε and δ are constants, we can apply [Theorem 5.4](#) to get a lower bound of $m = \Omega(\sqrt{D})$, as desired. $\square$

In [Section B](#) we apply parallel amplification to get a bound in the general (ε, δ) setting.

Corollary 6.3. *Fix constants $\varepsilon, \delta \geq 0$ such that $2/3 < \varepsilon + \delta < 1$ and let $\eta = 1 - \varepsilon - \delta$. Then, every (ε, δ) -disentangler with output dimension D requires*

$$m = \Omega_{\varepsilon, \delta}\left(D^{1/\lceil \log_{4/3}(2/\eta) \rceil}\right)$$

input qubits.

To summarize the results in this section, when $\varepsilon + \delta < 2/3$, we obtain

$$m = \Omega_{\varepsilon, \delta}(\sqrt{D}).$$

This is essentially optimal (up to a $\log(D)$ factor) in light of recent work of Jeronimo, Wu, and Xu [JWX26]. For every fixed $\varepsilon \in (0, 1)$, they construct an $(\varepsilon, 0)$ -disentangler with local output dimension D using

$$m = O_{\varepsilon}(\sqrt{D} \log D)$$

input qubits. When $\varepsilon + \delta > 2/3$, we get a lower bound that scales as $D^{1/\log(\eta^{-1})}$, where η is the difference between 1 and $\varepsilon + \delta$. This scaling might be the correct form for small η , but our bound is almost certainly not tight up to constants. The proof could be improved to achieve a slightly improved bound (in some regimes) of $D^{1/(\lceil 1/\eta \rceil - 1)}$ by modifying the construction to use antisymmetric subspace projectors on different sized registers; however, this is also probably not tight. We leave the task of finding the optimal upper and lower bounds on disentanglers as interesting future work.

7 Conclusion

What was needed to obtain this result. Our paper’s moral starting point is the work of She and Yuen [SY23]. There, the authors made the observation that for any unitarily invariant property, a q -query QMA verifier reduces to the polynomial,

$$F(X) = \sum_{t=0}^{2q} \sum_{\sigma, \tau \in S_t} \text{tr}[(R_{\pi} \otimes R_{\sigma})X^{\otimes t}]. \quad (9)$$

To conclude the proof, two other observations are necessary:

1. Instantiate X above as the antisymmetric subspace projector Π_d^- . Then, the expression $\text{tr}[(R_{\pi} \otimes R_{\sigma})(\Pi_d^-)^{\otimes t}]$ yields a univariate polynomial in d .
2. Extending the resulting polynomial to *negative* formal dimension $d < 0$ yields the same polynomials but evaluated on *symmetric* subspace projectors.

These two steps are what allow us to apply standard polynomial lower bounds to obtain the QMA lower bound. As mentioned in the introduction, moving from Step 1 to Step 2 above also requires the somewhat subtle final choice to set the X as *two* copies of the antisymmetric subspace projectors (in the proof, we use the more restrictive choice $\Lambda^4 \mathbb{C}^d \subseteq \Lambda^2 \mathbb{C}^d \otimes \Lambda^2 \mathbb{C}^d$ to obtain a smaller product value). Nonetheless, it seems the primary missing ingredient towards a full oracle separation was instantiating the construction of [SY23] with the symmetric and antisymmetric subspaces. Thus, we believe that subsequent to the publication of [SY23] most of the technical ingredients to construct the oracle separation were present. However, putting together the various ingredients required familiarity with [SY23], one’s membership in the “Church of the Symmetric Subspace” [Har13], as well as intuition to regard the symmetric subspace as a “negative” dimension antisymmetric subspace [Pen71].

Although these ingredients (particularly the first two) are not unknown in quantum information, connecting these pieces to formulate a complete proof is precisely a strong point of generative AI models. We speculate that this contributed to the relative ease with which ChatGPT 5.6 Sol was able to produce the correct scaffolding of a proof with limited interaction.

