An efficient test for product states, with applications to quantum Merlin-Arthur games

Aram W. Harrow* and Ashley Montanaro†

May 4, 2010

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, where we show that a witness from two unentangled provers can simulate a witness from arbitrarily many unentangled provers, up to a constant loss of soundness. 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. This result implies complexity-theoretic obstructions to finding a polynomial-time algorithm to determine separability of mixed quantum states, even up to constant error, and also to proving "weak" variants of the additivity conjecture for quantum channels.

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 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 [16, 26] reflect how entangled inputs can sometimes have advantages against even uncorrelated noise. For

*Department of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK; a.harrow@bris.ac.uk.

†Department of Computer Science, University of Bristol, Woodland Road, Bristol, BS8 1UB, UK; montanar@cs.bris.ac.uk.

quantum interactive proofs, the primary difficulty is in bounding the ability of provers to cheat using entangled strategies [30]. Even for $\mathsf{QMA}(k)$ (the variant of QMA with k unentangled Merlins [32, 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 [31] 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 [40, 47]. 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 Appendix C) is that it requires two copies of $|\psi\rangle$.

The parameters of our test resemble classical property-testing algorithms [18]. 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. [13, 6]) and quantum [38] oracles (a.k.a. unitary operators, although see Section 5 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 [24] 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 [12]. This test, which can be implemented efficiently, takes two (possibly mixed) states ρ, σ of equal dimension as input, and returns "same" with probability $\frac{1}{2} + \frac{1}{2}\operatorname{tr}\rho\,\sigma$, otherwise returning "different".

The product test has appeared before in the literature. It was originally introduced in [37] 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 [45]. Further, the test was proposed in [38] 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

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.*

theorem.

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

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

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

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

Furthermore, if $\epsilon \geq 11/32 > 0.343$*,* $P_{test}(|\psi\rangle\langle\psi|) \leq 501/512 < 0.979$*.*

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 Appendix C 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 [25, 19] (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 channel. It is known that the maximum output purity of this channel is achieved for product state inputs [4]; 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.

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

for ρ a arbitrary mixed state of one d-dimensional system, and define the product state output purity to be $P_{\text{prod}}(\delta) = \operatorname{tr}(\mathcal{D}_\delta^{\otimes n} |\phi\rangle\langle\phi|)^2$, where $|\phi\rangle$ is an arbitrary product state. Then

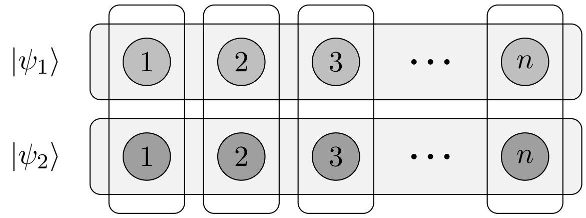

Figure 1: Schematic of the product test applied to an n-partite state $|\psi\rangle$. The swap test (vertical boxes) is applied to the n pairs of corresponding subsystems of two copies of $|\psi\rangle$ (horizontal boxes).

our main result, stated more precisely as Theorem 10 in Appendix A, is that for small enough $\delta > 0$, if $\mathrm{tr}(\mathcal{D}_\delta^{\otimes n} |\psi\rangle\langle\psi|)^2 \geq (1-\epsilon)P_{\mathrm{prod}}(\delta)$, then there is a product state $|\phi_1, \ldots, \phi_n\rangle$ such that $|\langle\psi|\phi_1, \ldots, \phi_n\rangle|^2 \geq 1 - O(\epsilon)$.

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 [32, 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. Our result also allows us to prove a weak form of amplification for $\mathsf{QMA}(2)$ and $\mathsf{QMA}(k)$ in which $1 - 1/\mathrm{poly}(n)$ soundness can be improved to $1 - \Omega(1)$ soundness.

As a further corollary, we can improve upon the results of [1, 10] to obtain a protocol in $\mathsf{QMA}(2)$ that verifies 3-SAT with constant soundness gap and $O(\sqrt{n}\,\mathrm{poly}\log(n))$ qubits (where n is the number of clauses). This in turn has consequences for the difficulty of approximating $\mathrm{SEP}(d, d)$, the set of separable quantum states on $d \times d$ dimensions. It was shown in Ref. [25] that SEP cannot be approximated to precision $\exp(-d)$ in time $\mathrm{poly}(d)$ unless $\mathsf{P} = \mathsf{NP}$. In Refs. [35, 19], 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 $\in \mathsf{DTIME}(\exp(\sqrt{n}\log^{O(1)}(n)))$.

Our result has two further corollaries, under the same assumption on the complexity of 3-SAT. First, we show that the minimum output entropy of a quantum channel cannot be estimated up to a constant in polynomial time. Second, we show a hardness result for approximating the ground-state energy of quantum systems under a mean-field approximation.

Our final application is that the product test can be used to determine whether a unitary operator is a tensor product. This can be seen [38] as one possible generalisation of the well-studied problem of testing whether a boolean function $\{0,1\}^n \to \{0,1\}$ is linear [11]. This application is described in Section 5.

These different applications of the product test reflect the many different interpretations of $P_{\mathrm{test}}(|\psi\rangle\langle\psi|)$. 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 [46] 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 Appendix C).
- The average overlap of $|\psi\rangle$ with a *random* product state, and a quantum variant of the Gowers uniformity norm [20] (discussed in Appendix G).
- The average purity of $|\psi\rangle$ across a random partition of $[n]$ into two subsets (also discussed in Appendix G).

