Testing product states, quantum Merlin-Arthur games and tensor optimisation

Aram W. Harrow[*] and Ashley Montanaro[†]

October 26, 2012

Abstract

We give a test that can distinguish efficiently between product states of n quantum systems and states which are far from product. If applied to a state $|\psi\rangle$ whose maximum overlap with a product state is $1-\epsilon$, the test passes with probability $1-\Theta(\epsilon)$, regardless of n or the local dimensions of the individual systems. The test uses two copies of $|\psi\rangle$. We prove correctness of this test as a special case of a more general result regarding stability of maximum output purity of the depolarising channel.

A key application of the test is to quantum Merlin-Arthur games with multiple Merlins, where we obtain several structural results that had been previously conjectured, including the fact that efficient soundness amplification is possible and that two Merlins can simulate many Merlins: $\mathsf{QMA}(k) = \mathsf{QMA}(2)$ for $k \geq 2$. Building on a previous result of Aaronson et al., this implies that there is an efficient quantum algorithm to verify 3-SAT with constant soundness, given two unentangled proofs of $\widetilde{O}(\sqrt{n})$ qubits. We also show how $\mathsf{QMA}(2)$ with log-sized proofs is equivalent to a large number of problems, some related to quantum information (such as testing separability of mixed states) as well as problems without any apparent connection to quantum mechanics (such as computing injective tensor norms of 3-index tensors). As a consequence, we obtain many hardness-of-approximation results, as well as potential algorithmic applications of methods for approximating $\mathsf{QMA}(2)$ acceptance probabilities.

Finally, our test can also be used to construct an efficient test for determining whether a unitary operator is a tensor product, which is a generalisation of classical linearity testing.

1 Introduction

Entanglement of quantum states presents both an opportunity and a difficulty for quantum computing. To describe a pure state of n qudits (d-dimensional quantum systems) requires a comparable number of parameters to a classical probability distribution on d^n items. Effective methods are known for testing properties of probability distributions. However, for quantum states many of these tools no longer work. For example, due to interference, the probability of a test passing cannot be simply written as an average over components of the state. Moreover, measuring one part of a state and conditioning on the measurement outcome may induce entanglement between other parts of the state that were not previously entangled with each other.

These counter-intuitive properties of entanglement account for many of the outstanding puzzles in quantum information. In quantum channel coding, the famous additivity violations of [27, 40]

[*] Department of Computer Science & Engineering, University of Washington; aram@cs.washington.edu.

[†] Department of Applied Mathematics and Theoretical Physics, University of Cambridge; am994@cam.ac.uk.

reflect how entangled inputs can sometimes have advantages against even uncorrelated noise. For quantum interactive proofs, the primary difficulty is in bounding the ability of provers to cheat using entangled strategies [48]. Even for $\mathsf{QMA}(k)$ (the variant of QMA with k unentangled Merlins [50, 2]), most important open questions could be resolved by finding a way to control entanglement within each proof. Here, the recently discovered failure of parallel repetition for entangled provers [49] is a sort of complexity-theoretic analogue of additivity violations.

The situation is different when we consider quantum states that are *product* across the n systems. In this case, while individual systems of course behave quantumly, the lack of correlation between the systems means that classical tools such as Chernoff bounds can be used. For example, in channel coding with product-state inputs, not only does the single-letter Holevo formula give the capacity, so that there is no additivity problem, but so-called strong converse theorems are known, which prove that attempting to communicate at a rate above the capacity results in an exponentially decreasing probability of successfully transmitting a message [61, 73]. Naturally, many of the difficulties in dealing with entangled proofs and quantum parallel repetition would also go away if quantum states were constrained to be in product form.

1.1 Our results

In this paper, we present a quantum test to determine whether an n-partite state $|\psi\rangle$ is a product state or far from any product state. We make no assumptions about the local dimensions of $|\psi\rangle$; in fact, the local dimension can even be different for different systems. The test passes with certainty if $|\psi\rangle$ is product, and fails with probability $\Theta(\epsilon)$ if the overlap between $|\psi\rangle$ and the closest product state is $1-\epsilon$. An essential feature of our test (or any possible such test, as we will argue in Section 5) is that it requires two copies of $|\psi\rangle$.

The parameters of our test resemble classical property-testing algorithms [30]. In general, these algorithms make a small number of queries to some object and accept with high probability if the object has some property P (*completeness*), and with low probability if the object is "far" from having property P (*soundness*). Crucially, the number of queries used and the success probability should not depend on the size of the object. The main result of this paper is a test for a property of a quantum state, in contrast to previous work on quantum generalisations of property testing, which has considered quantum algorithms for testing properties of classical (e.g. [19, 6]) and quantum [59] oracles (a.k.a. unitary operators, although see Section 7 for an application to this setting). In this sense, our work is closer to a body of research on determining properties of quantum states directly, without performing full tomography (e.g. the "pretty good tomography" of Aaronson [1]). The direct detection of quantities relating to entanglement has received particular attention; see [37] for an extensive review. However, previous work has generally focused on Bell inequalities and entanglement witnesses, which are typically designed to distinguish a *particular* entangled state from any separable state. By contrast, our product test is generic and will detect entanglement in any entangled state $|\psi\rangle$.

The product test is defined in Protocol 1 below, and illustrated schematically in Figure 1. It uses as a subroutine the *swap test* for comparing quantum states [18]. This test, which can be implemented efficiently, takes two (possibly mixed) states ρ, σ of equal dimension as input. The test uses an ancilla qubit initialised in state $|0\rangle$ and applies a Hadamard gate to this qubit to produce the state $|+\rangle\langle+| \otimes \rho \otimes \sigma$. The test proceeds by applying a controlled-SWAP operation to the latter two registers, controlled by the ancilla qubit, then applies a Hadamard gate on the ancilla qubit, followed by a computational basis measurement. If the outcome is 0, the output of the test is "same"; otherwise, the output is "different". It is easy to show that this test outputs "same" with probability $\frac{1}{2} + \frac{1}{2} \operatorname{tr} \rho \sigma$.

Protocol 1 (Product test).

The product test proceeds as follows.

1. *Prepare two copies of* $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$*; call these* $|\psi_1\rangle$*,* $|\psi_2\rangle$*.*
2. *Perform the swap test on each of the* n *pairs of corresponding subsystems of* $|\psi_1\rangle$*,* $|\psi_2\rangle$*.*
3. *If all of the tests returned "same", accept. Otherwise, reject.*

The product test has appeared before in the literature. It was originally introduced in [58] as one of a family of tests for generalisations of the concurrence entanglement measure, and has been implemented experimentally as a means of detecting bipartite entanglement directly [69] (but see also [66] for caveats). Further, the test was proposed in [59] as a means of determining whether a unitary operator is product. Our contribution here is to prove the correctness of this test for all n, as formalised in the following theorem.

Theorem 1. *Given* $|\psi\rangle \in B(\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n})$*, let*

$$1 - \epsilon = \max\{|\langle\psi|\phi_1, \dots, \phi_n\rangle|^2 : |\phi_i\rangle \in B(\mathbb{C}^{d_i}), 1 \leq i \leq n\}.$$

Let $P_{test}(|\psi\rangle\langle\psi|)$ *be the probability that the product test passes when applied to* $|\psi\rangle$*. Then*

$$1 - 2\epsilon + \epsilon^2 \leq P_{test}(|\psi\rangle\langle\psi|) \leq 1 - \epsilon + \epsilon^2 + \epsilon^{3/2}.$$

For some values of ϵ *this bound is trivial, but we have the following weaker bound which applies everywhere:*

$$P_{test}(|\psi\rangle\langle\psi|) \leq 1 - \frac{11}{512}\epsilon. \tag{1}$$

More concisely, $P_{test}(|\psi\rangle\langle\psi|) = 1 - \Theta(\epsilon)$*.*

This result is essentially best possible, in a number of ways. First, we show in Section 5 that the product test itself is optimal: among all tests for product states that use two copies and have perfect completeness, the product test has optimal soundness. We also show that there cannot exist any non-trivial test that uses only one copy of the test state. Second, our analysis of the test cannot be improved too much, without introducing dependence on n and the local dimensions. When ϵ is low, we give examples of states $|\psi\rangle$ which achieve the upper and lower bounds on $P_{\text{test}}(|\psi\rangle\langle\psi|)$, up to leading order. We also give an example of a bipartite state for which ϵ is close to 1, but $P_{\text{test}}(|\psi\rangle\langle\psi|) \approx 1/2$, implying that the constant in our bound cannot be replaced with a function of ϵ that goes to 0 as ϵ approaches 1. (The bounds on this constant obtained from our proof could easily be improved somewhat, but we have not attempted to do this.) See Appendix B for all these examples. Finally, it is unlikely that a similar test could be developed for separability of *mixed* states, as the separability problem for mixed states has been shown to be NP-hard [38, 31] (and indeed we improve on this result, as discussed below).

The proof of Theorem 1 is based on relating the probability of the test passing to the action of the qudit depolarising channel. In fact, we prove a considerably more general result regarding this

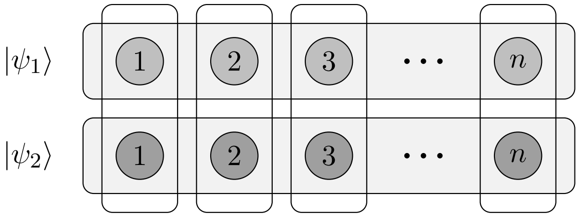

Figure 1: Schematic of the product test applied to an n-partite state $|\psi\rangle$. The swap test (vertical boxes) is applied to the n pairs of corresponding subsystems of two copies of $|\psi\rangle$ (horizontal boxes).

channel. It is known that the maximum output purity of this channel is achieved for product state inputs [5]; our result, informally, says that any state that is "close" to achieving maximum output purity must in fact be "close" to a product state. This is a *stability* result for this channel, which strengthens the previously known multiplicativity result.

Somewhat more formally, let $\mathcal{D}_\delta$ be the d-dimensional qudit depolarising channel with noise rate $1-\delta$, i.e.

$$\mathcal{D}_\delta(\rho) = (1-\delta)(\operatorname{tr}\rho)\frac{I}{d} + \delta\,\rho \tag{2}$$

for ρ a arbitrary mixed state of one d-dimensional system, and define the *O*utput *P*urity of *P*roduct states to be

$$\mathrm{OPP}(\delta) = \operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,|\phi\rangle\langle\phi|)^2 \tag{3}$$

where $|\phi\rangle$ is an arbitrary product state. Then our main result, stated more precisely as Theorem 18 in Section 6, is that for small enough $\delta > 0$, if $\operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,|\psi\rangle\langle\psi|)^2 \geq (1-\epsilon)\,\mathrm{OPP}(\delta)$, then there is a product state $|\phi_1,\dots,\phi_n\rangle$ such that $|\langle\psi|\phi_1,\dots,\phi_n\rangle|^2 \geq 1 - O(\epsilon)$.

1.2 Applications and interpretations of the product test

We describe several applications of the product test. The most important of these is that this test can be used to relate $\mathsf{QMA}(k)$ to $\mathsf{QMA}(2)$, as we will discuss in Section 3. The complexity class $\mathsf{QMA}(k)$ is defined to be the class of languages that can be decided with bounded error by a poly-time quantum verifier that receives poly-size witnesses from k unentangled provers[1] [50, 2]. To put $\mathsf{QMA}(k)$ inside $\mathsf{QMA}(2)$ with constant loss of soundness, we can have two provers simulate k provers by each submitting k unentangled proofs, whose lack of entanglement can be verified with our product test. Indeed, this gives an alternate way to understand our test as a method of using bipartite separability to certify k-partite separability.

Surprisingly, using this result as a building block also allows us to prove amplification for $\mathsf{QMA}(k)$ protocols. It has been conjectured [50, 2] that such protocols can be amplified to exponentially small soundness error. We completely resolve this conjecture, showing that $\mathsf{QMA}(k)$ protocols can be simulated in $\mathsf{QMA}(2)$ with exponentially small soundness error, and hence $\mathsf{QMA}(k) = \mathsf{QMA}(2)$ for $k \geq 2$. Indeed, we show that this result still holds if the verifier's "yes" measurement operator in a $\mathsf{QMA}(2)$ protocol is required to be separable (see Appendix C for the formal definition of this class of measurements).

As a further corollary, we can improve upon the results of [2, 12] to obtain a protocol in $\mathsf{QMA}(2)$ that verifies 3-SAT with constant soundness gap and $O(\sqrt{n}\operatorname{poly}\log(n))$ qubits (where n

[1] We assume throughout this paper that k is at most polynomial in the input size.

is the number of clauses). This in turn allows us to prove hardness of approximation results for 19 problems from quantum information theory and elsewhere which turn out to be closely related to $\mathsf{QMA}(2)$. The complete list of equivalent and related problems is given in Section 4.2; while most had previously been known, we believe that they had not been previously collected in one place.

One example of such a problem is detecting separability, or in other words the weak membership problem for $\mathrm{Sep}(d,d)$, the set of separable quantum states on $d \times d$ dimensions. It was shown in Ref. [38] that Sep cannot be approximated to precision $\exp(-d)$ in time $\mathrm{poly}(d)$ unless $\mathsf{P} = \mathsf{NP}$. In Refs. [52, 31], this result was improved to show that approximating Sep to precision $1/\mathrm{poly}(d)$ is similarly NP-hard. We show that there is a universal constant $\delta > 0$ such that, if K is a convex set that approximates SEP to within trace distance δ, then membership in K cannot be decided in polynomial time unless 3-SAT $\in \mathsf{DTIME}(\exp(\sqrt{n}\log^{O(1)}(n)))$. Other such problems for which we can prove that no polynomial-time algorithm exists, under the same assumption about the hardness of 3-SAT, are estimating the minimum output entropy of a quantum channel up to a constant, and estimating the ground-state energy of quantum systems under a mean-field approximation.

We also prove hardness results for some tensor optimisation problems which are not apparently related directly to quantum information theory, examples of which include approximating the injective tensor norm of 3-index tensors and estimating the $\ell_2 \to \ell_4$ norm of a matrix. Our proof that amplification of $\mathsf{QMA}(2)$ protocols is possible implies that one can derive stronger hardness results for all of these tasks, if one is willing to make stronger assumptions about the hardness of 3-SAT.

Our final application is that the product test can be used to determine whether a unitary operator is a tensor product or far from a tensor product in the Hilbert-Schmidt norm, promised that one of these is the case. This can be seen [59] as one possible generalisation of the well-studied problem of testing whether a boolean function $\{0,1\}^n \to \{0,1\}$ is linear [13]. This application is described in Section 7.

These different applications of the product test reflect the many different interpretations of $P_{\text{test}}(|\psi\rangle\langle\psi|)$. It is related in a precise sense to

- The purity of $|\psi\rangle$ after it is subjected to independent depolarising noise (see Appendix A).
- The maximum overlap of $|\psi\rangle$ with any product state (proved in Appendix B). The logarithm of this maximum overlap is an important entanglement measure known as the geometric measure of entanglement (see [71] and references therein).
- The overlap of $|\psi\rangle^{\otimes 2}$ with the tensor product of the symmetric subspaces of $\mathbb{C}^{d_1}\otimes\mathbb{C}^{d_1},\dots,\mathbb{C}^{d_n}\otimes\mathbb{C}^{d_n}$ (discussed in Section 5).
- The average overlap of $|\psi\rangle$ with a *random* product state, and a quantum variant of the Gowers uniformity norm [33] (discussed in Appendix G).
- The average purity of $|\psi\rangle$ across a random partition of $[n]$ into two subsets (also discussed in Appendix G).

1.3 Implications for classical computer science

The main result of our paper proposes a quantum solution to a quantum problem. Nevertheless, there are several implications of our product test that may be of interest to classical computing. Instead of viewing our results as concerning entangled states of many systems, they may be interpreted in terms of tensors with many indices. These tensors have been studied in the context of image processing [4], the planted clique problem [17], constraint satisfaction problems [25] and many other settings [67].

In this language, our results in Section 4 imply that many central tensor problems, such as the injective tensor norm (defined in Section 4.2), are hard to approximate even to within constant factors. On the positive side, our Theorem 11 (together with the equivalences in Section 4.2) implies that if a heuristic or approximation algorithm existed to optimise over trilinear forms, it could be extended with little loss of accuracy, to perform optimisations over k-linear forms for general k. These connections have been further explored in [8], which shows that the $\ell_2 \to \ell_4$ norm of a matrix is hard to approximate, and connects this problem to the small-set expansion problem.

1.4 Related work

Our paper addresses a central problem in multipartite entanglement, which is too vast a field to reasonably summarise here (one good recent survey is [44]). We therefore concentrate on reviewing work on quantum Merlin-Arthur games with multiple provers.

The class $\mathsf{QMA}(k)$ was first introduced by Kobayashi, Matsumoto, and Yamakami [50], who showed that amplification of the soundness gap of $\mathsf{QMA}(2)$ protocols implies that $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ (a result proven independently in [2]). Both these papers also showed that it is possible to amplify *completeness* to exponentially close to 1 (see Lemma 8 for a restatement). Blier and Tapp showed [12] that graph 3-colourability can be decided using a $\mathsf{QMA}(2)$ protocol with messages of length $O(\log n)$ qubits and soundness $1 - \Omega(1/n^6)$. This soundness gap was improved to $\Omega(1/n^{3+\epsilon})$ by Beigi [10], and has recently been improved again to $\Omega(1/(n \operatorname{polylog} n))$ by Le Gall, Nakagawa and Nishimura [51]. By contrast, Aaronson et al. showed [2] that 3-SAT can be solved by a $\mathsf{QMA}(k)$ protocol with *constant* soundness, at the expense of increasing k to $O(\sqrt{n} \operatorname{polylog}(n))$. Finally, Liu, Christandl and Verstraete have given a problem in $\mathsf{QMA}(2)$ which is not obviously in QMA [53].

Following the conference and arXiv versions of this work, there have been several interesting developments related to $\mathsf{QMA}(k)$. First, it has been shown by Brandão, Christandl and Yard [15, 16] that $\mathsf{QMA}(k)$ protocols, for constant k, are no stronger than QMA protocols if the verifier's measurement is restricted to be LOCC (implementable via local operations and classical communication). One consequence of their work is that, if there existed an efficient LOCC product state test, $\mathsf{QMA}(k) = \mathsf{QMA}$. However, we show in Appendix D that no such test can exist. In the same work, the authors give a subexponential-time algorithm for optimizing over the set of separable states [16]; an alternative algorithm for this task has been given by Shi and Wu [64], who also prove that several special cases of $\mathsf{QMA}(2)$ protocols can be simulated in polynomial space.

On the other hand, Chen and Drucker [21] have improved on the result of [2] and given an LOCC $\mathsf{QMA}(k)$ protocol that verifies 3-SAT with constant soundness gap for $k = O(\sqrt{n} \operatorname{poly}\log(n))$. (In fact, their protocol fits in the more restrictive class known as $\mathsf{BellQMA}(k)$.) Chiesa and Forbes [22] recently gave a tight soundness analysis of this protocol, showing that the soundness gap increases smoothly with k.

McKague has recently used one of our results (that the verifier's "yes" measurement operator may be taken to be separable) to prove that restricting the class $\mathsf{QMA}(2)$ to real Hilbert space does not change its computational power [57]. While it is natural to expect the real case to behave similarly to the complex case, we note that even for states with real coefficients the closest product state may be complex [23]. Finally, another application of our results was found by [20], who have presented the only known nontrivial $\mathsf{QMA}(2)$-complete problem: estimating the minimum energy of a sparse Hamiltonian over all bipartite product states.

1.5 Organisation

The remainder of this paper is organised as follows. In Section 2, we give an overview of the proofs of our main results (details are in Appendices A and B). In Section 3, we apply the product test to prove that $\mathsf{QMA}(k) = \mathsf{QMA}(2)$ for $k \geq 2$, and we give some complexity-theoretic applications of this result in Section 4, including an extensive discussion of problems related to $\mathsf{QMA}(2)$. In Section 5 we argue that the product test is essentially optimal, and in Section 6 we state our results for the depolarising channel. We discuss the use of the product test to test product unitaries in Section 7, and finish with some open questions in Section 8.

1.6 Notation

For a vector space V, define $B(V)$ to be the unit vectors in V, $L(V)$ to be the linear operators from V to V, and $\mathcal{B}(V)$ to be the density operators on V. More concisely, let $\mathcal{B}(d)$ denote the set of $d \times d$ density matrices. If $|\psi\rangle$ is a vector, let $\psi := |\psi\rangle\langle\psi|$. Define the set of separable states on $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ to be

$$\mathrm{Sep}(d_A, d_B) := \mathrm{conv}\{\alpha \otimes \beta : |\alpha\rangle \in B(\mathbb{C}^{d_A}), |\beta\rangle \in B(\mathbb{C}^{d_B})\}, \tag{4}$$

where $\mathrm{conv}(S)$ denotes the convex closure of a set S. The swap operator on $\mathbb{C}^d \otimes \mathbb{C}^d$ is denoted $\mathcal{F}$, and is defined to be $\sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |j\rangle\langle i|$.