What did we learn. It is not always straightforward to interpret the meaning of an oracle separation, especially a quantum oracle separation [AK26]. We can split our discussion into three parts: lessons that

would be true of any quantum oracle separation between QMA(2) and QMA, those that we learn from having arrived at our separation using the polynomial method, and those that are specific to our construction.

Rather than being definitive evidence that QMA(2) is not QMA, quantum oracle separations should be interpreted more as “barriers” to certain reductions from QMA(2) to QMA. One way to interpret the canonical QMA(2) complete problem (sparse separable Hamiltonian [CS12]) is as follows: Given an order 4 tensor T with entries $T_{i,j,k,\ell}$, determine its injective 2-norm, that is $\max_{\|\psi\|_2=1} \langle T, \psi^{\otimes 4} \rangle$ [BBHKSZ12]. We can also talk about the flattening of the tensor T to a matrix, $T_{\text{flat}} = \sum_{i,j,k,\ell} T_{i,j,k,\ell} |i,j\rangle\langle k,\ell|$. In some sense, separating QMA(2) from QMA relative to a quantum oracle tells us that finding the injective 2-norm of T is extremely hard, *if* you only get access to T through its matrix flattening, T_{flat} .

To better understand what this means, it is useful to think about what our proof says about the differences between matrices and tensors. At a high level, our proof exploits the fact that matrices can be diagonalized into eigenvectors, which gives a low-degree approximation to the spectral norm (via the trace power method, or in-place amplification) that does not exist for the injective 2-norm. This means that algorithms that get oracle access to the matrix flattening are (in some sense) not sensitive enough to changes in the entries of the matrix, whereas algorithms solving the sparse separable Hamiltonian problem must be extremely sensitive to certain changes in the structure of the underlying tensor. In some sense, this is not a new phenomenon. The idea that the injective 2-norm does not have a low-degree polynomial approximation appears in many works on problems like tensor PCA and sum-of-squares algorithms [HSS15]. We also point out that this oracle separation very much mirrors the seminal work of Aaronson and Kuperberg, who provided the first quantum oracle separation between QMA and QCMA [AK07]. In their work, the authors argued (via a geometric argument) that polynomial-sized classical strings are not sensitive enough to changes in the entries of a quantum state to describe a random state. While their work was extremely influential and impressive, we believe that both their work and ours serve to formalize and recast folklore knowledge in the language of quantum complexity theory, rather than teach us something fundamentally new about the nature of quantum proofs.

One additional lesson from the construction of this paper is that it pinpoints a very specific way in which tensor and matrix structure can differ. In particular, this paper shows that switching between symmetric and antisymmetric subspace projectors is a smooth transition in terms of the matrix (in that there is a single low-degree polynomial that interpolates between them), but yields a very sharp transition when viewed as a tensor. This surprising fact has been noted in other works on mathematical physics (see, for instance [Kin71; Pen71; Dun89; KKMT23; ECK05], among others), but to the best of our knowledge has not been noticed by the computer science community. As a result, this observation gives a potentially exciting way to solve problems related to higher-order tensors.

Future directions. We believe this work is only the beginning of a long investigation into the nature of un-entanglement and unitary property testing. We hope that researchers continue to be interested in understanding the complexity class QMA(2), and want to propose a few directions that we believe will lead to interesting insights about it.

1. **A classical oracle separation?** Perhaps the most natural follow-up to this line of work is to search for a classical oracle separation. This work heavily exploits the sensitivity of quantum algorithms to the entries of a matrix, when viewed as complex numbers. However, similar to work on QMA versus QCMA, classical oracle access to a tensor is fundamentally different than quantum oracle access to it. In particular, while changing a single entry of the matrix would lead to a small change in the underlying matrix, it leads to a huge change in the classical description of the matrix, meaning that a classical oracle separation cannot exploit the same kinds of arguments used in this work.

We note that the final resolution of the QMA versus QCMA problem ultimately relied on a completely

different idea (based on the sampling reduction of [BHNZ26]) than the quantum oracle separation (conditioning on a “popular” witness) of [AK07]. Indeed, many follow-up works focused on dequantizing [AK07] directly and subsequently ran into technical difficulties. The same may hold true for the QMA versus QMA(2) question: it is clear that the classical oracle separation cannot use a lot of the powerful invariant theory that was leveraged here, and our oracle construction does not have an obvious classical analogue. We thus believe that the problem of finding a classical oracle separation between QMA(2) and QMA is a fundamentally challenging and interesting problem, and likely to require completely different ideas than those presented in this work.

Given the history of the QMA versus QCMA problem [BHNZ26; BHV26], we speculate that examining the differing operational characteristics of QMA and QMA(2) may prove useful for separating the two classes relative to a classical oracle (or even in the unrelativized setting). In addition to the fact that QMA enjoys a strong form of in-place soundness amplification that seems to be missing from QMA(2), we note that prior literature [ABDFS09; BT12] has shown that QMA(2) allows for the use of surprisingly succinct proofs (in a way that we suspect should not hold for QMA). Exploiting these distinguishing behaviors seems crucial both to dequantizing our separation and to understanding the broader relationship of QMA(2) with other complexity classes.

2. **More unitary property testing?** Another interesting direction is to return to the original inspiration behind [SY23], and study other problems related to unitary property testing. The construction from this paper provides a very interesting way to smoothly transform (locally-invariant) matrices across a sharp boundary, and may be useful for understanding other problems in unitary property testing. Concretely, [SY23] leaves open the problem of finding tight bounds for the recurrence time and entropy estimation problems, as well as finding interesting problems related to unitaries that are invariant under conjugation by various groups (discussed in [Pro76]).
3. **State property testing?** One can formulate an analogue of QMA and QMA(2) for state property testing; these are propQMA and propQMA(2) [JMSW24]. In addition to copies of the target state, you are also provided with unentangled proof states. A single proof is not helpful, as the de-Merlinization of [Aar05] can be used to remove a single proof state, but this argument does not work for $k > 1$. Could the techniques in this paper be used to exhibit a property which has exponential copy-complexity, yet is efficiently testable given $k \geq 2$ copies of a proof? As a first attempt, one could consider the mixed states proportional to the symmetric and antisymmetric subspace projectors. Interestingly, the situation is a reversal of our oracle separation, where the QMA(2) algorithm is trivial but the QMA lower bound requires the most work. For property testing, the normalization makes the propQMA lower bound straightforward, but there is no longer a clear propQMA(k) algorithm for this task.
4. **UniqueQMA and approximate counting?** Using the same proof techniques as in this paper, one can show a *quantum* oracle separation between QMA and uniqueQMA, with the northstar problem in this setting being a classical oracle separation between QMA and $BQP^{\text{uniqueQMA}}$. This is not a quantitative improvement over [AHHN26] which exhibits a classical oracle separation between QMA and uniqueQMA, but the proof is simpler. We consider the “no” case to be the 0-dimensional symmetric subspace, and the “yes” instances to be symmetric subspaces of dimension at least 2 (this is a subfamily of the EmptyNonEmpty oracle problem from Anshu et al.). There is one subtlety: the polynomial constructed for the OR bound corresponds to the *trace* of the acceptance operator. This means in the “yes” case, the value could be exponentially large. The proof of Theorem 5.4 resolves this issue by including only a single, appropriately scaled “yes” case, but this does not work for uniqueQMA. Instead, the unique promise is crucially used to bound the trace by a constant; this is the differentiating feature between QMA and uniqueQMA, which yields the separation. The work of [AHHN26] also

constructs oracle separations between approximate counting and QMA, as well as QMA^{QMA} . Interestingly, approximate counting is related to a *classical* oracle separation between QMA and $\text{QMA}(2)$; more precisely, if one could show that approximate counting was in $\text{QMA}(2)$, this, combined with the existing QMA lower bounds [AKKT19] would yield a classical oracle separation between QMA and $\text{QMA}(2)$! Thus, understanding the precise relationship between approximate counting and $\text{QMA}(2)$ is an intriguing line of research.