For $\alpha \geq 1$, let $\|M\|_\alpha$ denote the Schatten α-norm of a matrix: $\mathrm{tr}(|M|^\alpha)^{1/\alpha}$. For a density matrix $\rho \in \mathcal{B}(\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n})$, and $S \subseteq \{1, \ldots, n\}$, ρ_S denotes the density matrix obtained by tracing out (discarding) the systems not in S. To avoid excessive parenthesization, we write $\rho_S^2 := (\rho_S)^2$ and $\mathrm{tr}\,\rho^2 := \mathrm{tr}(\rho^2)$.

2 Overview of the proof of correctness

In this section, we sketch the proof of Theorem 1; a full proof is given in Appendices A and B.

We discuss here only the upper bound on P_{test}, since the lower bound follows from continuity and the fact that product states pass the test with probability 1. First, we make precise the intuition that the product test is likely to pass precisely when the average subsystem is close to pure.

Lemma 2. *Let $P_{test}(\rho, \sigma)$ denote the probability that the product test passes when applied to two mixed states $\rho, \sigma \in \mathcal{B}(\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n})$. Define $P_{test}(\rho) := P_{test}(\rho, \rho)$. Then*

$$P_{test}(\rho, \sigma) = \frac{1}{2^n} \sum_{S \subseteq [n]} \mathrm{tr}\,\rho_S \sigma_S,$$

and in particular

$$P_{test}(\rho) = \frac{1}{2^n} \sum_{S \subseteq [n]} \mathrm{tr}\,\rho_S^2.$$

The proof itself is split into two parts, beginning with the case where ϵ is low. We write $|\psi\rangle = \sqrt{1-\epsilon}|0^n\rangle + \sqrt{\epsilon}|\phi\rangle$ without loss of generality, for some product state $|0^n\rangle$ and arbitrary state $|\phi\rangle$. This allows an explicit expression for $\operatorname{tr}\psi_S^2$ in terms of ϵ and $|\phi\rangle$ to be obtained. While the marginals of $|\phi\rangle$ can be complicated (and worse, we need to consider products of expressions of the form $\operatorname{tr}_{\bar{S}} |0\rangle\langle\phi|$), we can simplify things by only considering the reductions in $\operatorname{tr}\psi_S^2$ that occur when $|0\rangle$ is combined with a state orthogonal to $|0\rangle$. Thus, we do not need a detailed picture of $|\phi\rangle$, but instead will merely split it into a superposition of strings with different Hamming weight (i.e. number of positions orthogonal to $|0\rangle$). Intuitively, the contribution to $\mathbb{E}_{S\subseteq[n]} \operatorname{tr}\psi_S^2$ of a piece of $|\phi\rangle$ with Hamming weight k should be exponentially small in k, since each position that differs from 0 leads to a constant reduction in weight when we project onto the symmetric subspace. In order to obtain a non-trivial bound from this expression, the final stage of this part of the proof is to use the fact that $|0^n\rangle$ is the closest product state to $|\psi\rangle$ to argue that $|\phi\rangle$ cannot have any amplitude on basis states of Hamming weight 0 or 1. Ruling out basis states of Hamming weight 0 (i.e. $|0^n\rangle$) is obvious, since otherwise ϵ would be smaller. Less obvious is that $|\phi\rangle$ cannot have any amplitude on Hamming weight-1 states, but this too is contradicted by the fact that $|0^n\rangle$ has overlap with $|\phi\rangle$ that is a local maximum among product states, and nonzero amplitude on weight-1 states would mean an infinitesimal local rotation could reduce ϵ. As a result, we obtain a bound that is applicable when ϵ is small.

Theorem 3. *Given $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, let*

$$1 - \epsilon = \max\{|\langle\psi|\phi_1,\dots,\phi_n\rangle|^2 : |\phi_i\rangle \in \mathbb{C}^{d_i}, 1 \le i \le n\}.$$

Then $1 - 2\epsilon + \epsilon^2 \le P_{test}(|\psi\rangle\langle\psi|) \le 1 - \epsilon + \epsilon^2 + \epsilon^{3/2}$.

In the case where ϵ is high, this result does not yet give a useful upper bound. In the second part of the proof, we derive a constant bound on $P_{\text{test}}(|\psi\rangle\langle\psi|)$ based on considering $|\psi\rangle$ as a k-partite state, for some $k < n$. $P_{\text{test}}(|\psi\rangle\langle\psi|)$ can be shown to be upper bounded by the probability that the test for being product across any partition into k parties passes. Informally speaking, if $|\psi\rangle$ is far from product across the n subsystems, we show that one can find a partition such that the distance from the closest product state (with respect to this partition) falls into the regime where the first part of the proof works.

Theorem 4. *Given $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, let*

$$1 - \epsilon = \max\{|\langle\psi|\phi_1,\dots,\phi_n\rangle|^2 : |\phi_i\rangle \in \mathbb{C}^{d_i}, 1 \le i \le n\}.$$

Then, if $\epsilon \ge 11/32 > 0.343$, $P_{test}(|\psi\rangle\langle\psi|) \le 501/512 < 0.979$.

Between them, Theorems 3 and 4 imply Theorem 1. In fact, we can say precisely that $P_{\text{test}}(|\psi\rangle\langle\psi|) = 1 - c(\psi)\epsilon$ for $\frac{11}{512} \le c(\psi) \le 2$.

One feature of our proof that can be generalised is the expectation over $S \subseteq [n]$. We effectively choose S by flipping a fair coin, but if we use a biased coin then this has an interesting alternate interpretation in terms of the output purity of the depolarising channel. This yields a similar result, which is not only that product states maximise the output purity (as was previously known), but that any state which even approximately maximises the output purity must be approximately product. See Section 6 for a precise statement. Since the correctness of the product test is a special case of this more general theorem, we first prove the result about depolarising channels in Appendix A and then complete the details necessary for the product test in Appendix B.

This completes the overview of the proof; we now discuss some applications of the product test.

3 $\mathsf{QMA}(2)$ vs. $\mathsf{QMA}(k)$

In this section, we apply the product test to a problem in quantum complexity theory: whether k unentangled provers are stronger than 2 unentangled provers. This question can be formalised as whether the complexity classes $\mathsf{QMA}(k)$ and $\mathsf{QMA}(2)$ are equal [50, 2]. These classes are defined as follows.

Definition 1. *A language L is in $\mathsf{QMA}(k)_{s,c}$ if there exists a polynomial-time quantum algorithm $\mathcal{A}$ such that, for all inputs $x \in \{0,1\}^n$:*

1. **Completeness:** *If $x \in L$, there exist k witnesses $|\psi_1\rangle, \dots, |\psi_k\rangle$, each a state of $\operatorname{poly}(n)$ qubits, such that $\mathcal{A}$ outputs "accept" with probability at least c on input $|x\rangle|\psi_1\rangle \dots |\psi_k\rangle$.*
2. **Soundness:** *If $x \notin L$, then $\mathcal{A}$ outputs "accept" with probability at most s on input $|x\rangle|\psi_1\rangle \dots |\psi_k\rangle$, for all states $|\psi_1\rangle, \dots, |\psi_k\rangle$.*

We use $\mathsf{QMA}(k)$ as shorthand for $\mathsf{QMA}(k)_{1/3,2/3}$, and QMA as shorthand for $\mathsf{QMA}(1)$. We always assume $1 \le k \le \operatorname{poly}(n)$.

We also define $\mathsf{QMA}_m(k)_{s,c}$ to indicate that $|\psi_1\rangle, \dots, |\psi_k\rangle$ each involve m qubits, where m may be a function of n other than $\operatorname{poly}(n)$.

Two of the major open problems related to $\mathsf{QMA}(k)_{s,c}$ are to determine how the size of the complexity class depends on k and on s, c. It has been conjectured for some years [50, 2] that in fact $\mathsf{QMA}(k) = \mathsf{QMA}(2)$ for $2 \le k \le \operatorname{poly}(n)$, and that the soundness and completeness can be amplified by parallel repetition in a way similar to BPP, BQP, MA, QMA and other complexity classes with bounded error. In fact, these conjectures are related: $2k$ independent provers can simulate k independent realisations of a $\mathsf{QMA}(2)$ protocol in order to amplify the soundness-completeness gap, and conversely, [50, 2] proved that $\mathsf{QMA}(2)$ amplification implies that $\mathsf{QMA}(2) = \mathsf{QMA}(\mathrm{poly})$. In this section, we will fully resolve these conjectures, proving that $\mathsf{QMA}(2) = \mathsf{QMA}(\mathrm{poly})$ and that $\mathsf{QMA}(k)$ can have its soundness and completeness amplified by a suitable protocol.

The most direct way of putting $\mathsf{QMA}(k)$ inside $\mathsf{QMA}(2)$ is to ask two provers to each send the k unentangled proofs that correspond to a $\mathsf{QMA}(k)$ protocol. If $k = \operatorname{poly}(n)$, then each prover is still sending only polynomially many qubits. Then the product test can be used to verify that the states sent were indeed product states and can be used as valid inputs to a $\mathsf{QMA}(k)$ protocol. The specific protocol is described in Protocol 2.

First observe that for YES instances (instances in the language), k Merlins can achieve success probability $\ge c$, so by sending two copies of this optimal state, two Merlins can achieve completeness $\ge \frac{1+c}{2} \ge c$ in this modified protocol.

Now consider NO instances. Assume for now that the two Merlins always send the same state. Then according to Theorem 1, if the Merlins send states that are far from product, they are likely to fail the product test, whereas basic continuity arguments can show that if they send states that are nearly product then the success probability will not be much larger than the soundness of the original protocol. Thus, the soundness does not become too much worse. These ideas (with a detailed proof in Appendix E) establish

Lemma 5. *For any m, k, $0 \le s < c \le 1$,*

$$\mathsf{QMA}_m(k)_{s,c} \subseteq \mathsf{QMA}_{km}(2)_{s',c'}$$

where $c' = \frac{1+c}{2}$ and $s' = 1 - \frac{(1-s)^2}{100}$.

Protocol 2 ($\mathsf{QMA}(k)$ **to** $\mathsf{QMA}(2)$)**.**

The $\mathsf{QMA}(2)$ *protocol proceeds as follows.*

1. *Each of the two Merlins sends* $|\psi\rangle := |\psi_1\rangle \otimes \ldots \otimes |\psi_k\rangle$ *to Arthur.*
2. *Arthur performs each of the following tests with probability* $1/2$*.*
   - (a) *Arthur runs the product test with the two states as input and accepts iff the test outputs "product."*
   - (b) *Arthur randomly chooses one of the states from the two Merlins, runs the algorithm* $\mathcal{A}$ *on that state, and outputs the result.*

This is already strong enough to achieve amplification up to constant soundness. However, Protocol 2 has a salutary side effect that will allow us to achieve stronger amplification. To see this, we will first introduce a further generalisation of the $\mathsf{QMA}(k)$ family. Let $\mathbb{M}$ be a set of Hermitian operators M satisfying $0 \leq M \leq I$. Each $M \in \mathbb{M}$ defines a binary measurement with M corresponding to the "accept" outcome and $I - M$ corresponding to the "reject" outcome. Variants of $\mathsf{QMA}(2)$ have been considered in which M is not only restricted to be efficiently implementable on a quantum computer, but also with the further restriction that it belongs to some set $\mathbb{M}$. We will consider standard classes of measurements such as BELL, LOCC, SEP, ALL, etc., whose definitions we include in Appendix C.

Formally, define $\mathsf{QMA}^{\mathbb{M}}_m(k)_{s,c}$ to be the class $\mathsf{QMA}_m(k)_{s,c}$ with Arthur restricted to performing measurements from $\mathbb{M}$ in addition to being restricted to quantum polynomial time. For example, $\mathsf{QMA}^{\mathrm{BELL}}(k)$ has been introduced under the name $\mathsf{BellQMA}(k)$ and proven equal to QMA (for constant k) by Brandão [14, 2]. Our paper will focus on the case that $\mathbb{M} = \mathrm{SEP}$, the class of measurements such that M is a separable operator. Observe that $M \in \mathrm{SEP}$ does not imply that $I - M$ is separable.

Armed with the definition of $\mathsf{QMA}^{\mathrm{SEP}}$, we can now see that Protocol 2 produces a protocol that is not only in $\mathsf{QMA}(2)$, but also $\mathsf{QMA}^{\mathrm{SEP}}(2)$. More formally, we can strengthen Lemma 5 to:

Lemma 6. *For any* m, k, $0 \leq s < c \leq 1$,

$$\mathsf{QMA}_m(k)_{s,c} \subseteq \mathsf{QMA}^{\mathrm{SEP}}_{km}(2)_{s',c'}$$

where $c' = \frac{1+c}{2}$ *and* $s' = 1 - \frac{(1-s)^2}{100}$.

Proof. We again use Protocol 2. By Lemma 5, we know that this protocol has completeness $c' = (1 + c)/2$, and has soundness $s' = 1 - \Omega((1 - s)^2)$. It remains only to argue that the "accept" measurement outcome is a separable operator.

Suppose the first Merlin sends systems $A_1, \ldots, A_k$ and the second Merlin sends systems $B_1, \ldots, B_k$. The "accept" outcome of the product test corresponds to the tensor product of projectors onto the symmetric subspaces of $A_1B_1, A_2B_2, \ldots, A_kB_k$. Since the symmetric subspace is spanned by vectors of the form $|\psi\rangle^{\otimes 2}$, it follows that the projector onto each symmetric subspace is separable across the $A : B$ cut, and in turn that their tensor product is as well. The other test in the protocol is to simply apply a measurement either entirely on $A_1, \ldots, A_k$ or entirely on $B_1, \ldots, B_k$, which is automatically separable. Finally, performing a probabilistic mixture of separable measurements creates a composite measurement which is also separable. □

The advantage of $\mathsf{QMA}^{\mathrm{SEP}}(k)$ is that it removes the chief difficulty with $\mathsf{QMA}(k)$ amplification, which is that conditioning on measurement outcomes can induce entanglement between systems we have not yet measured. This phenomenon is known as entanglement swapping [45]. However, if we condition on the outcome of a measurement being M, for some $M \in \mathrm{SEP}$, then no entanglement will be produced in the unmeasured states. As a result, cheating provers cannot gain any advantage by sending entangled proofs, and we obtain the following lemma.

Lemma 7. *For any $\ell \geq 1$,*

$$\mathsf{QMA}_m^{\mathrm{SEP}}(k)_{s,c} \subseteq \mathsf{QMA}_{\ell m}^{\mathrm{SEP}}(k)_{s^\ell,c^\ell}.$$

The idea is to simply repeat the original protocol ℓ times in parallel and to accept iff each subprotocol accepts. Since we are considering $\mathsf{QMA}^{\mathrm{SEP}}$ protocols, obtaining an "accept" outcome on one proof will not induce any entanglement on the remaining proofs. We give a detailed proof of Lemma 7 in Appendix E.

From Lemma 6 and Lemma 7, we can almost conclude that strong amplification is possible. Indeed, when we start with protocols with perfect completeness, we can apply Protocol 2, repeat $p(n)$ times, and reduce the soundness from s to $s^{O(p(n))}$. For the case of $c < 1$, we need one additional argument to keep the completeness from being reduced too much at the same time. Here we will use a method for completeness amplification proved in both [50, Lemma 5] and [2, Lemma 6].

Lemma 8 ([50, 2])**.** *For any $\ell \geq 1$,*

$$\mathsf{QMA}_m(k)_{s,c} \subseteq \mathsf{QMA}_{\ell m}(k)_{1-\frac{c-s}{3},1-\exp(-\frac{\ell(c-s)^2}{2})}.$$

Our amplification procedure for general $c < 1$ is then to

1. Use Lemma 8 to bring the completeness exponentially close to 1.

2. Use Lemma 6 to convert a general $\mathsf{QMA}(k)$ protocol to a $\mathsf{QMA}^{\mathrm{SEP}}(2)$ protocol.

3. Repeat the protocol polynomially many times to make the soundness exponentially small.

This procedure then achieves

Theorem 9. *1. If $s \leq 1 - 1/\operatorname{poly}(n)$, $k = \operatorname{poly}(n)$ and $p(n)$ is an arbitrary polynomial, then $\mathsf{QMA}(k)_{s,1} = \mathsf{QMA}^{\mathrm{SEP}}(2)_{\exp(-p(n)),1}$.*

2. If $c - s \geq 1/\operatorname{poly}(n)$, $c < 1$, $k = \operatorname{poly}(n)$ and $p(n)$ is an arbitrary polynomial, then $\mathsf{QMA}(k)_{s,c} = \mathsf{QMA}^{\mathrm{SEP}}(2)_{\exp(-p(n)),1-\exp(-p(n))}$.

We prove correctness of Protocol 2 and the rest of Theorem 9 in Appendix E. There are obvious variants of Theorem 9 to cover the case of limited message size, whose statements we leave implicit.

3.1 $\mathsf{QMA}(2)$ and h_{Sep}

The complexity of $\mathsf{QMA}_m(2)$ stems both from the complexity of producing the measurement made by the verifier, and of maximising its acceptance probability over product states. To understand the complexity of this second step, we define the support function of the separable states to be

$$h_{\mathrm{Sep}(d,d)}(M) := \max\{\operatorname{tr} M\rho : \rho \in \mathrm{Sep}(d,d)\} = \max\{\operatorname{tr} M(\alpha \otimes \beta) : |\alpha\rangle, |\beta\rangle \in B(\mathbb{C}^d)\}, \qquad (5)$$

for any $M \in L(\mathbb{C}^d \otimes \mathbb{C}^d)$. Calculating h_{Sep} up to $1/\operatorname{poly}(d)$ accuracy was proven to be NP-hard by Gurvits [38]. One of our central results will be a weaker hardness result for the problem of estimating h_{Sep} to *constant* additive error; see Theorem 11 below.

Thus, $\mathsf{QMA}_m(2)_{s,c}$ is the class of languages that can be decided by determining whether $h_{\mathrm{Sep}(2^m,2^m)}(M)$ is $\geq c$ or $\leq s$, where M is a measurement operator that can be constructed in polynomial time on a quantum computer. It is instructive to compare the case of $\mathsf{QMA}_m(1)$, where the problem can be thought of as computing the largest eigenvalue of a $2^m \times 2^m$ matrix. There the hardness comes from the fact that the matrix is implicitly specified by a polynomial-size quantum circuit. By contrast, in the case of $\mathsf{QMA}(2)$, there is no known $\operatorname{poly}(d)$-time algorithm to compute $h_{\mathrm{Sep}(d,d)}(M)$. As a result, $\mathsf{QMA}_{\log}(1) = \mathsf{BQP}$ [55], but $\mathsf{QMA}_{\log}(2)$ is not known to be in BQP (since we do not know how to search over unentangled pairs of $\log(n)$-qubit states in quantum polynomial time) or NP (since the measurement can depend on a general quantum poly-time algorithm). The weakest class that we know contains $\mathsf{QMA}_{\log}(2)$ is $\mathsf{NP}^{\mathsf{BQP}}$, by using the BQP oracle to obtain an explicit description of M. This can be achieved up to error ϵ by running the verifier's circuit $\operatorname{poly}(2^m, 1/\epsilon)$ times and performing tomography. We therefore obtain that $\mathsf{QMA}_m(2)_{s,c} \subseteq \mathsf{NTIME}(\operatorname{poly}(2^m, n, 1/(c-s)))^{\mathsf{BQP}}$. In particular, $\mathsf{QMA}(2) \subseteq \mathsf{NEXP}$. Unfortunately this cannot be scaled down to place $\mathsf{QMA}_{\log}(2)$ in NP. This is because the verifier in a $\mathsf{QMA}_{\log}(2)$ protocol still can perform a poly-time quantum computation. Thus, we only have that $\mathsf{QMA}_{\log}(2) \subseteq \mathsf{NP}^{\mathsf{BQP}}$.

Via the connection between $\mathsf{QMA}_m(2)$ and $h_{\mathrm{Sep}(2^m,2^m)}$, all of the results in this section can be stated in terms of h_{Sep}. In particular, we have the following variants of Lemma 7 and Theorem 9.

Lemma 10. *Let $M \in SEP$ be a $d^2 \times d^2$-dimensional separable Hermitian matrix satisfying $0 \leq M \leq I$. Then $h_{\mathrm{Sep}(d^k,d^k)}(M^{\otimes k}) = h_{\mathrm{Sep}}(M)^k$.*

Theorem 11. *Let M be a $d^2 \times d^2$-dimensional Hermitian matrix satisfying $0 \leq M \leq I$. Assume that we are promised that $h_{\mathrm{Sep}(d,d)}(M)$ is either $\geq c$ or $\leq s$ for $1 \geq c > s > 0$. Call these two cases "Y" and "N." Choose $1 \geq c' > s' > 0$ such that $c' = 1$ if and only if $c = 1$. Then there exists a matrix M' of size d^k such that*

$$h_{\mathrm{Sep}(d^k,d^k)}(M') \begin{cases} \geq c' & \textit{in case } Y \\ \leq s' & \textit{in case } N \end{cases} \tag{6}$$

If $c = 1$, then $k = O((1-s)^{-2}\log(1/s'))$, and if $c < 1$, then $k = O((c-s)^{-3}\log(1/(1-c'))\log^2(1/s'))$. Additionally $M' \in \mathrm{SEP}$, and M' can be constructed efficiently from M, even by a classical log-space transducer.