References

- [Aar05] Scott Aaronson. *QMA/qpoly is contained in PSPACE/poly: de-Merlinizing quantum protocols*. 2005. arXiv: [quant-ph/0510230](#) [quant-ph] (page 20).
- [Aar12] Scott Aaronson. “Impossibility of Succinct Quantum Proofs for Collision-Freeness”. In: *Quantum Information & Computation* (2012). DOI: [10.5555/2231036.2231039](#). arXiv: [1101.0403](#) [quant-ph] (pages 15, 16).
- [Aar21] Scott Aaronson. “Open problems related to quantum query complexity”. In: *ACM Transactions on Quantum Computing* (2021). DOI: [10.1145/3488559](#). arXiv: [2109.06917](#) [quant-ph] (page 1).
- [ABDFS09] Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. “The power of unentanglement”. In: *Theory of Computing* (2009). DOI: [10.4086/toc.2009.v005a001](#). arXiv: [0804.0802](#) [quant-ph] (pages 1–3, 20).
- [AHHN26] Anurag Anshu, Jonas Haferkamp, Yeongwoo Hwang, and Quynh T Nguyen. “UniqueQMA vs QMA: oracle separation and eigenstate thermalization hypothesis”. In: *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*. 2026. DOI: [10.4230/LIPIcs.ITCS.2026.10](#). arXiv: [2410.23811](#) [quant-ph] (page 20).
- [AK07] Scott Aaronson and Greg Kuperberg. “Quantum versus classical proofs and advice”. In: *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC’07)*. IEEE. 2007. arXiv: [quant-ph/0604056](#) [quant-ph] (pages 1, 19, 20).
- [AK26] Avantika Agarwal and Srijita Kundu. *Non-Standard Oracles for Bounded-Error Complexity Classes*. 2026. arXiv: [2504.19470](#) [quant-ph]. URL: <https://arxiv.org/abs/2504.19470> (page 18).
- [AKKT19] Scott Aaronson, Robin Kothari, William Kretschmer, and Justin Thaler. “Quantum lower bounds for approximate counting via Laurent polynomials”. In: *arXiv preprint arXiv:1904.08914* (2019) (page 21).
- [AKT25] Seiseki Akibue, Go Kato, and Seiichiro Tani. “On the hardness of conversion from entangled proof into separable one”. In: *Mathematical Foundations for Post-Quantum Cryptography: Crypto-Math CREST*. 2025 (page 2).
- [BBCMD01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald De Wolf. “Quantum lower bounds by polynomials”. In: *Journal of the ACM (JACM)* (2001). DOI: [10.1145/502090.502097](#). arXiv: [quant-ph/9802049](#) [quant-ph] (page 3).
- [BBHKSZ12] Boaz Barak, Fernando GSL Brandão, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. “Hypercontractivity, sum-of-squares proofs, and their applications”. In: *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 2012 (page 19).
- [BCDZ99] Harry Buhrman, Richard Cleve, Ronald De Wolf, and Christof Zalka. “Bounds for small-error and zero-error quantum algorithms”. In: *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*. IEEE. 1999 (page 9).
- [BFL91] László Babai, Lance Fortnow, and Carsten Lund. “Non-deterministic exponential time has two-prover interactive protocols”. In: *Computational complexity* (1991) (page 1).
- [BHNZ26] John Bostanci, Jonas Haferkamp, Chinmay Nirkhe, and Mark Zhandry. “Separating QMA from QCMA with a Classical Oracle”. In: *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*. 2026. DOI: [10.1145/3798129.3800776](#). URL: <https://doi.org/10.1145/3798129.3800776> (page 20).