4 Complexity-theoretic implications

4.1 Evidence for the hardness of $\mathsf{QMA}_{\log}(2)$

A key application of Theorem 9 is to the protocol of Ref. [2] that puts 3-SAT on n clauses inside the complexity class $\mathsf{QMA}_{\log(n)}(\sqrt{n}\operatorname{poly}\log(n))_{1-\Omega(1),1}$. Applying Theorem 9 lets us simulate this using two provers with perfect completeness and arbitrary soundness, so that we obtain

Corollary 12. *Let $\ell : \mathbb{N} \to \mathbb{N}$ be polynomially bounded. Then*

$$\textit{3-SAT} \in \mathsf{QMA}_{\ell(n)\sqrt{n}\operatorname{poly}\log(n)}(2)_{2^{-\ell(n)},1}.$$

In other words, there is a protocol for 3-SAT instances with n clauses that uses two provers, $\ell(n)\sqrt{n}\operatorname{poly}\log(n)$-qubit proofs and has perfect completeness and soundness $2^{-\ell(n)}$.

Therefore, making assumptions about the hardness of 3-SAT allows us to prove hardness results for the complexity class $\mathsf{QMA}_{\log}(2)$, and stronger assumptions naturally imply stronger hardness results. We formalise this correspondence as the following corollary.

Corollary 13. *The following implications hold.*

(i) If 3-SAT on n clauses is not in $\mathsf{DTIME}(\exp(o(n)))$, then for arbitrary constant $\epsilon > 0$

$$\mathsf{QMA}_{\log(d)}(2)_{\frac{1}{2},1} \not\subseteq \mathsf{DTIME}(d^{\log^{1-\epsilon} d}).$$

(ii) If 3-SAT on n clauses is not in $\mathsf{DTIME}(\exp(o(n)))$, then

$$\mathsf{QMA}_{\log(d)}(2)_{2^{-\sqrt{\log d}/\operatorname{polylog}(\log d)},1} \not\subseteq \mathsf{DTIME}(\operatorname{poly}(d)).$$

(iii) If 3-SAT on n clauses is not in $\mathsf{DTIME}(\exp(\sqrt{n}\operatorname{polylog}(n)))$, then

$$\mathsf{QMA}_{\log(d)}(2)_{\frac{1}{2},1} \not\subseteq \mathsf{DTIME}(\operatorname{poly}(d)).$$

More generally, assume that for some functions $\ell, m : \mathbb{N} \to \mathbb{N}$, 3-SAT on n clauses is not contained in $\mathsf{DTIME}(m(\exp(\ell(n)\sqrt{n}\operatorname{polylog}(n))))$. Then, defining $d = 2^{\ell(n)\sqrt{n}\operatorname{polylog}(n)}$,

$$\mathsf{QMA}_{\log(d)}(2)_{2^{-\ell(n)},1} \not\subseteq \mathsf{DTIME}(m(d)).$$

Note that the assumptions on the hardness of 3-SAT made in the first two cases are essentially equivalent to the (not implausible) *Exponential Time Hypothesis* of Impagliazzo and Paturi [46], which states that 3-SAT $\notin \mathsf{DTIME}(\exp(\ell(n)))$ for any $\ell(n) = o(n)$.

4.2 $\mathsf{QMA}_{\log}(2)$ equivalences and reductions

We now discuss a number of problems for which Corollary 13 allows us to prove hardness results. As described in Section 3.1, $\mathsf{QMA}_{\log(d)}(2)$ is intimately connected to $h_{\mathrm{Sep}(d,d)}$. Here we focus solely on the hardness of estimating $h_{\mathrm{Sep}(d,d)}(M)$ when $0 \le M \le I$ is given explicitly as input. In other words, we will examine the part of the hardness of $\mathsf{QMA}_{\log}(2)$ that does *not* come from having access to a poly-time quantum computation.

One definition we will repeatedly use is that of the *weak membership problem.* If K is a convex set, $\epsilon > 0$ and d is a metric, then $\mathrm{WMEM}_\epsilon^{(d)}(K)$ denotes the following task: given an input x, determine whether $x \in K$ or $d(x, K) \ge \epsilon$, given the promise that one of these conditions holds. Here $d(x, K) := \inf_{y\in K} d(x, y)$. The reason for the ϵ is because the complexity of the problem can depend on the required precision, just as the size of $\mathsf{QMA}(k)_{s,c}$ depends on how close s and c are. See [35] for more background and equivalent formulations of the weak membership problem for convex sets. In many cases, d will be the trace norm distance; in this case, we will simply write $\mathrm{WMEM}_\epsilon(K)$ for the weak membership problem. We also define $B_d(K, \epsilon) := \{x : d(x, K) \le \epsilon\}$, and define the Hausdorff distance between (not necessarily convex) sets K, L to be $d_H(K, L) := \max\{\sup_{x\in K} d(x, L), \sup_{x\in L} d(x, K)\}$, or equivalently, $\inf\{\epsilon \ge 0 : X \subseteq B_d(Y, \epsilon) \text{ and } Y \subseteq B_d(X, \epsilon)\}$.

The following equivalences and reductions are a combination of known results ([72, 38, 28, 56, 29, 52, 17, 32, 8], and some unpublished and/or folklore) and consequences of our main theorems.

Even though many of the reductions are straightforward, we are not aware of any similar list in the literature, despite many of the quantities being discussed individually.

Equivalent problems

1. Given M with $0 \leq M \leq I$, determine whether

   $$h_{\mathrm{Sep}}(M) := \max_{\rho \in \mathrm{Sep}(d,d)} \operatorname{tr} M\rho = \max_{|\alpha\rangle, |\beta\rangle \in B(\mathbb{C}^d)} \operatorname{tr} M(\alpha \otimes \beta) \tag{7}$$

   is $\geq c$ or $\leq s$. As discussed above, this represents the acceptance probability of a $\mathsf{QMA}(2)$ protocol when the measurement is fixed and the provers use an optimal strategy.

   This is our reference problem, and we will compare the problems below to this one. However, we observe that this problem is equivalent (up to a polynomial change of dimension described below) to versions with different choices of c and s as long as $0 < s < c < 1$ are constants independent of dimension.

2. Define $\mathrm{ProdSym}(d) := \mathrm{conv}\{\psi \otimes \psi : |\psi\rangle \in B(\mathbb{C}^d)\}$. Given M with $0 \leq M \leq I$, determine whether $h_{\mathrm{ProdSym}(d)}(M)$ is $\geq c/4$ or $\leq s/4$.

3. Define $\mathrm{SepSym}(d) := \mathrm{conv}\{\rho \otimes \rho : \rho \in \mathcal{B}(d)\}$. Given M with $0 \leq M \leq I$, determine whether $h_{\mathrm{SepSym}(d)}(M)$ is $\geq c/4$ or $\leq s/4$.

4. The set EW:=EW(d,d) of *entanglement witnesses* [65] is the dual cone of Sep, meaning that

   $$\mathrm{EW}(d,d) = \{W : \operatorname{tr} W\rho \geq 0, \forall \rho \in \mathrm{Sep}(d,d)\}. \tag{8}$$

   Given M, determine whether $\min\{\|M + W\| : W \in \mathrm{EW}(d,d)\}$ is $\geq c$ or $\leq s$.

5. For a quantum channel $\mathcal{N}$, determine whether the superoperator $1 \to \infty$ norm $\|\mathcal{N}\|_{1\to\infty}$ is $\geq c$ or $\leq s$, where $\|\mathcal{N}\|_{1\to\infty} := \max_\rho \frac{\|\mathcal{N}(\rho)\|_\infty}{\|\rho\|_1}$.

6. For a quantum channel $\mathcal{N}$, determine whether the superoperator $1 \to 2$ norm $\|\mathcal{N}\|_{1\to 2}$ is $\geq 4c-3$ or $\leq 4s-3$, where $\|\mathcal{N}\|_{1\to 2} := \max_\rho \frac{\|\mathcal{N}(\rho)\|_2}{\|\rho\|_1}$. (This equivalence is only nontrivial for some values of c, s.)

7. Given $\mathcal{N}$, determine whether the minimum output Rènyi entropy $S_\infty^{\min}(\mathcal{N})$ is $\geq \log(1/s)$ or $\leq \log(1/c)$. Here $S_\infty^{\min} := \min_\rho S_\infty(\rho)$, where $S_\infty(\sigma) := -\log\|\sigma\|_\infty$.

8. Given $\mathcal{N}$, determine whether the minimum output Rènyi entropy $S_2^{\min}(\mathcal{N})$ is $\geq \log(2/\sqrt{s})$ or $\leq \log(2/\sqrt{c})$. Here $S_2^{\min} := \min_\rho S_2(\rho)$, where $S_2(\sigma) := -\log\|\sigma\|_2$.

9. Given a 3-index tensor $T \in \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^d$, determine whether the injective tensor norm $\|T\|_{\mathrm{inj}}$ is $\geq \sqrt{c}$ or $\leq \sqrt{s}$. The injective tensor norm is defined here to be

   $$\|T\|_{\mathrm{inj}} = \max_{x,y,z \in B(\mathbb{C}^d)} |\langle T| \cdot |x\rangle \otimes |y\rangle \otimes |z\rangle| \tag{9}$$

   and T should have the property that for some choice of indices it can be interpreted as a linear map from $\mathbb{C}^d \to \mathbb{C}^{d^2}$ with operator norm ≤ 1.

10. Given a linear map T from $\mathbb{C}^d$ to $L(\mathbb{C}^d)$ with operator norm ≤ 1, determine whether $\|T\|_{\ell_2 \to S_\infty}$ is $\geq \sqrt{c}$ or $\leq \sqrt{s}$. Here ℓ_2 is the usual vector 2-norm and we use S_∞ to emphasise that the output norm is the Schatten ∞-norm for operators.

11. Given a pure state $|\psi\rangle \in B(\mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^d)$, define the geometric measure of entanglement

$$E_{\text{geom}}(|\psi\rangle) = -\log \max_{x,y,z \in B(\mathbb{C}^d)} |\langle\psi| \cdot |x\rangle \otimes |y\rangle \otimes |z\rangle|^2. \tag{10}$$

Then determine whether $E_{\text{geom}}(|\psi\rangle)$ is $\leq \log(d/c)$ or $\geq \log(d/s)$, for $|\psi\rangle^{ABC}$ satisfying $\psi_A = I/d$.

12. Given a subspace $V \subseteq \mathbb{C}^d \otimes \mathbb{C}^d$, define the minimum entanglement of V to be

$$\nu_\infty(V) := \min_{|\psi\rangle \in B(V)} \| \operatorname{tr}_1 |\psi\rangle\langle\psi| \|_\infty, \tag{11}$$

where tr_1 means the partial trace over the first subsystem. Then determine whether $\nu_\infty(V)$ is $\geq c$ or $\leq s$).

13. Given a Hermitian $K \in L(\mathbb{C}^d \otimes \mathbb{C}^d)$ with $0 \leq K \leq I$, define the mean-field Hamiltonian $H_n \in L((\mathbb{C}^d)^{\otimes n})$ by

$$H_n := \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} K^{(i,j)}, \tag{12}$$

where $K^{(i,j)}$ indicates the operator with K acting on systems i, j and identity matrices elsewhere. Let $\lambda_{\min}(H_n)$ denote the smallest eigenvalue of H_n. Then determine whether $\lim_{n\to\infty} \lambda_{\min}(H_n)$ is $\geq 1 - s/2$ or $\leq 1 - c/2$.

The following problems can be reduced to and from estimating h_{Sep}, but unlike the above problems, the reductions no longer preserve the same completeness and soundness.

Approximately equivalent problems

14. Separability testing: given a state ρ and a promise that it is either separable or a constant distance away from separable in the trace norm, determine which is the case. In other words, solve $\text{WMEM}_\epsilon(\text{Sep}(d,d))$ for some $\epsilon > 0$.

15. Weak membership for entanglement witnesses (defined in Eq. (8)), with distance defined in operator norm; i.e. $\text{WMEM}_\epsilon^\infty(\text{EW})$.

16. Injective tensor norm for k-partite states with $k \geq 4$, geometric measure of entanglement for k-partite states with $k \geq 4$, mean field for interactions that are k-local for $k \geq 3$ and $h_{\text{Sep}(d,d,d,\ldots)}$ and weak membership in $\text{Sep}(d,d,d,\ldots)$ for more systems.

17. Estimating the $\ell_2 \to \ell_4$ norm of a matrix, defined as $\|A\|_{\ell_2 \to \ell_4} := \sup_{x \neq 0} \|Ax\|_{\ell_4} / \|x\|_{\ell_2}$, where $\|x\|_{\ell_p} := (\sum_{i=1}^d |x_i|^p)^{1/p}$.

The following problems are at least as easy as h_{Sep}, meaning that they can be reduced to estimating h_{Sep}. We will discuss below the specific parameters of the reductions.

Easier problems

18. Deciding $\text{3-SAT}_{\log^2}$, which is defined to be the class of 3-SAT instances with $\log^2(d)$ variables and $O(\log^2(d))$ clauses. By Corollary 12, this reduces to $\mathsf{QMA}_{\log}(2)_{1/2,1}$.

19. The *planted clique problem* is to distinguish a $G_{n,1/2}$ graph (i.e. an undirected graph with n vertices in which each edge is present with i.i.d. probability 1/2) from the union of a $G_{n,1/2}$ graph and a random clique of size n^β. For certain values of β, as we discuss below, this problem is known to reduce to estimating injective tensor norms.

The following problems are at least as hard as estimating h_{Sep}, meaning that h_{Sep} can be reduced to them, in the special case when $c = 1$.

Harder problems (when $c = 1$)

20. Given a channel $\mathcal{N}$, determine whether $S_\alpha^{\min}(\mathcal{N}) = 0$ or is $\geq \log(1/s)$. The minimum output Rényi α-entropy of $\mathcal{N}$ is defined to be $S_\alpha^{\min}(\mathcal{N}) = \min_\rho S_\alpha(\mathcal{N}(\rho))$, where $S_\alpha(\sigma) = \frac{1}{1-\alpha}\log \operatorname{tr}\sigma^\alpha$.

21. Determine whether the *regularised* minimum output Rènyi entropy $S_\alpha^{R,\min}(\mathcal{N})$ is 0 or $\geq \log(1/s)$. Here $S_\alpha^{R,\min}(\mathcal{N}) = \lim_{n\to\infty}\frac{1}{n}S_\alpha^{\min}(\mathcal{N}^{\otimes n})$.

Before explaining the connections between these problems, we note that Corollary 13 can be restated in terms of h_{Sep}, and thus also in terms of any of the equivalent or harder problems.

Corollary 14. *Tasks 1-13 and 20-21 cannot be completed in time* $\mathrm{poly}(d)$ *for any constants* $0 < s < c < 1$ *unless 3-SAT* $\in \mathsf{DTIME}(\exp(\sqrt{n}\,\mathrm{poly}\log(n)))$.

Explanations

1. *Changing c and s for h_{Sep}:* This claim follows from Theorem 11. Similarly difficult is the problem of producing an estimate X such that $|X - h_{\mathrm{Sep}}(M)| \leq \epsilon$ for some $\epsilon > 0$.

   One subtlety is that the $c = 1$ case is not known to be equivalent to the $c < 1$ case. Soundness, on the other hand, is always nonzero, since we always have $h_{\mathrm{Sep}(d,d)}(M) \geq \operatorname{tr} M/d^2$.

2. *Estimating h_{ProdSym}:* We show this has equivalent difficulty to estimating h_{Sep}. Initially assume that we have an algorithm for estimating h_{ProdSym}, and given $M \in L(\mathbb{C}^d \otimes \mathbb{C}^d)$, would like to compute $h_{\mathrm{Sep}}(M)$. Then define $M' = |01\rangle\langle 01| \otimes M$. M' is $4d^2$-dimensional, and if $0 \leq M \leq I$, then $0 \leq M' \leq I$.

   To calculate h_{ProdSym}, we can without loss of generality let

   $$|\psi\rangle = \sqrt{p_0}|0\rangle|\alpha\rangle + \sqrt{p_1}|1\rangle|\beta\rangle, \tag{13}$$

   where $p_0 + p_1 = 1$ and $|\alpha\rangle, |\beta\rangle \in B(\mathbb{C}^d)$. Then $\operatorname{tr} M'(\psi\otimes\psi) = p_0 p_1 \operatorname{tr} M(\alpha\otimes\beta)$. This is maximised when $p_0 = p_1 = 1/2$. Thus

   $$h_{\mathrm{ProdSym}(2d)}(M') = \frac{1}{4}h_{\mathrm{Sep}(d,d)}(M).$$

   Conversely, suppose we are given an arbitrary M and the ability to compute $h_{\mathrm{Sep}}(\cdot)$ and would like to estimate $h_{\mathrm{ProdSym}}(M)$. First we assume $\mathcal{F}M\mathcal{F} = M$. This can be done WLOG since $h_{\mathrm{ProdSym}}(M) = h_{\mathrm{ProdSym}}((M + \mathcal{F}M\mathcal{F})/2)$. Then define $M' = \frac{\Pi_{\mathrm{sym}}^{d,2} + M}{2}$. Our desired equivalence will follow from the following claim:

   $$h_{\mathrm{Sep}}(M') = \frac{1 + h_{\mathrm{ProdSym}}(M)}{2}. \tag{14}$$

   One direction is easy: if $h_{\mathrm{ProdSym}}(M) = \operatorname{tr} M(\psi\otimes\psi)$ then $h_{\mathrm{Sep}}(M') \geq \operatorname{tr} M'(\psi\otimes\psi) = (1 + h_{\mathrm{ProdSym}}(M))/2$. To upper-bound $h_{\mathrm{Sep}}(M') = \max_{\alpha,\beta}\operatorname{tr} M'(\alpha\otimes\beta)$, we define $\theta, |a\rangle, |b\rangle$ such that

   $$\begin{aligned}|\alpha\rangle &= \cos(\theta/2)|a\rangle + \sin(\theta/2)|b\rangle\\ |\beta\rangle &= \cos(\theta/2)|a\rangle - \sin(\theta/2)|b\rangle.\end{aligned}$$

To compute $\operatorname{tr} M'(\alpha\otimes\beta)$, first we see that $\operatorname{tr}\Pi_{\mathrm{sym}}^{d,2}(\alpha\otimes\beta) = (1+\operatorname{tr}\alpha\beta)/2 = 1-\sin^2(\theta)/2$. Next, we expand

$$|\alpha\rangle|\beta\rangle = \cos^2(\theta/2)|aa\rangle + \sin(\theta/2)\cos(\theta/2)(|ba\rangle - |ab\rangle) - \sin^2(\theta/2)|bb\rangle.$$

When we expand $\langle\alpha,\beta|M|\alpha,\beta\rangle$, the symmetry of M means that terms such as $\langle aa|M(|ba\rangle - |ab\rangle)$ vanish, and we are left with

$$\cos^4(\theta/2)\langle aa|M|aa\rangle + \sin^4(\theta/2)\langle bb|M|bb\rangle - \sin^2(\theta/2)\cos^2(\theta/2)(\langle aa|M|bb\rangle + \langle bb|M|aa\rangle) \\ + 2\sin^2(\theta/2)\cos^2(\theta/2)\frac{\langle ba| - \langle ab|}{\sqrt{2}} M \frac{|ba\rangle - |ab\rangle}{\sqrt{2}}.$$

Since $\|M\| \leq 1$, and using the definition of h_{ProdSym}, we have

$$\begin{aligned}\operatorname{tr} M'(\alpha\otimes\beta) &\leq 1 - \frac{\sin^2(\theta)}{2} + \frac{(\sin^4(\theta/2)+\cos^4(\theta/2))h_{\mathrm{ProdSym}}(M) + \sin^2(\theta)}{2} \\ &\leq \frac{1+h_{\mathrm{ProdSym}}(M)}{2}.\end{aligned}$$

Maximising over all unit vectors α, β, this establishes Eq. (14). We remark that Lemma 5 would also relate h_{Sep} and h_{ProdSym} but not in this exact fashion.

3. *Estimating h_{SepSym}:* Given M, let $M' = M^{A_1B_1} \otimes I^{A_2B_2}$. Then $h_{\mathrm{ProdSym}}(M') = h_{\mathrm{SepSym}}(M)$.