- [BHV26] John Bostanci, Andrew Huang, and Vinod Vaikuntanathan. “Separating Quantum and Classical Advice with Good Codes”. In: *arXiv preprint arXiv:2602.09385* (2026) (page 20).
- [Bra37] Richard Brauer. “On algebras which are connected with the semisimple continuous groups”. In: *Annals of Mathematics* (1937) (page 10).
- [BT12] Hugue Blier and Alain Tapp. “A quantum characterization of NP”. In: *computational complexity* (2012). DOI: [10.1007/s00037-011-0016-2](https://doi.org/10.1007/s00037-011-0016-2) (page 20).
- [CS12] Andre Chailloux and Or Sattath. “The Complexity of the Separable Hamiltonian Problem”. In: *Proceedings of the 2012 IEEE Conference on Computational Complexity*. 2012. DOI: [10.1109/CCC.2012.42](https://doi.org/10.1109/CCC.2012.42) (page 19).
- [CSW12] Matthias Christandl, Norbert Schuch, and Andreas Winter. “Entanglement of the Antisymmetric State”. In: *Communications in Mathematical Physics* (2012). DOI: [10.1007/s00220-012-1446-7](https://doi.org/10.1007/s00220-012-1446-7). URL: <http://dx.doi.org/10.1007/s00220-012-1446-7> (page 25).
- [Dun89] G V Dunne. “Negative-dimensional groups in quantum physics”. In: *Journal of Physics A: Mathematical and General* (1989). DOI: [10.1088/0305-4470/22/11/014](https://doi.org/10.1088/0305-4470/22/11/014). URL: <https://doi.org/10.1088/0305-4470/22/11/014> (page 19).
- [ECK05] Henriette Elvang, Predrag Cvitanovi, and Anthony D Kennedy. “Diagrammatic Young projection operators for $U(n)$ ”. In: *Journal of mathematical physics* (2005) (page 19).
- [Har13] Aram W Harrow. *The church of the symmetric subspace*. 2013. arXiv: [1308.6595 \[quant-ph\]](https://arxiv.org/abs/1308.6595). URL: <https://arxiv.org/abs/1308.6595> (page 18).
- [HNW19] Aram W Harrow, Anand Natarajan, and Xiaodi Wu. “Limitations of semidefinite programs for separable states and entangled games”. In: *Communications in Mathematical Physics* (2019) (page 2).
- [HSS15] Samuel B. Hopkins, Jonathan Shi, and David Steurer. *Tensor principal component analysis via sum-of-squares proofs*. 2015. arXiv: [1507.03269 \[cs.LG\]](https://arxiv.org/abs/1507.03269). URL: <https://arxiv.org/abs/1507.03269> (page 19).
- [JMSW24] Fernando Granha Jeronimo, Nir Magrafta, Joseph Slope, and Pei Wu. “Coherence in property testing: Quantum-classical collapses and separations”. In: *arXiv preprint arXiv:2411.15148* (2024) (page 20).
- [JNVWY20] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. $MIP^* = RE$. 2020. arXiv: [2001.04383 \[quant-ph\]](https://arxiv.org/abs/2001.04383). URL: <https://arxiv.org/abs/2001.04383> (page 1).
- [JWL26] Fernando Granha Jeronimo, Pei Wu, and Itai Leigh. “The QMA(2) Universe: Complexity, Entanglement, and Optimization”. In: *ACM SIGACT News* (2026). DOI: [10.1145/3802807.3802814](https://doi.org/10.1145/3802807.3802814) (page 1).
- [JWX26] Fernando Granha Jeronimo, Pei Wu, and Haochen Xu. *Optimal Quantum de Finetti Theorems via Argmax Rounding*. 2026. arXiv: [2604.27886 \[quant-ph\]](https://arxiv.org/abs/2604.27886) (pages 2, 9, 18).
- [Kin71] R. C. King. “The Dimensions of Irreducible Tensor Representations of the Orthogonal and Symplectic Groups”. In: *Canadian Journal of Mathematics* (1971). DOI: [10.4153/CJM-1971-017-2](https://doi.org/10.4153/CJM-1971-017-2) (page 19).
- [KKMT23] H Keppler, T Krajewski, T Muller, and A Tanasa. “Duality of $O(N)$ and $Sp(N)$ random tensor models: tensors with symmetries”. In: *Journal of Physics A: Mathematical and Theoretical* (2023). DOI: [10.1088/1751-8121/ad0af4](https://doi.org/10.1088/1751-8121/ad0af4). URL: <http://dx.doi.org/10.1088/1751-8121/ad0af4> (page 19).
- [KMY01] Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. *Quantum certificate verification: Single versus multiple quantum certificates*. 2001. arXiv: [quant-ph/0110006 \[quant-ph\]](https://arxiv.org/abs/quant-ph/0110006) (page 1).
- [KW00] Alexei Y. Kitaev and John Watrous. “Parallelization, amplification, and exponential time simulation of quantum interactive proof systems”. In: *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*. 2000. DOI: [10.1145/335305.335387](https://doi.org/10.1145/335305.335387). URL: <https://doi.org/10.1145/335305.335387> (page 1).
- [MW04] Chris Marriott and John Watrous. “Quantum Arthur-Merlin Games”. In: *19th Annual IEEE Conference on Computational Complexity (CCC 2004), 21-24 June 2004, Amherst, MA, USA*. 2004. DOI: [10.1109/CCC.2004.1313850](https://doi.org/10.1109/CCC.2004.1313850). URL: <https://doi.org/10.1109/CCC.2004.1313850> (page 16).
- [MW05] Chris Marriott and John Watrous. “Quantum Arthur-Merlin Games”. In: *Computational Complexity* (2005). DOI: [10.1007/s00037-005-0194-x](https://doi.org/10.1007/s00037-005-0194-x) (page 15).

- [Pen71] Roger Penrose. “Applications of negative dimensional tensors”. In: *Combinatorial Mathematics and its Applications*. 1971 (pages 18, 19).
- [Pro76] Claudio Procesi. “The invariant theory of $n \times n$ matrices”. In: *Advances in mathematics* (1976). DOI: [10.1016/0001-8708\(76\)90027-X](https://doi.org/10.1016/0001-8708(76)90027-X) (pages 10, 20).
- [SY23] Adrian She and Henry Yuen. “Unitary property testing lower bounds by polynomials”. In: *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. 2023. DOI: [10.4230/LIPIcs.ITCS.2023.96](https://doi.org/10.4230/LIPIcs.ITCS.2023.96). arXiv: [2210.05885 \[quant-ph\]](https://arxiv.org/abs/2210.05885) (pages 2–4, 10, 14, 18, 20).
- [Wat09] John Watrous. “Quantum Computational Complexity”. In: *Encyclopedia of Complexity and Systems Science*. 2009. DOI: [10.1007/978-0-387-30440-3_428](https://doi.org/10.1007/978-0-387-30440-3_428). arXiv: [0804.3401](https://arxiv.org/abs/0804.3401) (page 9).
- [Wat18] John Watrous. *The Theory of Quantum Information*. 2018. DOI: [10.1017/9781316848142](https://doi.org/10.1017/9781316848142). URL: <https://doi.org/10.1017/9781316848142> (page 9).

A Diagonalization argument

We use the query lower bound to construct the oracle, allowing us to prove [Theorem 5.1](#). The idea is standard: enumerate all polynomial-time QMA oracle verifiers and defeat them one at a time. At stage i , we choose a fresh input length n_i and use [Theorem 5.4](#) to choose the n_i -th oracle component on which the i -th verifier fails.

Theorem 5.1. *There exists a unitary oracle U such that*

$$\text{QMA}^U \neq \text{QMA}(2)^U.$$

Proof. We will construct an input-length-indexed unitary oracle $U = (U^{(n)})_{n \geq 1}$. The n -th component acts on one control qubit and two n -qubit registers. For a projector P_n on $\mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n}$, we set

$$U^{(n)} = U_{P_n} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes (I - 2P_n).$$

We now construct the projectors P_n together with a unary language L^U .