   For the converse, we use the same construction as h_{ProdSym}. Assume that we have an algorithm for estimating h_{SepSym}, and given $M \in L(\mathbb{C}^d\otimes\mathbb{C}^d)$, would like to compute $h_{\mathrm{Sep}}(M)$. Again we define $M' = |01\rangle\langle 01| \otimes M$. Let ρ achieve the maximum of $\operatorname{tr} M'(\rho\otimes\rho)$, and expand $\rho = |0\rangle\langle 0|\otimes\rho_{00} + |0\rangle\langle 1|\otimes\rho_{01} + |1\rangle\langle 0|\otimes\rho_{10} + |1\rangle\langle 1|\otimes\rho_{11}$, for some $\rho_{ij} \in L(\mathbb{C}^d)$. Then $\operatorname{tr} M'(\rho\otimes\rho) = \operatorname{tr} M(\rho_{00}\otimes\rho_{11})$. Since ρ_{00}, ρ_{11} are proportional to density matrices, and $\operatorname{tr}\rho = \operatorname{tr}\rho_{00} + \operatorname{tr}\rho_{11}$, the rest of the analysis proceeds identically to in the case of h_{ProdSym}.

4. *Entanglement witnesses:* $h_{\mathrm{Sep}}(M)$ is a convex program whose dual is given by the minimisation of $\|M+W\|$ over $W \in \mathrm{EW}$. See [32] for a discussion of this point.

5. *Estimating $\|\mathcal{N}\|_{1\to\infty}$:* This connection has been known for some time as folklore and has appeared before in Ref. [56] (which cites a personal communication from Watrous). Since the largest value of $\|\mathcal{N}(\rho)\|_\infty$ occurs when ρ is pure, finding it corresponds to optimising a trilinear form over unit vectors [72]; i.e. is equivalent to the injective tensor norm problem described in task 9. More concretely, define $V_{\mathcal{N}} : \mathbb{C}^d \to \mathbb{C}^d\otimes\mathbb{C}^d$ to be the isometric extension of $\mathcal{N}$, so that $\operatorname{tr}_E V_{\mathcal{N}}\rho V_{\mathcal{N}}^\dagger = N(\rho)$. Then

$$\|\mathcal{N}\|_{1\to\infty} = \max_{\alpha,\beta,\gamma\in B(\mathbb{C}^d)} |(\langle\beta|\otimes\langle\gamma|)V_{\mathcal{N}}|\alpha\rangle|^2. \tag{15}$$

   This expression equals $\|T\|_{\mathrm{inj}}^2$ (see task 9) for $|T\rangle = \sum_{i=1}^d |i\rangle\otimes V_{\mathcal{N}}|i\rangle$.

6. *Estimating $\|\mathcal{N}\|_{1\to 2}$:* Define $M := (\mathcal{N}^\dagger\otimes\mathcal{N}^\dagger)(\frac{I+\mathcal{F}}{2})$. Then $h_{\mathrm{ProdSym}}(M) = \max\{(1+\operatorname{tr}(\mathcal{N}(\psi))^2)/2 : |\psi\rangle \in B(\mathbb{C}^d)\} = (1+\|\mathcal{N}\|_{1\to 2}^2)/2$. By Eq. (14), there exists M' with $h_{\mathrm{Sep}}(M') = (3+\|\mathcal{N}\|_{1\to 2}^2)/4$.

7. *Estimating $S_\infty^{\min}(\mathcal{N})$:* Since $S_\infty^{\min}(\mathcal{N}) = -\log\|\mathcal{N}\|_{1\to\infty}$, this is equivalent to task 5.

8. *Estimating $S_2^{\min}(\mathcal{N})$:* Similarly, this is equivalent to task 6.

9. *Estimating* $\|T\|_{\text{inj}}$*:* This relates to h_{Sep} in a way that is analogous to the relation between the largest singular value of a matrix A and the largest eigenvalue of $A^\dagger A$.

$$\begin{aligned}
\|T\|_{\text{inj}}^2 &= \max_{x,y,z \in B(\mathbb{C}^d)} \left| \sum_{i,j,k=1}^{d} T_{i,j,k} x_i y_j z_k \right|^2 \\
&= \max_{x,y \in B(\mathbb{C}^d)} \left\| \sum_{i,j,k=1}^{d} T_{i,j,k} x_i y_j |k\rangle \right\|_2^2 \\
&= \max_{x,y \in B(\mathbb{C}^d)} \sum_{i,j,i',j',k=1}^{d} T_{i,j,k} T^*_{i',j',k} x_i y_j x^*_{i'} y^*_{j'} \\
&= h_{\text{Sep}} \left(\sum_{i,j,i',j',k=1}^{d} T_{i,j,k} T^*_{i',j',k} |i\rangle\langle i'| \otimes |j\rangle\langle j'| \right). \qquad (16)
\end{aligned}$$

We can think of T as a $d^2 \times d$ matrix by grouping indices i, j together. By doing so, Eq. (16) becomes simply $h_{\text{Sep}}(TT^\dagger)$ (and by our assumption about the operator norm of the matrix version of T, we have that $TT^\dagger \leq I$). To show the equivalence holds in both directions, observe that any $M \geq 0$ with rank $\leq d$ can be written as $TT^\dagger$ for some $d^2 \times d$ matrix T. This rank restriction can be removed either by taking T to be a $d \times d \times d^2$ tensor, or by suitable padding.

10. *Estimating* $\|T\|_{\ell_2 \to S_\infty}$*:* Observe that

$$\|T\|_{\ell_2 \to S_\infty} = \max_{x \in B(\mathbb{C}^d)} \|T(x)\|_{S_\infty} = \max_{x,y,z \in B(\mathbb{C}^d)} |\langle y|T(x)|z\rangle|.$$

This last expression is the maximum of a trilinear form over triples of unit vectors, and so is equivalent to computing an injective tensor norm (see task 9).

11. *Estimating* $E_{\text{geom}}(|\psi\rangle)$*:* Treating $|\psi\rangle$ as a 3-index tensor, it is apparent from the definitions that $E_{\text{geom}}(|\psi\rangle) = -\log \||\psi\rangle\|_{\text{inj}}^2$. The condition on ψ_A corresponds to the requirement that $\sqrt{d}$ times the resulting tensor should be be an isometry when interpreted as a map from $A \to BC$. Thus the estimation problems are equivalent. The $\sqrt{d}$ factor also explains why we need to distinguish the cases $E_{\text{geom}} \leq \log(d/c)$ and $\geq \log(d/s)$. Interestingly, Theorem 1 shows that it is *easy* to distinguish whether the geometric measurement of entanglement is $\leq \epsilon$ or $\geq C + \epsilon$ for a sufficiently large constant C.

12. *Estimating* $\nu_\infty(V)$*:* Suppose that $\dim V = m$. Define T to be an isometry from $\mathbb{C}^m$ to V. Then $\nu_\infty(V) = \|T\|_{\ell_2 \to S_\infty}$, and estimating $\nu_\infty(V)$ is equivalent to task 10. For simplicity, one can assume that $m = d$ by padding the appropriate dimensions; this does not affect the complexity by more than a polynomial factor.

13. *Mean-field Hamiltonians:* In Ref. [29], the quantum de Finetti theorem was used to show that when $n \gg d^2$, then the ground state of H is very close to a product state. In the limit, finding the ground-state energy density of H is equivalent to calculating the quantity

$$\max_{\rho \in \mathcal{B}(d)} \operatorname{tr} K(\rho \otimes \rho).$$

This task is therefore equivalent to task 3.

14. *Separability testing:* A classic result in convex optimisation [35] allows one to show that $\mathrm{WMEM}_\epsilon(\mathrm{Sep}(d,d))$ is roughly equivalent to estimating h_{Sep}. Unfortunately, known versions of this result give up $1/\operatorname{poly}(d)$ factors in the approximation guarantees. This fact has been used to show the NP-hardness of $\mathrm{WMEM}_{1/\operatorname{poly}}(\mathrm{Sep})$ in Refs. [52, 31, 10] and, previously, of $\mathrm{WMEM}_{1/\exp}(\mathrm{Sep})$ by Gurvits [38] (although the connection to $\mathsf{QMA}_{\log}(2)$ was only observed by [10]). We conjecture that $\mathrm{WMEM}_\epsilon(\mathrm{Sep}(d,d))$ should be $\mathsf{NP}_{\log^2}$-hard for some $\epsilon > 0$; i.e. that $\mathrm{Sep}(d,d)$ cannot be approximated to (sufficiently small) constant accuracy in time $\operatorname{poly}(d)$.

    However, we are able to rule out only algorithms that have the further restriction of recognizing a nearly convex set that in turn approximates Sep to constant accuracy. The following result is an immediate consequence of Corollary 4.3.12 of [35] and Corollary 13.

    Proposition 15. *Suppose that there exists a constant $\epsilon > 0$ such that for all d, there exists a convex set K_d with Hausdorff distance ϵ to $\mathrm{Sep}(d,d)$ such that $\mathrm{WMEM}_{1/\operatorname{poly}(d)}(K)$ can be solved in time $\operatorname{poly}(d)$. Then 3-SAT $\in \mathsf{DTIME}(\exp(\sqrt{n}\operatorname{poly}\log(n)))$.*

    As a result, one possible alternate title for our paper could have been:

    Detecting pure entanglement is easy, so detecting mixed entanglement is hard.

    In fact, reductions between WMEM and approximating h_{Sep} go in both directions. We have to be careful not to assume (as does [47]) that approximation algorithms for h_{Sep} output an approximately optimal density matrix. Indeed, some approximations (e.g. [16]) only output a scalar value approximating h_{Sep}. However, we can prove the following reduction.

    Proposition 16. *Let $f(M)$ be a convex function such that $f(0) = 0$ and $|f(M) - h_{\mathrm{Sep}(d,d)}(M)| \leq \epsilon\|M\|_\infty$. Given oracle access to f, we can solve $\mathrm{WMEM}_{2\epsilon}(\mathrm{Sep}(d,d))$ in time $\operatorname{poly}(d)$.*

    Proof. Suppose we are given a density matrix ρ for which we would like to solve $\mathrm{WMEM}_\epsilon(\mathrm{Sep}(d,d))$. The algorithm computes

    $$Z := \max\{\operatorname{tr} M\rho - f(M) : -I \leq M \leq I\}.$$

    This can be done in polynomial time [35]. If $Z \leq \epsilon$, then we declare that $\rho \in \mathrm{Sep}(d,d)$, and if $Z > \epsilon$ then we declare that $\rho \notin B_1(\mathrm{Sep}(d,d),\epsilon)$.

    To analyze the correctness of the algorithm, we prove rather that it is *not wrong*. In other words, we need to give the correct answer in the cases: (1) when $\rho \notin B_1(\mathrm{Sep}(d,d), 2\epsilon)$, and (2) when $\rho \in \mathrm{Sep}(d,d)$. In case (1), then ρ has trace distance $> 2\epsilon$ from every point in $\mathrm{Sep}(d,d)$ and so there exists a M with $\|M\|_\infty \leq 1$ for which $\operatorname{tr} M\rho > h_{\mathrm{Sep}}(M) + 2\epsilon$. This implies that $\operatorname{tr} M\rho > f(M) + \epsilon$, and that $Z > \epsilon$.

    On the other hand, in case (2), we have $\rho \in \mathrm{Sep}(d,d)$, which implies $\operatorname{tr} M\rho \leq h_{Sep(d,d)}(M) \leq f(M) + \epsilon$ for all M, and thus $Z \leq \epsilon$. □

15. *Weak membership for entanglement witnesses:* For a Hermitian matrix M, we have $h_{\mathrm{Sep}}(M) \leq \epsilon$ if and only if $M \in B_\infty(\mathrm{EW},\epsilon)$. Here $B_\infty(S,\epsilon)$ refers to the points within ϵ of a set S in the Schatten-∞ norm. This shows that if we can approximate h_{Sep} then we can solve the weak membership problem for EW. Conversely, if we are given an algorithm for $\mathrm{WMEM}^\infty_\epsilon(\mathrm{EW})$, then on input M we can use binary search to find approximately the smallest γ such that $M - \gamma I \in \mathrm{EW}$. This γ will be within ϵ of $h_{\mathrm{Sep}}(M)$.

16. *k-partite tensor norm problems:* By adding more systems, we will not make any of the problems any easier. To reduce from an injective tensor norm on k-tensors to the injective tensor norm on 3-tensors, we can use Lemma 5. The other reductions claimed in this point are similar.

    When performing these mappings, there is no direct penalty that depends on k. However, the dimensions of the spaces involved will scale exponentially with k. For example, estimating the support function of $\mathrm{Sep}(d_1, d_2, \ldots, d_k)$ is harder than estimating $h_{\mathrm{Sep}(d_1,d_2)}$, and by Lemma 5 can be reduced to estimating $h_{\mathrm{Sep}(d,d)}$, where $d := \mathrm{poly}(d_1 d_2 \cdots d_k)$.

17. *$\ell_2 \to \ell_4$ norm:* If $A = \sum_{i=1}^m |i\rangle\langle\alpha_i|$ with each $|\alpha_i\rangle \in \mathbb{C}^n$, then

    $$\|A\|_{2\to 4}^4 = \max_{|\psi\rangle\in S(\mathbb{C}^n)} |\langle\alpha_i|\psi\rangle|^2 = h_{\mathrm{ProdSym}}(\sum_i \alpha_i^{\otimes 2}) \tag{17}$$

    To show a reduction in the other direction, we need to convert any measurement M into an M' with similar h_{Sep} such that M' can be written as $\sum_i \alpha_i \otimes \alpha_i$. Theorem 11 instead allows us to write M in the form $\sum_i \alpha_i \otimes \beta_i$, but the desired $\sum_i \alpha_i \otimes \alpha_i$ form can be achieved by taking $M' = (\sqrt{M}^{A_1B_1} \otimes \sqrt{M}^{A_2B_2})(P_{\mathrm{sym}}^{A_1B_1} \otimes P_{\mathrm{sym}}^{A_2B_2})(\sqrt{M}^{A_1B_1} \otimes \sqrt{M}^{A_2B_2})$. This construction is analyzed in [8].

    As a result, it is NP-hard to approximate the $2 \to 4$ norm to $1/\mathrm{poly}(d)$ accuracy, and it is $\mathsf{NP}_{\log^2}$-hard to approximate it to within a constant multiplicative error.

18. *3-SAT on $\log^2$ variables:* Corollary 12 explains how 3-SAT instances with $O(\log^2(n))$ variables can be reduced to determining whether $h_{\mathrm{Sep}(n,n)}(M) = 1$ or is $\leq 1/2$, for some efficiently-computable matrix M. It is an extremely interesting open question to determine whether the reverse reduction is also possible.

19. *Planted clique problem:* In [17], the problem of finding a clique of size $\Omega(n^{1/r} r^5 \log^3(n)\alpha^2)$ planted in a $G_{n,1/2}$ graph was reduced to determining whether $\|T\|_{\mathrm{inj}}$ is $\geq \alpha^r$ or ≤ 1. Here $\alpha \leq 1$ and T is a tensor in $(\mathbb{C}^n)^{\otimes r}$ with all ± 1 entries. Thus, we can trivially bound $\|TT^\dagger\|_\infty \leq n^r$ (although we suspect that the norm should typically be $O(n^{r-1})$).

    For concreteness, let us focus on the case of $r = 3$. In this case, [17] reduce the problem of finding a clique of size $n^{1/3+o(1)}$ to $\mathsf{QMA}_{\log}(2)_{1/n^{1.5},2/n^{1.5}}$ (or $\mathsf{QMA}_{\log}(2)_{1/n,2/n}$, if one assumes that $\mathbb{E}\|TT^\dagger\|_\infty \leq O(n^{r-1})$). Since random graphs typically have no clique of size larger than $2\log n + 1$, the planted clique problem can always be reduced to a Circuit-SAT instance of size $\mathrm{poly}(n)$ with $O(\log^2(n))$ input variables. Since

    $$\mathsf{QMA}_{\log}(2)_{1/n^{1.5},2/n^{3/2}} \supseteq \mathsf{QMA}_{\log}(2)_{1/2,1} \ni \text{3-SAT}_{\log^2},$$

    this implies that the reduction of [17] achieves a reduction that is comparable to the previously known reduction to Circuit-SAT. It is an interesting open question to determine whether Circuit-SAT instances with size $\mathrm{poly}(n)$ and $O(\log^2(n))$ input variables can be placed in $\mathsf{QMA}_{\log}(2)_{1/2,1}$. If this were possible, then it would imply that the reduction of [17] would be strictly subsumed by the previous reduction of planted clique to Circuit-SAT.

20. *Minimum output Rènyi entropies:* For any $\alpha \geq 0$, we have $S_\alpha^{\min}(\mathcal{N}) \geq S_\infty^{\min}(\mathcal{N})$ but also $S_\alpha^{\min}(\mathcal{N}) = 0$ iff $S_\infty^{\min}(\mathcal{N}) = 0$. Thus, for any $c > 0$, distinguishing between $S_\alpha^{\min} = 0$ and $S_\alpha^{\min} \geq c$ is at least as hard as distinguishing between $S_\infty^{\min} = 0$ and $S_\infty^{\min} \geq c$.

21. *Regularised minimum output Rènyi entropies:* Our hardness result for $S_\alpha^{\min}$ immediately gives us the equivalent hardness result for $S_\alpha^{R,\min}$. The reason is that our proof of amplification for $\mathsf{QMA}(2)$ protocols (see Lemma 7) essentially works by constructing a channel $\mathcal{N}$ for which $S_\infty^{R,\min}(\mathcal{N}) = S_\infty^{\min}(\mathcal{N})$ by design.

4.2.1 Additional remarks

- *Additivity violations:* As a result of the connection between $\mathsf{QMA}(2)$ and estimating $S_\infty^{\min}$, the question of whether $\mathsf{QMA}(2)$ protocols can be amplified to exponentially small error is directly related to the question of additivity of the minimum output min-entropy (equivalently, multiplicativity of the maximum output infinity norm). Indeed, additivity violations for $S_\infty^{\min}$ (e.g. [42, 41, 36]) translate directly into $\mathsf{QMA}(2)$ protocols for which perfect parallel repetition fails[2]. Conversely, [56] observed that $\mathsf{QMA}(2)$ protocols obey perfect parallel repetition when the corresponding channel $\mathcal{N}$ is known to have additive $S_\infty^{\min}$, for example when $\mathcal{N}$ is entanglement breaking. Indeed our Lemma 7 is a restatement of this point.

- *Minimum output entropy:* Beigi and Shor previously showed that it is NP-hard to compute the minimum output entropy up to $1/\operatorname{poly}(d)$ accuracy [11]. Our result improves their accuracy requirement, but under a more restrictive complexity assumption. For general channels, we automatically have $S_\alpha^{R,\min}(\mathcal{N}) \leq S_\alpha^{\min}(\mathcal{N})$; however, the famous failures of the additivity conjecture imply that sometimes this inequality can be strict, with examples known for $\alpha \geq 1$ [40, 41] and for α near 0 [24]. Still, these examples only demonstrate that $S_\alpha^{R,\min}$ can deviate very slightly from $S_\alpha^{\min}$. On the other hand, various lower bounds for $S_\alpha^{R,\min}$ are known [68, 74, 26, 75], and it may be that one of these bounds could be related to $S_\alpha^{\min}$, thereby proving that $S_\alpha^{R,\min}$ cannot be far from $S_\alpha^{\min}$. Our results do not rule out the possibility that $S_\alpha^{\min}$ may be fruitfully related to $S_\alpha^{R,\min}$. However, they do imply that these lower bounds on $S_\alpha^{R,\min}$ (and thereby on $S_\alpha^{\min}$) are unlikely to be efficiently computable, or if they are, they are likely to be extremely loose bounds in general.

- *Mean-field approximation:* Previous work on the hardness of approximating ground-state energy of quantum systems generally had d constant and only ruled out the possibility of $1/\operatorname{poly}(n)$ approximation error. By addressing the case when approximation error is a constant fraction of the overall energy of the system, our result achieves one of the goals of the conjectured quantum PCP theorem [3]. However, we require d to grow asymptotically, and we achieve a hardness result much weaker than QMA-hardness. Indeed, due to the *classical* PCP theorem combined with the Exponential Time Hypothesis, finding the ground state of a system of $d^2 \log(d)$ bits (without any symmetry constraint) is likely to require time $\exp(d^2 \log(d))$, while our results merely imply an $\exp(\Omega(\log^2(d)))$ lower bound. Still, our result provides a superpolynomial bound on an important class of Hamiltonians that had been previously considered to be computationally easy to work with.

5 Optimality of the product test

Our product test has perfect completeness in the sense that if $|\psi\rangle$ is exactly a product state then it will always pass the product test. Soundness could be in principle described by a functional relation

[2] Note that, taking the standard definition of $\mathsf{QMA}(2)$, this is strictly speaking only true if the corresponding $\mathsf{QMA}(2)$ protocol can be implemented in polynomial time.

between maximum acceptance probability and distance to the nearest product state. However, for our purposes, we can say that our test has constant soundness in that if $|\psi\rangle$ has overlap at most $1-\epsilon$ with any product state then it will pass the product test with probability at most $1-\Theta(\epsilon)$.

In fact, if we consider only product-state tests with perfect completeness, then we can show that our test has optimal soundness: that is, it rejects as often as possible given the constraint of always accepting product states. More generally, suppose that a product-state test T is given $|\psi\rangle^{\otimes k}$ as input. Since the outcome of the test is binary, we can say that T is an operator on the nk-qudit Hilbert space with $0 \leq T \leq I$ and that the test accepts with probability $\operatorname{tr} T\psi^{\otimes k}$.

Let S be the set of product states in $\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, and define S^k to be the span of $\{|\phi\rangle^{\otimes k} : |\phi\rangle \in S\}$. For a single system $\mathbb{C}^d$, the span of $\{|\phi\rangle^{\otimes k} : |\phi\rangle \in \mathbb{C}^d\}$ is denoted $\operatorname{Sym}^k \mathbb{C}^d$. This is the symmetric subspace of $(\mathbb{C}^d)^{\otimes k}$, meaning that it can be equivalently defined to be the set of vectors in $(\mathbb{C}^d)^{\otimes k}$ that is invariant under permutation by the symmetric group S_k. This fact allows the projector onto $\operatorname{Sym}^k \mathbb{C}^d$, which we denote $\Pi^{\text{sym}}_{d,k}$, to be implemented efficiently [9]. Also, it implies that $S^k = \operatorname{Sym}^k \mathbb{C}^{d_1} \otimes \cdots \otimes \operatorname{Sym}^k \mathbb{C}^{d_n}$ and that the projector onto S^k, denoted Π_{S^k}, is $\Pi^{\text{sym}}_{d_1,k} \otimes \cdots \otimes \Pi^{\text{sym}}_{d_n,k}$.

Now we return to our discussion of product-state tests. If $\operatorname{tr} T\phi^{\otimes k} = 1$ for all $\phi \in S$, then $T \geq \Pi_{S^k}$. Thus, T will always accept at least as often as Π_{S^k} will on any input, or equivalently, taking $T = \Pi_{S^k}$ yields the test which rejects as often as possible given the constraint of accepting every state in S^k.

To understand Π_{S^k}, note that the projector onto $\operatorname{Sym}^k \mathbb{C}^d$ is given by $\frac{1}{k!}\sum_{\pi \in \mathcal{S}_k} P_d(\pi)$, where

$$P_d(\pi) = \sum_{i_1,\ldots,i_k \in [d]} |i_1,\ldots,i_k\rangle\langle i_{\pi(1)},\ldots,i_{\pi(k)}|. \tag{18}$$

For $k=1$, $\operatorname{Sym}^1 \mathbb{C}^d$ simply equals $\mathbb{C}^d$, and Π_{S^1} is the identity operator on $(\mathbb{C}^d)^{\otimes n}$. Thus, no non-trivial product-state test is possible when given one copy of $|\psi\rangle$.

When $k=2$, $\operatorname{Sym}^2 \mathbb{C}^d$ is the $+1$ eigenspace of $(I+\mathcal{F})/2$, which is the space that passes the swap test. Thus, the product test (in Protocol 1) performs the projection onto S^2 and therefore rejects non-product states as often as possible for a test on $|\psi\rangle^{\otimes 2}$ that always accepts when $|\psi\rangle$ is a product state. These arguments also imply that given $|\psi\rangle^{\otimes k}$, projecting onto S^k yields an optimal k-copy product-state test of $|\psi\rangle$. The strength of these tests is strictly increasing with k, but we leave the problem of analysing them carefully to future work.

Finally, this interpretation of the product test allows us to consider generalisations to testing membership in other sets S. The general prescription for a test that is given k copies of a state is simply to project onto the span of $\{|\psi\rangle^{\otimes k} : |\psi\rangle \in S\}$. We will not explore these possibilities further in this paper, but see [70] for a subsequent paper that considers variations on this theme for the related problem of testing properties of unitary operators.

6 Stability of the depolarising channel

As discussed in Section 2, the correctness proof of the product test in fact applies to a larger class of processes acting on two copies of n-partite states. In general, choosing S according to a binomial distribution on $[n]$ and taking the expectation of $\operatorname{tr} \rho_S^2$ is equivalent to evaluating the output purity of n uses of the depolarising channel $\mathcal{D}_\delta$ (as defined in (2)). A special case of this corresponds to the probability of the product test passing.

Lemma 17. *We have*

$$\operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,\rho)^2 = \left(\frac{1-\delta^2}{d}\right)^n \sum_{S\subseteq[n]} \left(\frac{d\delta^2}{1-\delta^2}\right)^{|S|} \operatorname{tr}(\rho_S^2),$$

and in particular

$$\operatorname{tr}(\mathcal{D}_{1/\sqrt{d+1}}^{\otimes n}\,\rho)^2 = \frac{1}{(d+1)^n} \sum_{S\subseteq[n]} \operatorname{tr}(\rho_S^2),$$

and for pure product states,

$$\mathrm{OPP}(\delta) := \operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,(|\psi_1\rangle\langle\psi_1| \otimes\cdots\otimes |\psi_n\rangle\langle\psi_n|))^2 = \left(\frac{d-1}{d}\delta^2 + \frac{1}{d}\right)^n.$$

We will see that it is possible to prove a more general version of Theorem 1.

Theorem 18. *Given $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, let*

$$1-\epsilon = \max\{|\langle\psi|\phi_1,\dots,\phi_n\rangle|^2 : |\phi_1\rangle,\dots,|\phi_n\rangle \in \mathbb{C}^d\}.$$

Then (recalling the definitions of $\mathcal{D}_\delta$ and OPP(δ) *from equations (2) and (3)),*

$$\operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,|\psi\rangle\langle\psi|)^2 \le \mathrm{OPP}(\delta)\left(1 - 4\epsilon(1-\epsilon)\frac{d\delta^2(1-\delta^2)}{(1{+}(d-1)\delta^2)^2} + 4\epsilon^{3/2}\left(\frac{(1-\delta^2)^2+d^2\delta^4}{(1{+}(d-1)\delta^2)^2}\right)^2\right).$$

In particular,

$$\operatorname{tr}(\mathcal{D}_{1/\sqrt{d+1}}^{\otimes n}|\psi\rangle\langle\psi|)^2 \le \mathrm{OPP}(1/\sqrt{d+1})\left(1{-}\epsilon{+}\epsilon^2{+}\epsilon^{3/2}\right).$$

The idea of the proof is more or less the same as the outline sketched in Section 2 and the details can be found in Appendix A.

7 Testing for product unitaries

As well as being useful for testing quantum states, the product test has applications to testing properties of unitary operators. The results we obtain will be in terms of the normalised Hilbert-Schmidt inner product, which is defined as $\langle M,N\rangle := \frac{1}{d}\operatorname{tr} M^\dagger N$ for $M,N \in M(d)$, where $M(d)$ denotes the set of $d\times d$ matrices. Note that, with this normalisation, $|\langle U,V\rangle| \le 1$ for unitary operators U, V.

We consider the problem of testing whether a unitary operator is a tensor product. That is, we are given access to a unitary U on the space of n qudits (for simplicity, restricting to the case where each of the qudits has the same dimension d), and we would like to decide whether $U = U_1\otimes\cdots\otimes U_n$. This is one possible generalisation of the classical problem of testing linearity of functions $f:\{0,1\}^n \to \{0,1\}$ [13]. To see this, observe that f is linear (i.e. $f(x\oplus y) = f(x)\oplus f(y)$ for all x and y) if and only if the function $g:\{0,1\}^n \to \{\pm 1\}$ defined by $g(x) = (-1)^{f(x)}$ is a product of individual functions $g_i(x) = (-1)^{a_i x_i}$, for $a_i \in \{0,1\}$. Thus the diagonal unitary operator U on n qubits defined by $U_{xx} = g(x)$ is a tensor product if and only if f is linear.

In Protocol 3 we give a test that solves this problem using the product test. The test always accepts product unitaries, and rejects unitaries that are far from product, as measured by the normalised Hilbert-Schmidt inner product. Several papers have proposed property tests with similar performance for other sets of unitary matrices: e.g. Pauli matrices [59], Clifford gates [54, 70] and many other sets [70].

The following correspondence (also known as the Choi-Jamiołkowski isomorphism) underlies our ability to apply the product test to unitaries. Let $|\Phi\rangle$ be a maximally entangled state of two d-dimensional qudits, written as $\frac{1}{\sqrt{d}}\sum_{i=1}^{d}|i,i\rangle$ in terms of some basis $\mathcal{B} = (|1\rangle, \dots, |d\rangle)$. For any matrix $M \in M(d^n)$, define $|v(M)\rangle := (M \otimes I)|\Phi\rangle^{\otimes n}$. Then $\langle j|\langle k|v(M)\rangle = \frac{\langle j|M|k\rangle}{\sqrt{d^n}}$. In particular, for any matrices $M, N \in M(d^n)$, $\langle M, N\rangle = \langle v(M)|v(N)\rangle = \operatorname{tr} M^\dagger N/d^n$.

Protocol 3 (Product unitary test).

The product unitary test proceeds as follows.

1. *Prepare two copies of the state $|\Phi\rangle^{\otimes n}$, then in both cases apply U to the n first halves of each pair of qudits to create two copies of the state $|v(U)\rangle \in (\mathbb{C}^{d^2})^{\otimes n}$.*
2. *Return the result of applying the product test to the two copies of $|v(U)\rangle$, with respect to the partition into n d^2-dimensional subsystems.*

Let the probability that this test passes when applied to some unitary U be $P_{\text{test}}(U)$. Then we have the following theorem, which proves a conjecture from [59].

Theorem 19. *Given $U \in U(d^n)$, let*

$$1 - \epsilon = \max\{|\langle U, V_1 \otimes \cdots \otimes V_n\rangle|^2 : V_1, \dots, V_n \in U(d)\}.$$

Then, if $\epsilon = 0$, $P_{test}(U) = 1$. If $\epsilon \lesssim 0.106$, then $P_{test}(U) \le 1 - \frac{1}{4}\epsilon + \frac{1}{16}\epsilon^2 + \frac{1}{8}\epsilon^{3/2}$. If $0.106 \lesssim \epsilon \le 1$, $P_{test}(U) \le 501/512$. More concisely, $P_{test}(U) = 1 - \Theta(\epsilon)$.

The proof is given in Appendix F. It is not quite immediate from the previous results; the key problem is that the closest product state to $|v(U)\rangle$ may not correspond to the closest unitary operator to U.

Our test is sensitive to the Hilbert-Schmidt distance of a unitary from the set of product unitaries. One might hope to design a similar test that instead uses a notion of distance based on the operator norm. However, this is not possible. For example, if we could detect a constant difference in the operator norm between an arbitrary unitary U and the set of product unitaries then we could find a single marked item in a set of size d^n. By the optimality of Grover's algorithm, this requires $\Omega(d^{n/2})$ queries to U. More generally, any test that uses only a constant number of black-box queries to U can only detect an $\Omega(1)$ difference in an $\Omega(1)$ fraction of the d^n dimensions that U acts upon.

8 Open problems

We conclude with a discussion of open problems related to our work.

1. Our main result can be seen as a "stability" theorem for the output purity of the depolarising channel (cf. Section 6).. It is an interesting problem to determine whether a similar result

holds for all output Rényi entropies for the depolarising channel, or even for all channels where additivity holds.

2. Can Theorem 1 be tightened further, perhaps by improving the constant in the $\epsilon^{3/2}$ term? It would also be interesting to improve the constants in Theorem 1 in the regime of large ϵ, as at present they are extremely pessimistic. The regime of large ϵ is generally somewhat mysterious: for example, we do not know the minimum value of P_{test}, or the largest distance from any product state that can be achieved by a state of n qudits. This is equivalent to determining the maximal value of the geometric measure of entanglement [71] which can be achieved by a pure state of n qudits; see the PhD thesis [7] for a review of recent work concerning bounds on this quantity.

3. Suppose the goal is to test whether a given state is of the form $|\varphi\rangle^{\otimes n}$ for some unknown $|\varphi\rangle$. Can we substantially improve on the performance of the product test, say with a test whose acceptance probability decreases exponentially in the number of positions not equal to $|\varphi\rangle$? Ideally we would achieve performance comparable to the exponential de Finetti theorem [63], but without any dependence on dimension. The natural test for this problem is to project onto the symmetric subspace of all $2n$ positions.

4. The relationship between QMA and $\mathsf{QMA}(2)$ remains unresolved. Our Theorem 9 proves that $\mathsf{QMA}^{\text{SEP}}(2) = \mathsf{QMA}(2)$, while the result of [15] implies that $\mathsf{QMA}^{\text{LOCC}}(2) = \mathsf{QMA}$. Can this gap be closed? One possible way to do this would be to improve our results to show that $\mathsf{QMA}^{\text{LOCC}}(2) = \mathsf{QMA}(2)$; but see also Appendix D for a proof that an efficient LOCC product test does not exist. Alternatively, one might improve the simulation of [15] to apply to separable measurements instead of only LOCC measurements, but the obvious approaches to modifying their proof do not appear to work. Finally, if $\mathsf{QMA}(2)$ is not shown to be in QMA, one might hope for any upper bound on its complexity that is better than NEXP.

5. Is there an oracle separation between QMA and $\mathsf{QMA}(2)$? The equalities in the previous point relativise, so this is equivalent to showing a separation between $\mathsf{QMA}^{\text{SEP}}(2)$ and $\mathsf{QMA}^{\text{LOCC}}(2)$.

A The depolarising channel

Let $\mathcal{D}_\delta$ be the qudit depolarising channel as defined in equation (2). We will be interested in applying the n-fold product $\mathcal{D}_\delta^{\otimes n}$ to states of n qudits, and in particular in the purity of the resulting states. This has the following characterisation.

Lemma 17. *We have*

$$\operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,\rho)^2 = \left(\frac{1-\delta^2}{d}\right)^n \sum_{S\subseteq[n]} \left(\frac{d\delta^2}{1-\delta^2}\right)^{|S|} \operatorname{tr}(\rho_S^2),$$

and in particular

$$\operatorname{tr}(\mathcal{D}_{1/\sqrt{d+1}}^{\otimes n}\,\rho)^2 = \frac{1}{(d+1)^n} \sum_{S\subseteq[n]} \operatorname{tr}(\rho_S^2),$$

and for pure product states,

$$\operatorname{OPP}(\delta) := \operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,(|\psi_1\rangle\langle\psi_1| \otimes \cdots \otimes |\psi_n\rangle\langle\psi_n|))^2 = \left(\frac{d-1}{d}\delta^2 + \frac{1}{d}\right)^n.$$

Proof. Consider some Hermitian operator basis for $\mathcal{B}(\mathbb{C}^d)$ which contains the identity and is orthonormal with respect to the normalised Hilbert-Schmidt inner product $\langle A, B\rangle = \frac{1}{d}\operatorname{tr} A^\dagger B$, and extend this basis to $\mathcal{B}((\mathbb{C}^d)^{\otimes n})$ by tensoring. Expand ρ in terms of the resulting basis as

$$\rho = \sum_{\mathbf{t}\in\{0,\dots,d^2-1\}^n} \hat{\rho}_{\mathbf{t}}\chi_{\mathbf{t}}.$$

where $\hat{\rho}_{\mathbf{t}} \in \mathbb{R}$, $\chi_{\mathbf{t}}$ represents an element of the tensor product basis corresponding to the string $\mathbf{t} \in \{0,\dots,d^2-1\}^n$, and the identity is indexed by 0 at each position. Then we have

$$\operatorname{tr}(\rho_S^2) = d^{2n-|S|}\left(\sum_{\mathbf{t},\, \mathbf{t}_i=0,\, \forall i\in\bar{S}} \hat{\rho}_{\mathbf{t}}^2\right),$$

and hence, for any δ,

$$\begin{aligned}
\sum_{S\subseteq[n]} \delta^{|S|}\operatorname{tr}(\rho_S^2) &= d^{2n}\sum_{S\subseteq[n]}(\delta/d)^{|S|}\left(\sum_{\mathbf{t},\, \mathbf{t}_i=0,\, \forall i\in\bar{S}} \hat{\rho}_{\mathbf{t}}^2\right) = d^{2n}\sum_{\mathbf{t}}\hat{\rho}_{\mathbf{t}}^2\left(\sum_{\substack{S\subseteq[n],\\ \mathbf{t}_i=0,\, \forall i\in\bar{S}}}(\delta/d)^{|S|}\right)\\
&= d^{2n}\sum_{\mathbf{t}}\hat{\rho}_{\mathbf{t}}^2\left(\sum_{x=0}^{n-|\mathbf{t}|}\binom{n-|\mathbf{t}|}{x}(\delta/d)^{x+|\mathbf{t}|}\right)\\
&= d^{2n}\sum_{\mathbf{t}}\hat{\rho}_{\mathbf{t}}^2\,(\delta/d)^{|\mathbf{t}|}(1+\delta/d)^{n-|\mathbf{t}|}\\
&= (d(d+\delta))^n\sum_{\mathbf{t}}\hat{\rho}_{\mathbf{t}}^2\,(\delta/(\delta+d))^{|\mathbf{t}|}\\
&= (d+\delta)^n\operatorname{tr}(\mathcal{D}^{\otimes n}_{\sqrt{\delta/(\delta+d)}}\rho)^2.
\end{aligned}$$

Rearranging completes the proof; the two special cases in the statement of the lemma can be verified directly. □

Using the above lemma, we can see that maximal output purity is obtained only for product states, since only product states saturate the inequality $\operatorname{tr}\rho_S^2 \le 1$ for all $S \subseteq [n]$. We will now prove our main result, which is a "stability" theorem for the depolarising channel: if a state achieves close to maximal output purity, it must be close to a product state.

Theorem 18. *Given $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, let*

$$1-\epsilon = \max\{|\langle\psi|\phi_1,\dots,\phi_n\rangle|^2 : |\phi_1\rangle,\dots,|\phi_n\rangle \in \mathbb{C}^d\}. \tag{19}$$

Then

$$\operatorname{tr}(\mathcal{D}_\delta^{\otimes n}|\psi\rangle\langle\psi|)^2 \le \mathrm{OPP}(\delta)\left(1-4\epsilon(1-\epsilon)\frac{d\delta^2(1-\delta^2)}{(1+(d-1)\delta^2)^2}+4\epsilon^{3/2}\left(\frac{(1-\delta^2)^2+d^2\delta^4}{(1+(d-1)\delta^2)^2}\right)^2\right).$$

In particular,

$$\operatorname{tr}(\mathcal{D}^{\otimes n}_{1/\sqrt{d+1}}|\psi\rangle\langle\psi|)^2 \le \mathrm{OPP}(1/\sqrt{d+1})\left(1-\epsilon+\epsilon^2+\epsilon^{3/2}\right).$$

Proof. Without loss of generality assume that one of the states achieving the maximum in Eq. (19) is $|0\rangle^{\otimes n}$, which we will abbreviate simply as $|0^n\rangle$, or $|0\rangle$ when there is no ambiguity. We thus have

$$|\psi\rangle = \sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|\phi\rangle$$

for some state $|\phi\rangle$ such that $\langle 0|\phi\rangle = 0$, and $|\phi\rangle = \sum_{x\neq 0} \alpha_x |x\rangle$ for some $\{\alpha_x\}$. We write down explicitly

$$\psi := |\psi\rangle\langle\psi| = (1-\epsilon)|0\rangle\langle 0| + \sqrt{\epsilon(1-\epsilon)}(|0\rangle\langle\phi| + |\phi\rangle\langle 0|) + \epsilon|\phi\rangle\langle\phi|.$$

By Lemma 17,

$$\operatorname{tr}(\mathcal{D}_\delta^{\otimes n}\,\psi)^2 = \left(\frac{1-\delta^2}{d}\right)^n \sum_{S\subseteq[n]} \gamma^{|S|} \operatorname{tr}\psi_S^2,$$

where we set $\gamma = d\delta^2/(1-\delta^2)$ for brevity. Now

$$\sum_{S\subseteq[n]} \gamma^{|S|} \operatorname{tr}\psi_S^2 = \sum_{S\subseteq[n]} \gamma^{|S|} \left(\operatorname{tr}((1-\epsilon)|0\rangle\langle 0|_S + \sqrt{\epsilon(1-\epsilon)}(|0\rangle\langle\phi|_S + |\phi\rangle\langle 0|_S) + \epsilon|\phi\rangle\langle\phi|_S)^2\right),$$

and for any subset S,

$$\begin{aligned}\operatorname{tr}\psi_S^2 &= (1-\epsilon)^2 \operatorname{tr}|0\rangle\langle 0|_S^2 + \epsilon(1-\epsilon)\operatorname{tr}(|0\rangle\langle\phi| + |\phi\rangle\langle 0|)_S^2 + \epsilon^2 \operatorname{tr}|\phi\rangle\langle\phi|_S^2\\ &+ 2\sqrt{\epsilon}(1-\epsilon)^{3/2}\operatorname{tr}|0\rangle\langle 0|_S(|0\rangle\langle\phi| + |\phi\rangle\langle 0|)_S + 2\epsilon(1-\epsilon)\operatorname{tr}|0\rangle\langle 0|_S|\phi\rangle\langle\phi|_S\\ &+ 2\epsilon^{3/2}\sqrt{1-\epsilon}\operatorname{tr}|\phi\rangle\langle\phi|_S(|0\rangle\langle\phi| + |\phi\rangle\langle 0|)_S.\end{aligned}$$

We now bound the sum over S (weighted by $\gamma^{|S|}$) of each of these terms, in order. Note that we repeatedly use the notation $[E]$ for a term which evaluates to 1 if the expression E is true, and 0 if E is false.