Enumerate all polynomial-time QMA oracle verifiers $V_1, V_2, \dots$ together with explicit polynomial bounds on their running time. For each V_i , let $m_i(n)$ and $q_i(n)$ be polynomial bounds on its witness length and number of oracle queries on input 1^n , respectively. We also let $a_i(n)$ be a polynomial bound on the largest oracle component that V_i can query on input 1^n . Such a bound must exist because the r -th oracle component itself acts on $2r + 1$ qubits.

We construct the oracle in stages. After stage i , we maintain a number B_i such that every oracle component of index at most B_i has been permanently fixed. Initially, set $B_0 = 0$. Suppose stages $1, \dots, i - 1$ have been completed. We first choose a fresh length $n_i > B_{i-1}$. By [Theorem 5.4](#), there is a constant $C > 0$ such that any QMA verifier for AES_{2^n} with completeness $2/3$ and soundness $1/3$, using an m -qubit witness and q oracle queries, satisfies

$$q^2(m + 1) \geq C\sqrt{2^n}$$

for all sufficiently large n . We choose n_i sufficiently large that

$$q_i(n_i)^2(m_i(n_i) + 1) < CN(\sqrt{2^{n_i}}).$$

where $N(\sqrt{2^n})$ is the function defined in [Equation \(7\)](#). Such a choice is always possible because the left-hand side grows polynomially in n_i , whereas the right-hand side grows exponentially in n_i .

We now fix all other oracle components that V_i could access on this input. Set

$$B_i := \max\{B_{i-1}, a_i(n_i), n_i\}.$$

For every still-undefined component of index at most B_i , except for n_i itself, set $P_n = 0$, so that $U^{(n)} = I$. At this point, every oracle component that V_i can query on input 1^{n_i} has been fixed except $U^{(n_i)}$.

Thus, with all other queried oracle components fixed, V_i on input 1^{n_i} becomes a verifier for the promise problem $\text{AES}_{2^{n_i}}$. It remains only to choose the projector P_{n_i} . By our choice of n_i and [Theorem 5.4](#), V_i cannot satisfy completeness $2/3$ and soundness $1/3$ for every $P \in \mathcal{Y}_{2^{n_i}} \cup \mathcal{N}_{2^{n_i}}$. Hence there exists some such projector P_{n_i} on which V_i fails. Fix P_{n_i} to be such a projector, and define

$$1^{n_i} \in L^U \iff P_{n_i} \in \mathcal{Y}_{2^{n_i}}.$$

This completes the construction at the i -th stage. Note that all oracle components of index at most B_i are now permanently fixed. Since $n_i > B_{i-1}$ and $B_i \geq n_i$, we have $B_i \rightarrow \infty$. Thus every oracle component is eventually fixed. Since V_i cannot query any component of index larger than $a_i(n_i) \leq B_i$ on input 1^{n_i} , no later stage can change its computation on this input. Thus V_i remains incorrect on 1^{n_i} in the final oracle. We reiterate that the promise problem AES_D is only used at a sequence of selected input lengths. At all other lengths we set $P_n = 0$ and declare 1^n to be a NO instance of the resulting unary language.

To complete the oracle, we repeat this construction for every i . By construction, for every i , the verifier V_i fails to decide L^U on input 1^{n_i} . Since every polynomial-time QMA oracle verifier appears in the enumeration, no such verifier decides L^U . Therefore,

$$L^U \notin \text{QMA}^U.$$

All that remains is to show that $L^U \in \text{QMA}(2)^U$. Following [Proposition 5.3](#), there is a single uniform verifier. On input 1^n , the two provers each send an n -qubit state, and the verifier uses one query to $U^{(n)}$ to perform the measurement $\{P_n, I - P_n\}$, accepting on outcome P_n .