1. As $|0\rangle$ is product, clearly

$$\sum_{S\subseteq[n]} \gamma^{|S|} \operatorname{tr}|0\rangle\langle 0|_S^2 = \sum_{S\subseteq[n]} \gamma^{|S|} = (1+\gamma)^n.$$

2. We have

$$\operatorname{tr}(|0\rangle\langle\phi| + |\phi\rangle\langle 0|)_S^2 = \operatorname{tr}|0\rangle\langle\phi|_S^2 + \operatorname{tr}|\phi\rangle\langle 0|_S^2 + 2\operatorname{tr}|0\rangle\langle\phi|_S|\phi\rangle\langle 0|_S.$$

It is easy to see that the first two terms must be 0 for all S (as only the off-diagonal entries of the first row of the matrix $|0\rangle\langle\phi|$ can be non-zero). For the third, we explicitly calculate

$$|0\rangle\langle\phi|_S|\phi\rangle\langle 0|_S = \sum_{x\neq 0} |\alpha_x|^2 [x_i = 0, \forall i \in \bar{S}]|0\rangle\langle 0|^{\otimes k},$$

and hence

$$\begin{aligned}\sum_{S\subseteq[n]} \gamma^{|S|} \operatorname{tr}|0\rangle\langle\phi|_S|\phi\rangle\langle 0|_S &= \sum_{x\neq 0} |\alpha_x|^2 \sum_{S\subseteq[n]} \gamma^{|S|}[x_i = 0, \forall i \in \bar{S}]\\ &= \sum_{x\neq 0} |\alpha_x|^2 \sum_{k=|x|}^{n} \gamma^k \binom{n-|x|}{n-k}\\ &= (1+\gamma)^n \sum_{x\neq 0} |\alpha_x|^2 \left(\frac{\gamma}{1+\gamma}\right)^{|x|}.\end{aligned}$$

3. It clearly holds that $\operatorname{tr}|\phi\rangle\langle\phi|_S^2 \leq 1$, so as in part (1),
$$\sum_{S\subseteq[n]} \gamma^{|S|} \operatorname{tr}|\phi\rangle\langle\phi|_S^2 \leq (1+\gamma)^n,$$
and this will be tight if and only if $|\phi\rangle$ is product itself.

4. Using the same argument as in part (2), $\operatorname{tr}|0\rangle\langle 0|_S|0\rangle\langle\phi|_S = \operatorname{tr}|0\rangle\langle 0|_S|\phi\rangle\langle 0|_S = 0$.

5. Write the state $\phi = |\phi\rangle\langle\phi|$ as
$$\phi = \sum_{x,y} \phi_{x_1,\dots,y_n}|x_1\rangle\langle y_1| \otimes \cdots \otimes |x_n\rangle\langle y_n|.$$
Then, for any $S = \{i_1,\dots,i_k\}$,
$$\phi_S = \sum_{x,y}[x_i = y_i, \forall i \in \bar{S}]\phi_{x_1,\dots,y_n}|x_{i_1}\rangle\langle y_{i_1}| \otimes \cdots \otimes |x_{i_k}\rangle\langle y_{i_k}|,$$
which implies
$$\operatorname{tr}|0\rangle\langle 0|_S|\phi\rangle\langle\phi|_S = \sum_x [x_i = 0, \forall i \in S]|\alpha_x|^2,$$
and hence, similarly to part (2),
$$\begin{aligned}\sum_{S\subseteq[n]} \gamma^{|S|}\operatorname{tr}|0\rangle\langle 0|_S|\phi\rangle\langle\phi|_S &= \sum_{x\neq 0}|\alpha_x|^2\sum_{k=0}^{n-|x|}\gamma^k\binom{n-|x|}{k}\\ &= (1+\gamma)^n\sum_{x\neq 0}|\alpha_x|^2\left(\frac{1}{1+\gamma}\right)^{|x|}.\end{aligned}$$

6. The last term can be trivially bounded using
$$|\operatorname{tr}|\phi\rangle\langle\phi|_S(|0\rangle\langle\phi| + |\phi\rangle\langle 0|)_S| \leq 2.$$
However, it is possible to get a better bound with a bit more work. We expand
$$\begin{aligned}&\sum_{S\subseteq[n]}\gamma^{|S|}\operatorname{tr}|\phi\rangle\langle\phi|_S|0\rangle\langle\phi|_S =\\ &\sum_{S\subseteq[n]}\gamma^{|S|}\sum_{x,y,z}\alpha_x\alpha_y^*\alpha_z^*[z_i = 0, i\in\bar{S}][x_i = y_i, i \in \bar{S}]\operatorname{tr}|x_1\rangle\langle y_1|0\rangle\langle z_1| \otimes\cdots\otimes |x_n\rangle\langle y_n|0\rangle\langle z_n|\\ &\qquad= \sum_{S\subseteq[n]}\gamma^{|S|}\sum_{x,y,z}\alpha_x\alpha_y^*\alpha_z^*[z_i = 0, i\in\bar{S}][x_i = y_i, i\in\bar{S}][y_i = 0, i \in S][x_i = z_i, i\in S]\\ &\qquad= \sum_{|y\wedge z|=0}\alpha_{y\vee z}\alpha_y^*\alpha_z^*\sum_{S\subseteq[n]}\gamma^{|S|}[y_i = 0, i\in S][z_i = 0, i\in\bar{S}]\\ &\qquad= \sum_{|y\wedge z|=0}\alpha_{y\vee z}\alpha_y^*\alpha_z^*\gamma^{|z|}(1+\gamma)^{n-|y|-|z|}.\end{aligned}$$

This expression can be upper bounded as follows:

$$\begin{aligned}
&\sum_{|y\wedge z|=0} \alpha_{y\vee z}\alpha_y^*\alpha_z^*\gamma^{|z|}(1+\gamma)^{-(|y|+|z|)} \\
&\le \sqrt{\sum_{|y\wedge z|=0}|\alpha_y|^2|\alpha_z|^2}\sqrt{\sum_{|y\wedge z|=0}\frac{\gamma^{2|z|}}{(1+\gamma)^{2|y\vee z|}}|\alpha_{y\vee z}|^2} \\
&\le \left(\sum_x (1+\gamma)^{-2|x|}|\alpha_x|^2\left(\sum_{|y\wedge z|=0}\gamma^{2|z|}[y\vee z=x]\right)\right)^{1/2} \\
&= \left(\sum_x\left(\frac{1+\gamma^2}{(1+\gamma)^2}\right)^{|x|}|\alpha_x|^2\right)^{1/2}.
\end{aligned} \tag{20}$$

Combining these terms, we have

$$\sum_{S\subseteq[n]}\gamma^{|S|}\operatorname{tr}\psi_S^2 \le (1+\gamma)^n((1-\epsilon)^2+2\epsilon(1-\epsilon)\sum_{x\neq 0}|\alpha_x|^2(1+\gamma)^{-|x|}(\gamma^{|x|}+1)+\epsilon^2+$$
$$4\epsilon^{3/2}\sqrt{1-\epsilon}\left(\sum_x\left(\frac{1+\gamma^2}{(1+\gamma)^2}\right)^{|x|}|\alpha_x|^2\right)^{1/2}).$$

Note that $(1+\gamma)^{-|x|}(\gamma^{|x|}+1)$ decreases with $|x|$ for all $\gamma>0$, as does $(1+\gamma^2)^{|x|}(1+\gamma)^{-2|x|}$. To complete the proof, we will show that $|\phi\rangle$ has no weight 1 components (i.e. $\alpha_x=0$ for $|x|<2$). In the contribution from Eq. (20), this implies that only the $|x|\ge 4$ terms contribute (since $x=y\vee z$ and $y\wedge z=\emptyset$). Therefore, $|\phi\rangle$ having no weight 1 components would imply that

$$\sum_{S\subseteq[n]}\gamma^{|S|}\operatorname{tr}\psi_S^2 \le (1+\gamma)^n\left(1-\frac{4\epsilon}{(1+\gamma)^2}\left(\gamma(1-\epsilon)-\left(\frac{(1+\gamma^2)^2}{(1+\gamma)^2}\right)\epsilon^{1/2}\right)\right),$$

which would imply the theorem. Now, for any θ, φ, we have $1-\epsilon\ge|(\cos\theta\langle 0|+e^{i\varphi}\sin\theta\langle 1|)\otimes\langle 0|^{\otimes n-1}|\psi\rangle|^2$. Picking θ such that

$$\cos\theta=\frac{|\langle 0|\psi\rangle|}{\sqrt{|\langle 0|\psi\rangle|^2+|\langle 10^{n-1}|\psi\rangle|^2}},$$

and φ such that $e^{i\varphi}\langle 10^{n-1}|\psi\rangle>0$, it is easy to see that

$$1-\epsilon\ge|\cos\theta\langle 0|\psi\rangle+e^{i\varphi}\sin\theta\langle 10^{n-1}|\psi\rangle|^2=|\langle 0|\psi\rangle|^2+|\langle 10^{n-1}|\psi\rangle|^2.$$

However, we have assumed that $1-\epsilon=|\langle 0|\psi\rangle|^2$, so this implies that $\langle 10^{n-1}|\psi\rangle=0$. Repeating the argument for the other $n-1$ subsystems shows that $|\psi\rangle$ is indeed orthogonal to every state with Hamming weight at most 1, so $|\phi\rangle$ has no weight 1 components. □

B Proof of Theorem 1: correctness of the product test

In this appendix, we prove correctness of the product test (Theorem 1). Let the test be defined as in Protocol 1. The following lemma from [59] expresses the probability of passing in terms of the partial traces of the input states; we include a proof for completeness.

Lemma 2. *Let $P_{test}(\rho, \sigma)$ denote the probability that the product test passes when applied to two mixed states $\rho, \sigma \in \mathcal{B}(\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n})$. Define $P_{test}(\rho) := P_{test}(\rho, \rho)$. Then*

$$P_{test}(\rho, \sigma) = \frac{1}{2^n} \sum_{S \subseteq [n]} \operatorname{tr} \rho_S \sigma_S,$$

and in particular

$$P_{test}(\rho) = \frac{1}{2^n} \sum_{S \subseteq [n]} \operatorname{tr} \rho_S^2.$$

If $d_1 = d_2 = \cdots = d_n = d$, for some d, then

$$P_{test}(\rho) = \left(\frac{d+1}{2} \right)^n \operatorname{tr}(\mathcal{D}_{1/\sqrt{d+1}}^{\otimes n}\, \rho)^2.$$

Note that we can in fact assume that $d_1 = d_2 = \cdots = d_n = d$ without loss of generality by setting $d = \max(d_1, \dots, d_n)$, and embedding each of $\mathbb{C}^{d_1}, \dots, \mathbb{C}^{d_n}$ into $\mathbb{C}^d$ in the natural way. This padding operation neither affects the probability of the swap tests passing nor changes the distance to the closest product state.

Proof. Let $\mathcal{F}$ denote the swap (or flip) operator that exchanges two quantum systems of equal but arbitrary dimension, with $\mathcal{F}_S$ denoting the operator that exchanges only the qudits in the set S. Then we have

$$P_{\text{test}}(\rho, \sigma) = \operatorname{tr}(\rho \otimes \sigma) \left(\frac{I + \mathcal{F}}{2} \right)^{\otimes n} = \frac{1}{2^n} \sum_{S \subseteq [n]} \operatorname{tr}(\rho \otimes \sigma)\, \mathcal{F}_S = \frac{1}{2^n} \sum_{S \subseteq [n]} \operatorname{tr} \rho_S \sigma_S.$$

The second part then follows from Lemma 17. □

We now analyse the probability of the product test passing for general n. We first note that, in the special case where $n = 2$, it is possible to analyse the probability of passing quite tightly. The proof of the following result, which is implicit in previous work of Wei and Goldbart [71], is essentially immediate from Lemma 2.

Lemma 20. *Let $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$, where $d_1 \leq d_2$, be a bipartite pure state with Schmidt coefficients $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_{d_1}}$. Then*

$$P_{test}(|\psi\rangle\langle\psi|) = \frac{1}{2} \left(1 + \sum_i \lambda_i^2 \right),$$

while

$$1 - \epsilon := \max_{|\phi_1\rangle, |\phi_2\rangle} |\langle\psi|\phi_1\rangle|\phi_2\rangle|^2 = \lambda_1.$$

In particular,

$$1 - \epsilon + \frac{d_1}{2(d_1 - 1)} \epsilon^2 \leq P_{test}(|\psi\rangle\langle\psi|) \leq 1 - \epsilon + \epsilon^2.$$

We are finally ready to prove Theorem 1. The proof is split into two parts, which we formalise as separate theorems. The first part holds when ϵ is small, and depends on the results proven in Appendix A. The second part holds when ϵ is large, and is proved using the first part.

Theorem 3. *Given $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, let*

$$1 - \epsilon = \max\{|\langle\psi|\phi_1, \dots, \phi_n\rangle|^2 : |\phi_i\rangle \in \mathbb{C}^{d_i}, 1 \le i \le n\}.$$

Then

$$1 - 2\epsilon + \epsilon^2 \le P_{test}(|\psi\rangle\langle\psi|) \le 1 - \epsilon + \epsilon^2 + \epsilon^{3/2}.$$

Proof. The lower bound holds by general arguments. It is immediate that, if applied to $|\phi_1, \dots, \phi_n\rangle$, the product test succeeds with probability 1. As the test acts on two copies of $|\psi\rangle$, which has overlap $1 - \epsilon$ with $|\phi_1, \dots, \phi_n\rangle$, it must succeed when applied to $|\psi\rangle$ with probability at least $(1 - \epsilon)^2$. The upper bound follows from Lemma 2 and Theorem 18. The statement of Theorem 18 only explicitly covers the case where the dimensions of all the subsystems are the same; however, as noted above, we can assume this without loss of generality. □

This result is close to optimal. At the low end, the state $|\psi\rangle = \sqrt{1-\epsilon}|0^n\rangle + \sqrt{\epsilon}|1^n\rangle$ has $P_{\text{test}}(|\psi\rangle\langle\psi|) = 1 - 2\epsilon + 2\epsilon^2 + o(1)$. At the high end, for $|\psi\rangle = \sqrt{1-\epsilon}|00\rangle + \sqrt{\epsilon}|11\rangle$, $P_{\text{test}}(|\psi\rangle\langle\psi|) = 1 - \epsilon + \epsilon^2$. We also note that this result does not extend to a test for separability of mixed states; the maximally mixed state on n qudits is separable but it is easy to verify that $P_{\text{test}}(I/d^n) = ((d+1)/2d)^n$, which approaches zero for large n.

Theorem 3 only gives a non-trivial upper bound on the probability of passing when ϵ is small (up to $\epsilon = \frac{1}{2}(3 - \sqrt{5}) \approx 0.38$). We now show that the product test also works in the case where the state under consideration is far from any product state. We will need two lemmas.

Lemma 21. *Given $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, let $P^P_{test}(|\psi\rangle\langle\psi|)$ be the probability that the P-product test – the test for being product across partition P – passes. Then, for all P, $P^P_{test}(|\psi\rangle\langle\psi|) \le P_{test}(|\psi\rangle\langle\psi|)$.*

Proof. The subspace corresponding to the usual product test passing is contained within the subspace corresponding to the P-product test passing. □

Lemma 22. *Let $|\psi\rangle$, $|\phi\rangle$ be pure states such that $|\langle\psi|\phi\rangle|^2 = 1 - \epsilon$, and let P satisfy $0 \le P \le I$; e.g. P might be a projector. Then $|\langle\psi|P|\psi\rangle - \langle\phi|P|\phi\rangle| \le \sqrt{\epsilon}$.*

Proof. We can directly calculate $\frac{1}{2}\|\,|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\,\|_1 = \sqrt{\epsilon}$. This then gives the claimed upper bound on $|\operatorname{tr} P(|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|)|$ (see [60, Chapter 9]). □

Theorem 4. *Given $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, let*

$$1 - \epsilon = \max\{|\langle\psi|\phi_1, \dots, \phi_n\rangle|^2 : |\phi_i\rangle \in \mathbb{C}^{d_i}, 1 \le i \le n\}.$$

Then, if $\epsilon \ge 11/32 > 0.343$, $P_{test}(|\psi\rangle\langle\psi|) \le 501/512 < 0.979$.

Proof. For simplicity, in the proof we will use a quadratic upper bound on $P_{\text{test}}(|\psi\rangle\langle\psi|)$ that follows by elementary methods from Theorem 1: $P_{\text{test}}(|\psi\rangle\langle\psi|) \le 1 - \frac{3}{4}\epsilon + 2\epsilon^2$. For a contradiction, assume that $P_{\text{test}}(|\psi\rangle\langle\psi|) > p := 501/512$, while $\epsilon \ge 11/32$.

For any partition P of $[n]$ into $1 \le k \le n$ parts, let $|\phi_P\rangle$ be the product state (with respect to P) that maximises $|\langle\psi|\phi\rangle|^2$ over all product states $|\phi\rangle$ (with respect to P). If

$$1 - h \le |\langle\psi|\phi_P\rangle|^2 \le 1 - \ell,$$

where for readability we define $\ell := 1/32$ and $h := 11/32$, then by the quadratic bound given above the P-product test passes with probability $P^P_{\text{test}}(|\psi\rangle\langle\psi|) \le p$, implying by Lemma 21 that

$P_{\text{test}}(|\psi\rangle\langle\psi|) \leq p$. Therefore, as we are assuming that $|\psi\rangle$ is a counterexample to the present theorem, there exists a k such that $|\langle\psi|\phi\rangle|^2 > 1-\ell$ for some $|\phi\rangle$ that is product across k parties, and yet $|\langle\psi|\phi\rangle|^2 < 1-h$ for all $|\phi\rangle$ that are product across $k+1$ parties.

So, for this k, let $|\phi_1\rangle \cdots |\phi_k\rangle$ be a state that maximises $|\langle\psi|\phi_1,\dots,\phi_k\rangle|^2$. Thus there is some $\epsilon' < \ell$ such that we can write $|\psi\rangle$ as

$$|\psi\rangle = \sqrt{1-\epsilon'}|\phi_1\rangle\cdots|\phi_k\rangle + \sqrt{\epsilon'}|\xi\rangle.$$

If $k=1$, then trivially $|\phi_1\rangle = |\psi\rangle$ and $\epsilon' = 0$. Assume without loss of generality that $|\phi_1\rangle$ is a state of two or more qudits. Now we know that

$$\max_{|\phi'_{1,1}\rangle,|\phi'_{1,2}\rangle} |\langle\phi_1|\phi'_{1,1}\rangle|\phi'_{1,2}\rangle|^2(1-\epsilon') < 1-h, \tag{21}$$

or $|\phi'_{1,1}\rangle|\phi'_{1,2}\rangle|\phi_2\rangle\cdots|\phi_k\rangle$ would be a $(k+1)$-partite state with overlap at least $1-h$ with $|\psi\rangle$. (Here we have used the fact that for $k>1$, by the arguments at the end of Theorem 18, $|\xi\rangle$ is orthogonal to $|\phi'_{1,1}\rangle|\phi'_{1,2}\rangle|\phi_2\rangle\cdots|\phi_k\rangle$ for any choice of $|\phi'_{1,1}\rangle$, $|\phi'_{1,2}\rangle$.) Let $1-\delta = \max_{|\phi'_{1,1}\rangle,|\phi'_{1,2}\rangle} |\langle\phi_1|\phi'_{1,1}\rangle|\phi'_{1,2}\rangle|^2$. Then Eq. (21) implies that

$$1-\delta < \frac{1-h}{1-\epsilon'} < \frac{1-h}{1-\ell} = \frac{21}{31}.$$

Using the exact expression given in Lemma 20, we find that $P_{\text{test}}(|\phi_1\rangle\langle\phi_1|) < 751/961$ (if $10/31 < \delta \leq 1/2$, this follows from $P_{\text{test}}(|\phi_1\rangle\langle\phi_1|) \leq 1+\delta+\delta^2$; if $\delta \geq 1/2$, then $P_{\text{test}}(|\phi_1\rangle\langle\phi_1|) \leq 3/4$ always). Next we use Lemma 22 to obtain

$$\begin{aligned} P_{\text{test}}(|\psi\rangle\langle\psi|) &\leq P_{\text{test}}(|\phi_1\rangle\langle\phi_1| \otimes \cdots \otimes |\phi_k\rangle\langle\phi_k|) + \sqrt{\epsilon'} \\ &< P_{\text{test}}(|\phi_1\rangle\langle\phi_1|) + \sqrt{\ell} \\ &< \frac{751}{961} + \sqrt{\frac{1}{32}} < 0.96. \end{aligned}$$

But we previously assumed that $P_{\text{test}}(|\psi\rangle\langle\psi|) > p > 0.978$. We have reached a contradiction, so the proof is complete. □

One might hope that this theorem could be improved to show that, as $\epsilon \to 1$, $P_{\text{test}}(|\psi\rangle\langle\psi|)$ necessarily approaches 0. However, this is not possible. Consider the $d \times d$-dimensional bipartite state $|\Phi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |ii\rangle$. It is easy to verify using Lemma 20 that $P_{\text{test}}(|\Phi\rangle\langle\Phi|) = 1/2(1+1/d)$ while $\max_{|\phi_1\rangle,|\phi_2\rangle} |\langle\Phi|\phi_1\rangle|\phi_2\rangle|^2 = 1/d$.

Combining Theorems 3 and 4, we obtain Theorem 1 and thus have proven correctness of the product test. The constants in Theorem 4 have not been optimised as far as possible and could be improved somewhat.

C Classes of measurement operators

In this appendix, we define the classes of measurement operators used in our paper and other relevant literature on QMA(2), such as [15]. Our definitions mostly follow the conventions of quantum information theory. Each class of measurement operators describes operators on $\mathbb{C}^d \otimes \mathbb{C}^d$.

- BELL is the set of M that can be expressed as

$$M = \sum_{(i,j)\in S} \alpha_i \otimes \beta_j, \tag{22}$$

  where $\sum_i \alpha_i = I$ and $\sum_j \beta_j = I$, and S is a set of pairs of indices. In other words, the systems are locally measured, obtaining outcomes i and j, and then the verifier accepts if $(i, j) \in S$.

- LOCC_1 is the set of M that can be realised by measuring the first system and then choosing a measurement on the second system conditional on the outcome of the first measurement. Such M can be written as

$$M = \sum_i \alpha_i \otimes M_i, \tag{23}$$

  where $\sum_i \alpha_i = I$ and $0 \leq M_i \leq I$ for each i.

- LOCC is the set of M that can be realised by alternating partial measurements on the two systems a finite number of times, choosing each measurement conditioned on the previous outcomes. An inductive definition is that M is in LOCC if there exist operators $\{E_i\}, \{M_i\}$, with $\sum_i E_i \leq I$ and each $M_i \in \text{LOCC}$, such that either $M = \sum_i(\sqrt{E_i} \otimes I)M_i(\sqrt{E_i} \otimes I)$ or $M = \sum_i(I \otimes \sqrt{E_i})M_i(I \otimes \sqrt{E_i})$. For the base case, it suffices to take $I \in \text{LOCC}$.

- SEP is the set of M such that

$$M = \sum_i \alpha_i \otimes \beta_i \tag{24}$$

  for some positive semidefinite (WLOG rank one) matrices $\{\alpha_i\}, \{\beta_i\}$. (Note: other works define SEP to be the smaller set of M for which both M and $I - M$ can be decomposed as in Eq. (24), and use the term SEP_{YES} to describe the measurements for which only M has to satisfy Eq. (24).)

- SEP-BOTH is the set of M for which $M \in \text{SEP}$ and $I - M \in \text{SEP}$.

- PPT (positive partial transpose [62, 43]) is the set of M for which $M^\Gamma \geq 0$, where Γ is the partial transpose map defined by $(|i\rangle\langle j| \otimes |k\rangle\langle l|)^\Gamma = (|i\rangle\langle j| \otimes |l\rangle\langle k|)$. Again note that this definition does not require $I - M \in \text{PPT}$.

- PPT-BOTH is the set of M for which $M \in \text{PPT}$ and $I - M \in \text{PPT}$.

- ALL has no restrictions on M other than $0 \leq M \leq I$.

We note that SEP-BOTH and PPT-BOTH are natural relaxations of LOCC because they preserve the property that both M and $I - M$ must be realisable through local operations and classical communication. On the other hand, SEP and PPT are more natural when we consider M by itself and do not wish to consider additional constraints on $I - M$.

These sets satisfy the following inclusions, all of which are known to be strict

$$\begin{array}{ccccccccccc}
\text{BELL} & \subset & \text{LOCC}_1 & \subset & \text{LOCC} & \subset & \text{SEP-BOTH} & \subset & \text{PPT-BOTH} & & \\
 & & & & & & \cap & & \cap & & \\
 & & & & & & \text{SEP} & \subset & \text{PPT} & \subset & \text{ALL}
\end{array}$$

D Nonexistence of an LOCC product test

A natural extension of the idea of product state testing is to a distributed setting where two parties, each of whom receives one copy of an n-partite state $|\psi\rangle$, must determine whether $|\psi\rangle$ is product using only local operations and classical communication (LOCC). Indeed, following the completion of an initial version of this work, it was shown by Brandão, Christandl and Yard that, if there were an efficient LOCC protocol for product state testing, then $\mathsf{QMA}(k) = \mathsf{QMA}$ [15, 16].

In this appendix, we show that unfortunately no such LOCC protocol exists. In fact, we rule out the larger class of PPT-BOTH measurements (defined in Appendix C). Our impossibility result holds for the easiest version of this task, in which $n = 2$. For simplicity, here we only consider the case where the test uses 2 copies of $|\psi\rangle$; one can show a similar result when the number of copies is larger but the proof is significantly more complicated [39].

Formally, we define a product test as a measurement $\{M, I - M\}$ that acts on $|\psi\rangle^{\otimes 2} = |\psi\rangle^{A_1B_1} \otimes |\psi\rangle^{A_2B_2}$ with outcome M corresponding to "product" and $I - M$ corresponding to "not product." There is no good canonical way to express the validity of a product test. One rather general way we might do this is to say that there are functions $f(\epsilon)$ and $g(\epsilon)$ such that if $|\psi\rangle$ has overlap $1 - \epsilon$ with the closest product state then its probabity P_{test} of passing the product test satisfies

$$f(\epsilon) \leq P_{\text{test}} \leq g(\epsilon). \tag{25}$$

For example, Theorem 1 shows that our product test satisfies Eq. (25) with $f(\epsilon) = 1 - c_1\epsilon$ and $g(\epsilon) = 1 - c_2\epsilon$ with $c_1 > c_2 > 0$.

For our impossibility result, we will use a different and simpler success measure. Define the completeness c of a product test to be the average probability of accepting a random product state $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$, and define the soundness s to be the average probability of accepting a random bipartite state $|\psi\rangle$. While strictly speaking a random bipartite state may sometimes be close to a product state, it has an overwhelmingly high probability of being close to maximally entangled. Thus, this demand for a product test is nearly as undemanding as possible. Finally, define the bias of the test as $b = c - s$.

Theorem 23. *Any 2-copy PPT-BOTH product test for bipartite $d \times d$-dimensional product states has bias which is $O(1/d)$.*

Proof. Let $|\psi\rangle$ be a bipartite $d \times d$-dimensional state on the system AB. Imagine we have a protocol which takes as input two copies of $|\psi\rangle$, written $|\psi\rangle_1|\psi\rangle_2$, and attempts to determine whether $|\psi\rangle$ is product across systems A and B. Consider two distributions $\mathcal{D}_0$, $\mathcal{D}_1$ on bipartite $d \times d$ states. Let M be a measurement operator which accepts states drawn from $\mathcal{D}_1$ with probability at least c, and rejects states from $\mathcal{D}_0$ with probability at least s. Then

$$\mathbb{E}_{\psi\sim\mathcal{D}_1}\mathbb{E}_{\phi\sim\mathcal{D}_0} \operatorname{tr} M(\psi \otimes \psi - \phi \otimes \phi) \geq c - s,$$

implying

$$\operatorname{tr} M(\mathbb{E}_{\psi\sim\mathcal{D}_1}(\psi \otimes \psi) - \mathbb{E}_{\phi\sim\mathcal{D}_0}(\phi \otimes \phi)) \geq c - s.$$

Taking $\mathcal{D}_1$ to be the uniform distribution over product states, via a standard calculation we obtain

$$\mathbb{E}_{\psi\sim\mathcal{D}_1}(\psi \otimes \psi) = \mathbb{E}_\psi\mathbb{E}_\phi(\psi_A \otimes \phi_B)^{\otimes 2} = \left(\frac{1}{d(d+1)}(I + \mathcal{F})_{A_1A_2}\right) \otimes \left(\frac{1}{d(d+1)}(I + \mathcal{F})_{B_1B_2}\right),$$

where as elsewhere $\mathcal{F}$ is the swap operator. Now let $\mathcal{D}_0$ be the uniform distribution on bipartite $d \times d$ states. In this case we have

$$\mathbb{E}_{\phi\sim\mathcal{D}_0}(\phi\otimes\phi) = \frac{1}{d^2(d^2+1)}(I+\mathcal{F})_{12}.$$

Thus

$$\begin{aligned}\Delta &:= \mathbb{E}_{\psi\sim\mathcal{D}_1}(\psi\otimes\psi) - \mathbb{E}_{\phi\sim\mathcal{D}_0}(\phi\otimes\phi)\\ &= \frac{1}{d^2(d+1)^2}\left(I_{A_1A_2}\otimes\mathcal{F}_{B_1B_2} + \mathcal{F}_{A_1A_2}\otimes I_{B_1B_2}\right) - \frac{2}{d(d^2+1)(d+1)^2}\left(I+\mathcal{F}_{12}\right).\end{aligned}$$

Assume that M is PPT across the 1:2 split. We want to maximise $\operatorname{tr} M\Delta$ assuming that $-I \leq M \leq I$ and $-I \leq M^\Gamma \leq I$, where the second is the PPT constraint. Further, as distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ are invariant under product unitaries, we can without loss of generality assume that M commutes with $U_{A_1}\otimes U_{A_2}\otimes V_{B_1}\otimes V_{B_2}$ for all unitaries U and V, which implies that

$$M = wI + x(\mathcal{F}_{A_1A_2}\otimes I_{B_1B_2}) + y(I_{A_1A_2}\otimes\mathcal{F}_{B_1B_2}) + z\mathcal{F}_{12}$$

for some w, x, y, z. By direct calculation

$$\begin{aligned}\operatorname{tr} M\Delta &= \frac{1}{d^2(d+1)^2}\left(2d^3w + (d^2+d^4)x + (d^2+d^4)y + 2d^3z\right) + O(1/d)\\ &= x + y + O(1/d).\end{aligned}$$

On the other hand,

$$M^\Gamma = wI + xd(\Phi_{A_1A_2}\otimes I_{B_1B_2}) + yd(I_{A_1A_2}\otimes\Phi_{B_1B_2}) + zd^2(\Phi_{A_1A_2}\otimes\Phi_{B_1B_2}),$$

where $|\Phi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |ii\rangle$. So $|x+y| = O(1/d)$, and we are done. □

E Proof of correctness of the protocol to put $\mathsf{QMA}(k)$ in $\mathsf{QMA}(2)$

This section proves several of the claims made in Section 3. First we prove Lemma 5 by showing the validity of Protocol 2.

Lemma 5 (restatement). *For any m, k, $0 \leq s < c \leq 1$,*

$$\mathsf{QMA}_m(k)_{s,c} \subseteq \mathsf{QMA}_{km}(2)_{s',c'}$$

where $c' = \frac{1+c}{2}$ and $s' = 1 - \frac{(1-s)^2}{100}$.

Proof. It is obvious that this protocol achieves completeness $(1+c)/2$: if the Merlins follow the protocol, the product test passes with certainty, so Arthur either accepts with probability 1, or with the same probability that $\mathcal{A}$ accepts, which is at least c. Showing soundness is somewhat more complicated.

Assume that Arthur receives states $|\phi_1\rangle$ and $|\phi_2\rangle$ such the maximal overlap of $|\phi_1\rangle$ (resp. $|\phi_2\rangle$) with a product state is $1-\epsilon_1$ (resp. $1-\epsilon_2$), and set $\epsilon = \frac{1}{2}(\epsilon_1+\epsilon_2)$. Further assume that the product test would accept $|\phi_1\rangle^{\otimes 2}$ (resp. $|\phi_2\rangle^{\otimes 2}$) with probability $1-\delta_1$ (resp. $1-\delta_2$).

Let $1-\delta$ be the probability that the product test would accept $|\phi_1\rangle \otimes |\phi_2\rangle$. We first show that this can be upper bounded in terms of the probabilities of accepting $|\phi_1\rangle^{\otimes 2}$ and $|\phi_2\rangle^{\otimes 2}$. The probability that the product test accepts is

$$\begin{aligned}
\frac{1}{2^k}\sum_{S\subseteq[k]} \mathrm{tr}(\phi_1)_S(\phi_2)_S &\leq \frac{1}{2^k}\sum_{S\subseteq[k]} \sqrt{\mathrm{tr}(\phi_1)_S^2}\sqrt{\mathrm{tr}(\phi_2)_S^2} \\
&\leq \frac{1}{2^k}\sum_{S\subseteq[k]} \frac{\mathrm{tr}(\phi_1)_S^2+\mathrm{tr}(\phi_2)_S^2}{2} \\
&= \frac{1}{2}\left(P_{\mathrm{test}}(\phi_1)+P_{\mathrm{test}}(\phi_2)\right) \\
&= 1-\frac{1}{2}\left(\delta_1+\delta_2\right).
\end{aligned}$$

Thus we have the bound from Theorem 1 that $\delta \geq \frac{11}{512}\epsilon$. On the other hand, if Arther chooses to measure M, then by Lemma 22 his probability of accepting is $\leq s+\frac{\sqrt{\epsilon_1}+\sqrt{\epsilon_2}}{2} \leq s+\sqrt{\epsilon}$. Combining the two tests, we find that the acceptance probability is

$$s' \leq \max_{\epsilon\leq\frac{512}{11}\delta} \frac{1-\delta+\min(1,s+\sqrt{\epsilon})}{2} \leq 1-\frac{(1-s)^2}{100}. \tag{26}$$

To obtain the last inequality, we observe that the worst case is obtained when $\sqrt{\epsilon}=1-s=\sqrt{\frac{512}{11}\delta}$.

As a result of Eq. (26), a k-prover soundness-s protocol can be simulated by a 2-prover protocol with soundness s'. If $k \leq \mathrm{poly}(n)$, then the messages will still have a polynomial number of qubits. □

Remark. Our choice of protocol in Protocol 2 was carefully designed to yield a separable protocol. One subtlety in doing so is that measurements in SEP do not compose the same way that separable operations do. Indeed, if $A,B \in$ SEP, then it does *not* follow that $A^{1/2}BA^{1/2} \in$ SEP, even in the case when A is a projector. For example, let A project onto the symmetric subspace and let $B=|0,1\rangle\langle 0,1|$. On the other hand, other choices of protocol could also yield separable measurements. For example, $(M\otimes I)P(M\otimes I)$ would work (after using Lemma 8 to amplify completeness), where P denotes the product test, M is a measurement on Alice's k systems and I acts on Bob's k systems. We are grateful to Fernando Brandão for pointing out this issue to us.

Next, we prove Lemma 7.

Lemma 7 (restatement). *For any $\ell \geq 1$,*

$$\mathsf{QMA}_m^{\mathrm{SEP}}(k)_{s,c} \subseteq \mathsf{QMA}_{\ell m}^{\mathrm{SEP}}(k)_{s^\ell,c^\ell}.$$

Proof. In the original protocol, Arthur performs a measurement M on k states, each comprising m qubits. If the input is a YES instance, then there exists a product state on which M has expectation value $\geq c$, whereas if the input is a NO instance, then for all product states, M has expectation value $\leq s$.

For the modified protocol, each of the k provers submits ℓm qubits, and Arthur's measurement is $M^{\otimes\ell}$. If the input is a YES instance, then the provers can submit ℓ copies of the optimal input to the original protocol. This state is still product (across the k provers) and has probability $\geq c^\ell$ of being accepted.

The more interesting case is when the input is a NO instance. In general, the provers can submit states that are entangled across the ℓ different parts of their message. However, there cannot be any entanglement between the k different provers.

For the new protocol, we can imagine Arthur sequentially performing the measurement $\{M, I - M\}$ ℓ times and accepting only if the outcome is M each time. Since the input is a NO instance, if this measurement is applied to a state, pure or mixed, that is separable across the k parties, then outcome M will occur with probability $\leq s$. Therefore the probability that all ℓ measurements have outcome M will be $\leq s^\ell$ as long as conditioning on outcome M does not induce any entanglement in the unmeasured states. We need not consider the state that remains after outcome $I - M$, since Arthur rejects immediately when this outcome occurs.

The final step of the proof is to show that applying the measurement $\{M, I - M\}$ to the first system of a multipartite product state and obtaining outcome M will not create any entanglement across the k provers. Suppose that the i^{th} Merlin supplies the ℓm-qubit state $|\varphi^{(i)}\rangle^{P_1^i \cdots P_\ell^i}$, where each P_j^i is an m-qubit system. The original measurement M is separable, so can be written as

$$M = \sum_j a_j |\alpha_j^{(1)}\rangle\langle\alpha_j^{(1)}| \otimes \cdots \otimes |\alpha_j^{(k)}\rangle\langle\alpha_j^{(k)}|,$$

for $a_j \geq 0$ and $|\alpha_j^{(i)}\rangle$ unit vectors on m qubits. This measurement is applied sequentially to $P_1^{1,\ldots,k}$, then $P_2^{1,\ldots,k}$, and so on until $P_\ell^{1,\ldots,k}$. (Here we write $P_j^{1,\ldots,k}$ as shorthand for $P_j^1 \cdots P_j^k$.) When the measurement is applied to $P_1^{1,\ldots,k}$ and the outcome is M, the residual state is proportional to

$$\begin{aligned}\operatorname{tr}_{P_1^{1,\ldots,k}} M^{P_1^{1,\ldots,k}} \left(|\varphi^{(1)}\rangle\langle\varphi^{(1)}|^{P^1_{1,\ldots,\ell}} \otimes \cdots \otimes |\varphi^{(k)}\rangle\langle\varphi^{(k)}|^{P^k_{1,\ldots,\ell}} \right) \\ = \sum_j a_j \bigotimes_{i=1}^k \langle\alpha_j^{(1)}|^{P_1^i} \otimes I^{P^i_{2,\ldots,\ell}} |\varphi^{(i)}\rangle\langle\varphi^{(i)}|^{P^i_{1,\ldots,\ell}} |\alpha_j^{(1)}\rangle^{P_1^i} \otimes I^{P^i_{2....,\ell}},\end{aligned}$$

which is separable across the $P^1_{2,\ldots,\ell} : P^2_{2,\ldots,\ell} : \cdots : P^k_{2,\ldots,\ell}$ cut. Therefore, by induction, the state always remain separable as long as outcome M always occurs. □

For the reader's convenience, we briefly summarise here the proof of Lemma 8 (originally due to [50, 2]).

Lemma 8 (restatement). *For any $\ell \geq 1$,*

$$\mathsf{QMA}_m(k)_{s,c} \subseteq \mathsf{QMA}_{\ell m}(k)_{1-\frac{c-s}{3}, 1-\exp(-\frac{\ell(c-s)^2}{2})}.$$

Proof. The idea is to repeat the basic protocol ℓ times, and accept if there are $\geq \frac{c+s}{2}\ell$ "accept" outcomes or reject otherwise. For YES instances, the provers can send ℓ copies of the same proofs, each of which will be accepted with probability $\geq c$. Then a Chernoff bound yields that the completeness is $\geq 1 - \exp(-\ell(c-s)^2/2)$. For NO instances, each of the ℓ copies has probability $\leq s$ of being accepted. Since the provers may submit entangled states, we can no longer guarantee that these events are independent, but still Markov's inequality implies that the probability of $\geq \frac{c+s}{2}\ell$ accept outcomes is $\leq \frac{2s}{c+s} = \frac{1}{1+(c-s)/2} \leq 1 - \frac{c-s}{3}$. This last step uses the fact that $1/(1 + x/2) \leq 1 - x/3$, for $0 \leq x \leq 1$. □

We now complete the proof of Theorem 9.

Theorem 9 (restatement). *1. If $s \leq 1 - 1/\operatorname{poly}(n)$, $k = \operatorname{poly}(n)$ and $p(n)$ is an arbitrary polynomial, then $\mathsf{QMA}(k)_{s,1} = \mathsf{QMA}^{\mathrm{SEP}}(2)_{\exp(-p(n)),1}$.*

2. If $c - s \geq 1/\operatorname{poly}(n)$, $c < 1$, $k = \operatorname{poly}(n)$ and $p(n)$ is an arbitrary polynomial, then $\mathsf{QMA}(k)_{s,c} = \mathsf{QMA}^{\mathrm{SEP}}(2)_{\exp(-p(n)),1-\exp(-p(n))}$.

Proof. For the case of perfect completeness, we use Lemma 6 to replace the original k-prover protocol with a 2-prover protocol that still has perfect completeness, still has $1 - 1/q(n)$ soundness (where $q(n)$ is still a polynomial of n) and now has a separable measurement performed by the verifier. Now we simply repeat (using Lemma 7) $q(n) \cdot p(n)$ times, and obtain soundness $\leq (1 - 1/q(n))^{q(n)p(n)} \leq e^{-p(n)}$.