If $1^n \in L^U$, then P_n is a rank-one product projector, so the provers can make the verifier accept with probability 1. If $1^n \notin L^U$, then either

$$P_n \in \mathcal{N}_{2^n} \quad \text{or} \quad P_n = 0.$$

In the former case, [Proposition 5.3](#) implies that every product witness is accepted with probability at most $1/3$. In the latter case, the verifier rejects with probability 1. Thus, we have established

$$L^U \in \text{QMA}(2)^U.$$

Combining with $L^U \notin \text{QMA}^U$, we conclude that

$$\text{QMA}^U \neq \text{QMA}(2)^U. \quad \square$$

B Improved disentangler bounds

We apply parallel amplification to the $\text{QMA}(2)$ verifier to obtain a (admittedly loose) lower bound for arbitrary ε, δ .

Lemma B.1. *For every $s \leq 1$ and $k = \lceil \log_{4/3}(1/s) \rceil$, there is a $\text{QMA}(2)$ verifier that makes k -queries to an oracle, takes as input a pair of un-entangled witnesses each of dimension D^k , and has completeness 1 and soundness s .*

Proof. The amplified $\text{QMA}(2)$ verifier takes a witness $|\psi\rangle_{A_1 \dots A_k} \otimes |\phi\rangle_{B_1 \dots B_k}$, where each register A_i, B_i has

dimension D . Then, it uses its oracle to measure a projector P k -many times on each pair $A_i B_i$. This verifier makes k oracle queries and takes as input a pair of registers of dimension D^k , as in the lemma statement.

In the YES case, the separable value of a tensor product of many un-entangled subspaces is still 1, so there is a proof that causes the verifier to accept with probability 1. In the NO case, we note that similar to the proof of [Lemma 4.1](#), the projector $\Pi_d^{\otimes k}$ can be written as $\Pi_d^{\otimes k} (\Pi_{\text{anti}}^{d,2})^{\otimes 2k}$. Lemma 12 from [\[CSW12\]](#) shows that the separable value of n copies of the antisymmetric subspace has a $h_{\text{sep}} \leq (3/4)^{n/2}$. Setting $n = 2k$, and setting $k = \lceil \log_{4/3}(1/s) \rceil$ gives us the desired soundness. $\square$

Corollary 6.3. *Fix constants $\epsilon, \delta \geq 0$ such that $2/3 < \epsilon + \delta < 1$ and let $\eta = 1 - \epsilon - \delta$. Then, every (ϵ, δ) -disentangler with output dimension D requires*

$$m = \Omega_{\epsilon, \delta} \left(D^{1/\lceil \log_{4/3}(2/\eta) \rceil} \right)$$

input qubits.

Proof. We first apply [Lemma B.1](#) to the QMA(2) verifier from [Proposition 5.3](#) to get a QMA(2) verifier for the AES_D problem that has soundness $\eta/2$ and completeness 1. Since our original verifier took a witness with dimension D , this new witness will be dimension $D^{\lceil \log_{4/3}(2/\eta) \rceil}$, and the algorithm will make $\lceil \log_{4/3}(2/\eta) \rceil$ queries. Now we set $D' = D^{1/\lceil \log_{4/3}(2/\eta) \rceil}$, so that the solving the $\text{AES}_{D'}$ problem requires a witness of size D .

From [Lemma 6.1](#), we now have that applying the disentangler to the m -qubit witness for the QMA verifier yields a QMA algorithm for the $\text{AES}_{D'}$ problem with completeness-soundness gap $\eta/2$ (which is a constant), making $O(\log(1/\eta))$ queries to the oracle. Using [Theorem 5.4](#), we have the bound of

$$\log(1/\eta)^2(m+1) = \Omega(\sqrt{D'}).$$

Substituting $D' = D^{1/\lceil \log_{4/3}(2/\eta) \rceil}$, we have that

$$m = \Omega \left(\frac{1}{\log(1/\eta)^2} D^{1/\lceil \log_{4/3}(2/\eta) \rceil} \right),$$

as desired. $\square$