When we merely have $c - s > 1/q(n)$, for some polynomial $q(n)$, then we first have to apply Lemma 8 with $\ell = p(n) + \log(243q(n)^2 p(n))$ to replace the soundness with $1 - 1/3q(n)$ and the completeness with $1 - \frac{\exp(-p(n))}{243q(n)^2 p(n)}$. Next, we apply Lemma 6 to leave the completeness the same, reduce the number of provers to 2, guarantee the measurement is in SEP and replace the soundness with $1 - 1/243q(n)^2$. Finally, we repeat $243q(n)^2 p(n)$ times and obtain soundness $\exp(-p(n))$ and completeness $1 - \exp(-p(n))$. □

F Proof of correctness of the product unitary test

This appendix is devoted to the proof of Theorem 19. In order to analyse the product unitary test in Protocol 3, we will need to relate the maximum overlap of an n-qudit unitary with a product operator to the maximum overlap of that unitary with a product unitary.

Lemma 24. *Given $U \in U(d^n)$, let*

$$1 - \epsilon = \max\{|\langle U, A_1 \otimes \cdots \otimes A_n \rangle|^2 : A_i \in M(d), \langle A_i, A_i \rangle = 1, 1 \leq i \leq n\}.$$

Then, if $\epsilon \leq 1/2$, there exist $V_1, \dots, V_n \in U(d)$ such that $|\langle U, V_1 \otimes \cdots \otimes V_n \rangle|^2 \geq (1 - 2\epsilon)^2$.

Proof. For all $1 \leq i \leq n$, let the polar decomposition of A_i be $|A_i| C_i$, where $|A_i| = \sqrt{A_i A_i^\dagger}$ and $C_i \in U(d)$. Set $A = \bigotimes_{i=1}^n A_i$, $C = \bigotimes_{i=1}^n C_i$. Then

$$\langle C, A \rangle = \frac{1}{d^n} \prod_{i=1}^n \operatorname{tr} C_i^\dagger |A_i| C_i = \frac{1}{d^n} \prod_{i=1}^n \operatorname{tr} |A_i| = \frac{1}{d^n} \max_{V \in U(d^n)} |\operatorname{tr} V A| \geq \sqrt{1 - \epsilon}.$$

This implies that we can expand

$$U = \sqrt{1-\epsilon}\, A + D, \;\; C = \sqrt{1 - \epsilon'}\, A + E$$

for some $\epsilon' \leq \epsilon$ and matrices D, E such that $\langle D, D \rangle = \epsilon$, $\langle E, E \rangle = \epsilon'$, $\langle A, D \rangle = 0$, $\langle A, E \rangle = 0$. So

$$|\langle U, C \rangle| = |\sqrt{1-\epsilon}\sqrt{1-\epsilon'} + \langle D, E \rangle| \geq |\sqrt{1-\epsilon}\sqrt{1-\epsilon'} - \sqrt{\epsilon}\sqrt{\epsilon'}| \geq 1 - 2\epsilon,$$

for $\epsilon \leq 1/2$. This implies the lemma. □

We are now ready to prove correctness of the product unitary test.

Theorem 19 (restatement). *Given $U \in U(d^n)$, let*

$$1 - \epsilon = \max\{|\langle U, V_1 \otimes \cdots \otimes V_n\rangle|^2 : V_1, \dots, V_n \in U(d)\}.$$

Then, if $\epsilon = 0$, $P_{test}(U) = 1$. If $\epsilon \lesssim 0.106$, then $P_{test}(U) \leq 1 - \frac{1}{4}\epsilon + \frac{1}{16}\epsilon^2 + \frac{1}{8}\epsilon^{3/2}$. If $0.106 \lesssim \epsilon \leq 1$, $P_{test}(U) \leq 501/512$. More concisely, $P_{test}(U) = 1 - \Theta(\epsilon)$.

Proof. By the Choi-Jamiołkowski isomorphism, there is a direct correspondence between operators $M \in M(d)$ with $|\langle M, M\rangle| = 1$ and normalised quantum states $|v(M)\rangle$. If we define

$$1 - \epsilon' := \max\{|\langle U, A_1 \otimes \cdots \otimes A_n\rangle|^2 : A_i \in M(d), \langle A_i, A_i\rangle = 1, 1 \leq i \leq n\},$$

then by Theorem 1, if $\epsilon' \lesssim 0.0265$, $P_{\text{test}}(U) \leq 1 - \epsilon' + \epsilon'^2 + \epsilon'^{3/2}$, and if $\epsilon' \gtrsim 0.0265$, $P_{\text{test}}(U) \leq 501/512$. If $\epsilon' \geq 1/2$, then the result follows immediately. On the other hand, by Lemma 24, if $\epsilon' \leq 1/2$, there exist $V_1, \dots, V_n \in U(d)$ such that $|\langle U, V_1 \otimes \cdots \otimes V_n\rangle|^2 \geq (1 - 2\epsilon')^2 \geq 1 - 4\epsilon'$. Thus we have $\frac{1}{4}\epsilon \leq \epsilon' \leq \epsilon$. The theorem follows by combining the bound on ϵ and the bound on $P_{\text{test}}(U)$. □

G Interpretation of the product test as an average over product states

We have seen (via Lemma 2) that the probability of the product test passing when applied to some state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ is equal to the average purity, across all choices of subsystem $S \subseteq [n]$, of $\operatorname{tr} |\psi\rangle\langle\psi|_S$. One interpretation of the proof of correctness of the product test is therefore that, if the average entanglement of $|\psi\rangle$ across all bipartite partitions of $[n]$ is low, as measured by the purity, then $|\psi\rangle$ must in fact be close to a product state across all subsystems.

In this appendix, we discuss a similar interpretation of our results in terms of an average over product states, via the following proposition.

Proposition 25. *Given $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$,*

$$P_{test}(|\psi\rangle\langle\psi|) = \left(\frac{d(d+1)}{2}\right)^n \mathbb{E}_{|\phi_1\rangle,\dots,|\phi_n\rangle}\left[|\langle\psi|\phi_1 \dots \phi_n\rangle|^4\right].$$

Proof. Similarly to before, let the input to the product test be two copies ψ_A, ψ_B of a state $\psi := |\psi\rangle\langle\psi|$, and let $\mathcal{F}$ denote the swap operator that exchanges systems A and B. Then

$$\begin{aligned}
&\mathbb{E}_{|\phi_1\rangle,\dots,|\phi_n\rangle}\left[|\langle\psi|\phi_1, \dots, \phi_n\rangle|^4\right] \\
&\quad= \mathbb{E}_{|\phi_1\rangle,\dots,|\phi_n\rangle}\left[\operatorname{tr}(\psi_A \otimes \psi_B)((\phi_1 \otimes \cdots \otimes \phi_n)_A \otimes (\phi_1 \otimes \cdots \otimes \phi_n)_B)\right] \\
&\quad= \operatorname{tr}(\psi_A \otimes \psi_B)\left(\mathbb{E}_{|\phi\rangle}\left[\phi_A \otimes \phi_B\right]\right)^{\otimes n} \\
&\quad= \operatorname{tr}(\psi_A \otimes \psi_B)\left(\frac{I + \mathcal{F}}{d(d+1)}\right)^{\otimes n} = \left(\frac{2}{d(d+1)}\right)^n P_{\text{test}}(|\psi\rangle\langle\psi|).
\end{aligned}$$

□

We note that, in this interpretation, our main result is reminiscent of the so-called inverse theorem for the second Gowers uniformity norm [33, 34], which we briefly outline. Let $f : \{0,1\}^n \to \mathbb{R}$ be some function such that $\frac{1}{2^n}\sum_x f(x)^2 = 1$, and let the p-norms of f on the Fourier side be defined as $\|\hat{f}\|_p = \left(\sum_{x\in\{0,1\}^n} \left|\frac{1}{2^n}\sum_{y\in\{0,1\}^n}(-1)^{x\cdot y}f(y)\right|^p\right)^{1/p}$. Then it is straightforward to show that $\|\hat{f}\|_\infty^4 \le \|\hat{f}\|_4^4 \le \|\hat{f}\|_\infty^2$, where the quantity in the middle is known as the (fourth power of) the second Gowers uniformity norm of f. That is, $\|\hat{f}\|_\infty^2$ (representing the largest overlap of f with a parity function) is well approximated by $\|\hat{f}\|_4^4$ (the *average* of the squared overlaps with parity functions). This simple approximation has proven useful in arithmetic combinatorics [33].

Via the correspondence of Proposition 25, Theorem 1 shows that a similar result holds if we replace the cube $\{0,1\}^n$ with the space $(\mathbb{C}^d)^{\otimes n}$: the largest overlap with a product state can be well approximated by the average squared overlap with product states. Note that if one attempts to use the classical proof technique for the Gowers uniformity norm to prove this result, one does not obtain Theorem 1, but a considerably weaker result containing a term exponentially large in n. Intuitively, this is because the set of overlaps with parity functions for some function $f : \{0,1\}^n \to \mathbb{R}$ is essentially arbitrary, whereas the set of overlaps of some state $|\psi\rangle$ with product states is highly constrained.

Acknowledgements

We did most of this research while working at the University of Bristol. AM was supported by the EC-FP6-STREP network QICS and an EPSRC Postdoctoral Research Fellowship. AWH was supported by the EPSRC grant "QIP-IRC", by NSF grants 0916400, 0829937, 0803478, DARPA QuEST contract FA9550-09-1-0044 and the IARPA MUSIQC and QCS contracts. We would like to thank many people for inspiring discussions, including Boaz Barak, Salman Beigi, Fernando Brandão, Toby Cubitt, Sevag Gharibian, Leonid Gurvits, Julia Kempe, Hirotada Kobayashi, Harumichi Nishimura, Thomas Vidick, Andreas Winter and Xiaodi Wu. We would also like to thank the FOCS and JACM referees for their helpful comments.

References

[1] S. Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A*, 463:2088, 2007. `quant-ph/0608142`.

[2] S. Aaronson, S. Beigi, A. Drucker, B. Fefferman, and P. Shor. The power of unentanglement. *Theory of Computing*, 5(1):1–42, 2009. `arXiv:0804.0802`.

[3] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani. The detectability lemma and quantum gap amplification. In *Proc. 41st Annual ACM Symp. Theory of Computing*, pages 417–426, New York, NY, USA, 2009. ACM. `arXiv:0811.3412`.

[4] S. Aja-Fernández, R. García, D. Tao, and X. Li. *Tensors in Image Processing and Computer Vision*. Advances in pattern recognition. Springer, 2009.

[5] G. G. Amosov, A. S. Holevo, and R. F. Werner. On some additivity problems in quantum information theory. *Problems Inform. Transmission*, 36(4):305–313, 2000. `math-ph/0003002`.

[6] A. Atici and R. A. Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6:323–348, 2007. `arXiv:0707.3479`.

[7] M. Aulbach. *Classification of Entanglement in Symmetric States*. PhD thesis, University of Leeds, 2011. `arXiv:1110.5200`.

[8] B. Barak, F. G. S. L. Brandão, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proc. 44th Annual ACM Symp. Theory of Computing*, pages 307–326, 2012. `arXiv:1205.4484`.

[9] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello. Stabilisation of quantum computations by symmetrisation. *SIAM J. Comput.*, 26(5):1541–1557, 1997. `quant-ph/9604028`.

[10] S. Beigi. NP vs QMA_log(2). *Quantum Inf. Comput.*, 10(1&2):0141–0151, 2010. `arXiv:0810.5109`.

[11] S. Beigi and P. Shor. On the complexity of computing zero-error and Holevo capacity of quantum channels, 2007. `arXiv:0709.2090`.

[12] H. Blier and A. Tapp. All languages in NP have very short quantum proofs. In *First International Conference on Quantum, Nano, and Micro Technologies*, pages 34–37, Los Alamitos, CA, USA, 2009. IEEE Computer Society. `arXiv:0709.0738`.

[13] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993.

[14] F. G. S. L. Brandão. *Entanglement Theory and the Quantum Simulation of Many-Body Physics*. PhD thesis, Imperial College, London, 2008. `arXiv:0810.0026`.

[15] F. G. S. L. Brandão, M. Christandl, and J. Yard. Faithful squashed entanglement. *Comm. Math. Phys.*, 306(3):805–830, 2011. `arXiv:1010.1750`.

[16] F. G. S. L. Brandão, M. Christandl, and J. Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proc. 43rd Annual ACM Symp. Theory of Computing*, pages 343–351, 2011. `arXiv:1011.2751`.

[17] S. Brubaker and S. Vempala. Random tensors and planted cliques. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 5687 of *Lecture Notes in Computer Science*, pages 406–419. Springer-Verlag, Berlin, Heidelberg, 2009. `arXiv:0905.2381`.

[18] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87(16):167902, 2001. `quant-ph/0102001`.

[19] H. Buhrman, L. Fortnow, I. Newman, and H. Röhrig. Quantum property testing. *SIAM J. Comput.*, 37(5):1387–1400, 2008. `quant-ph/0201117`.

[20] A. Chailloux and O. Sattath. The complexity of the separable Hamiltonian problem, 2011. `arXiv:1111.5247`.

[21] J. Chen and A. Drucker. Short multi-prover quantum proofs for SAT without entangled measurements, 2010. `arXiv:1011.0716`.

[22] A. Chiesa and M. Forbes. Improved soundness for QMA with multiple provers, 2011. `arXiv:1108.2098`.

[23] F. Cobos, T. Kühn, and J. Peetre. Remarks on symmetries of trilinear forms. *Rev. R. Acad. Cienc. Exact. Fis.Nat. (Esp)*, 94(4):441–449, 2000.

[24] T. Cubitt, A. W. Harrow, D. Leung, A. Montanaro, and A. Winter. Counterexamples to additivity of minimum output p-Renyi entropy for p close to 0. *Comm. Math. Phys.*, 284:281–290, 2008. `arXiv:0712.3628`.

[25] W. F. de la Vega, M. Karpinski, R. Kannan, and S. Vempala. Tensor decomposition and approximation schemes for constraint satisfaction problems. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, STOC '05, pages 747–754, 2005.

[26] I. Devetak, M. Junge, C. King, and M. B. Ruskai. Multiplicativity of completely bounded p-norms implies a new additivity result. *Comm. Math. Phys.*, 266:37–63, 2006. `quant-ph/0506196`.

[27] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum channel capacity of very noisy channels. *Phys. Rev. A.*, 57:830, 1998. `quant-ph/9706061`.

[28] J. Eisert, P. Hyllus, O. Gühne, and M. Curty. Complete hierarchies of efficient approximations to problems in entanglement theory. *Phys. Rev. A*, 70:062317, Dec 2004. `quant-ph/0407135`.

[29] M. Fannes and C. Vandenplas. Finite size mean-field models. *J. Phys. A: Math. Gen.*, 39(45):13843, 2006. `quant-ph/0605216`.

[30] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the European Association for Theoretical Computer Science*, 75:97–126, 2001.

[31] S. Gharibian. Strong NP-hardness of the quantum separability problem. *Quantum Inf. Comput.*, 10(3&4):343–360, 2010. `arXiv:0810.4507`.

[32] S. Gharibian, J. Sikora, and S. Upadhyay. QMA variants with polynomially many provers, 2011. `arXiv:1108.0617`.

[33] W. T. Gowers. A new proof of Szeméredi's theorem for progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998.

[34] W. T. Gowers. A new proof of Szeméredi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001.

[35] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*. Springer-Verlag, 1993.

[36] A. Grudka, M. Horodecki, and L. Pankowski. Constructive counterexamples to additivity of minimum output Rényi entropy of quantum channels for all $p > 2$, 2009. `arXiv:0911.2515`.

[37] O. Gühne and G. Toth. Entanglement detection. *Physics Reports*, 471(1), 2009. `arXiv:0811.2803`.

[38] L. Gurvits. Classical deterministic complexity of Edmonds' problem and quantum entanglement. In *Proc. 35th Annual ACM Symp. Theory of Computing*, pages 10–19, 2003. `quant-ph/0303055`.

[39] A. W. Harrow. Permutations are nearly orthogonal, 2012. In preparation.

[40] M. B. Hastings. A counterexample to additivity of minimum output entropy. *Nature Physics*, 5, 2009. `arXiv:0809.3972`.

[41] P. Hayden and A. Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all p > 1. *Comm. Math. Phys.*, 284(1):263–280, 2008.

[42] A. S. Holevo and R. F. Werner. Counterexample to an additivity conjecture for output purity of quantum channels. *J. Math. Phys.*, 43:4353–4357, 2002. quant-ph/0203003.

[43] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1–2):1–8, 1996. quant-ph/9605038.

[44] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009. quant-ph/0702225.

[45] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert. "Event-ready-detectors" Bell experiment via entanglement swapping. *Phys. Rev. Lett.*, 71:4287–4290, 1993.

[46] R. Impagliazzo and R. Paturi. On the complexity of k-SAT. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001.

[47] L. M. Ioannou. Computational complexity of the quantum separability problem. *Quantum information and computation*, 7:335, 2007. quant-ph/0603199.

[48] T. Ito, H. Kobayashi, and K. Matsumoto. Oracularization and two-prover one-round interactive proofs against nonlocal strategies, 2008. arXiv:0810.0693.

[49] J. Kempe and O. Regev. No strong parallel repetition with entangled and non-signaling provers, 2009. arXiv:0911.0201.

[50] H. Kobayashi, K. Matsumoto, and T. Yamakami. Quantum Merlin-Arthur proof systems: are multiple Merlins more helpful to Arthur? In *Proc. ISAAC '03*, pages 189–198, 2003. quant-ph/0306051.

[51] F. Le Gall, S. Nakagawa, and H. Nishimura. On QMA protocols with two short quantum proofs, 2011. arXiv:1108.4306.

[52] Y.-K. Liu. *The Complexity of the Consistency and N-representability Problems for Quantum States.* PhD thesis, Univ. of California, San Diego, 2007. arXiv:0712.3041.

[53] Y.-K. Liu, M. Christandl, and F. Verstraete. N-representability is QMA-complete. *Phys. Rev. Lett.*, 98:110503, 2007. quant-ph/0609125.

[54] R. A. Low. Learning and testing algorithms for the Clifford group. *Phys. Rev. A*, 80:052314, 2009. arXiv:0907.2833.

[55] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005. cs/0506068.

[56] K. Matsumoto. Some new results and applications of additivity problem of quantum channel. Poster at QIP'05 conference, 2005.

[57] M. McKague. On the power of quantum computation over real Hilbert spaces, 2011. arXiv:1109.0795.

[58] F. Mintert, M. Kuś, and A. Buchleitner. Concurrence of mixed multipartite quantum states. *Phys. Rev. Lett.*, 95(26):260502, 2005. quant-ph/0411127.

[59] A. Montanaro and T. Osborne. Quantum boolean functions. *Chicago Journal of Theoretical Computer Science*, 2010. arXiv:0810.2435.

[60] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

[61] T. Ogawa and H. Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Trans. Inform. Theory*, 45(7):2486–2489, 1999. quant-ph/9808063.

[62] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77(8):1413–1415, 1996.

[63] R. Renner. *Security of quantum key distribution*. PhD thesis, ETH Zurich, 2005. quant-ph/0512258.

[64] Y. Shi and X. Wu. Epsilon-net method for optimizations over separable states. In *Proc. 39th International Conference on Automata, Languages and Programming (ICALP'12)*, pages 798–809, 2012. arXiv:1112.0808.

[65] B. M. Terhal. Bell inequalities and the separability criterion. *Phys. Lett. A*, 271:319, 2000. quant-ph/9911057.

[66] S. J. van Enk, N. Lütkenhaus, and H. J. Kimble. Experimental procedures for entanglement verification. *Phys. Rev. A*, 75:052318, May 2007. arXiv:quant-ph/0611219.

[67] C. van Loan. Future directions in tensor-based computation and modeling, 2009. Unpublished NSF Workshop Report.

[68] G. Vidal and J. I. Cirac. Irreversibility in asymptotic manipulations of entanglement. *Phys. Rev. Lett.*, 86:022308, 2001. quant-ph/0102036.

[69] S. Walborn, P. Ribeiro, L. Davidovich, F. Mintert, and A. Buchleitner. Experimental determination of entanglement with a single measurement. *Nature*, 440(7087):1022–1024, 2006.

[70] G. Wang. Property testing of unitary operators. *Phys. Rev. A*, 84:052328, Nov 2011. arXiv:1110.1133.

[71] T. Wei and P. Goldbart. Geometric measure of entanglement and applications to bipartite and multipartite quantum states. *Phys. Rev. A.*, 68(4):42307, 2003. quant-ph/0307219.

[72] R. F. Werner and A. S. Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43(9):4353–4357, 2002. quant-ph/0203003.

[73] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inform. Theory*, 45(7):2481–2485, 1999.

[74] D. Yang, M. Horodecki, R. Horodecki, and B. Synak-Radtke. Irreversibility for all bound entangled states. *Phys. Rev. Lett.*, 95:190501, 2005. quant-ph/0506138.

[75] D. Yang, M. Horodecki, and Z. D. Wang. An additive and operational entanglement measure: conditional entanglement of mutual information. *Phys. Rev. Lett.*, 101:140501, 2008. arXiv:0804.3683.