2 Overview of the proof of correctness

A full proof of correctness of the product test (Theorem 1) is given in Appendices A and B. In this section, we sketch the proof strategy.

Our results are based on the following expression for the probability that an arbitrary state $|\psi\rangle$ passes the product test: $P_{\text{test}}(|\psi\rangle\langle\psi|) = \frac{1}{2^n}\sum_{S\subseteq[n]} \operatorname{tr}\psi_S^2$. The connection with the depolarising channel follows from observing that the output purity of a state to which n copies of the depolarising channel with arbitrary noise rate have been applied is given by a similar but more complicated sum, with weights on each term.

The proof itself is split into two parts, beginning with the case where ϵ is low. The difficult part is the upper bound on $P_{\text{test}}(|\psi\rangle\langle\psi|)$. 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. Each term of this expression is then carefully upper bounded to give an upper bound in terms of a sum of the amplitudes of $|\phi\rangle$, with weights that decrease exponentially with the Hamming weight of basis states. 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 1.

In the case where ϵ is high (a regime essential to the application to $\mathsf{QMA}(k)$), 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.

This completes the overview of the proof; we now turn to 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 better than 2 unentangled provers. This question can be formalised as whether the complexity classes $\mathsf{QMA}(k)$ and $\mathsf{QMA}(2)$ are equal [32, 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 $|\psi_1\rangle, \ldots, |\psi_k\rangle$, each a state of $\text{poly}(n)$ qubits, such that $\mathcal{A}$ outputs "accept" with probability at least c on input $|x\rangle|\psi_1\rangle \ldots |\psi_k\rangle$.*
2. **Soundness:** *If $x \notin L$, then $\mathcal{A}$ outputs "accept" with probability at most s on input $|x\rangle|\psi_1\rangle \ldots |\psi_k\rangle$, for all states $|\psi_1\rangle, \ldots, |\psi_k\rangle$.*

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 $1 \le k \le \text{poly}(n)$.

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

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 [32, 2] that in fact $\mathsf{QMA}(k) = \mathsf{QMA}(2)$, 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. We do not fully resolve these conjectures, but instead prove them only for s larger than a constant. The idea is that using the product test, any $\mathsf{QMA}(k)$ protocol can be simulated by a $\mathsf{QMA}(2)$ protocol, as long as one is willing to accept some loss of soundness. Access to $k/(c-s)^2$ unentangled provers can be used in turn to amplify $\mathsf{QMA}(k)$ protocols to a constant soundness/completeness gap[1]. These results are expressed in the following theorem:

Theorem 2. *For any polynomials $p(n), q(n)$ that are ≥ 2 for all n,*

1. $\mathsf{QMA}(2)_{0.99,1} = \mathsf{QMA}(p(n))_{1-1/q(n),1}$
2. *If $c - s \ge 1/\text{poly}(n)$, then $\mathsf{QMA}(2)_{0.99,1-\exp(-p(n))} = \mathsf{QMA}(p(n))_{s,c}$.*

We suspect that a tighter analysis of the product test could be used to obtain amplification even up to exponentially small soundness. However, for many applications the important amplification step is from inverse-polynomial soundness gap to constant soundness gap.

To prove Theorem 2, we will first show that a $\mathsf{QMA}(k)$ protocol with soundness s and completeness c can be simulated using two unentangled proofs with soundness $1 - \Theta((1-s)^2)$ and completeness c. Suppose the proofs in the original protocol are $|\psi_1\rangle, \ldots, |\psi_k\rangle$ and Arthur's original verification algorithm is $\mathcal{A}$. Then the $\mathsf{QMA}(2)$ protocol is as specified in Protocol 2 below.

We prove correctness of Protocol 2 and the rest of Theorem 2 in Appendix D.

4 Complexity-theoretic implications

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

[1] We are grateful to Salman Beigi for pointing this out to us.

Protocol 2 (QMA(k) **to** QMA(2)**).**

The 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 runs the product test with the two states as input.*
3. *If the test fails, Arthur rejects. Otherwise, Arthur runs the algorithm* $\mathcal{A}$ *on one of the two states, picked uniformly at random, and outputs the result.*

Corollary 3.

$$\text{3-SAT} \in \mathsf{QMA}_{\sqrt{n}\,\mathrm{poly}\log(n)}(2)_{1-\Omega(1),1}.$$

In other words, there is a 3-SAT protocol with two provers, $\sqrt{n}\,\mathrm{poly}\log(n)$*-qubit proofs, perfect completeness and constant soundness.*

Corollary 3 has several easy, but useful, consequences. First, we observe that we can scale down by an exponential to obtain

Corollary 4. *Define 3-SAT*$(\log^2)$ *to be the language of satisfiable 3-CNF formulas with length* n *and* $\leq \log^2(n)$ *variables. Then 3-SAT*$(\log^2) \in \mathsf{QMA}_{\log}(2)_{1-\Omega(1),1}$.

If we assume that there are no poly-time deterministic (or randomised or quantum) algorithms for 3-SAT$(\log^2)$, then this implies that $\mathsf{QMA}_{\log}(2)_{1-\Omega(1),1}$ is not contained in P (or BPP or BQP, respectively). Equivalently, we can assume that 3-SAT $\notin \mathsf{DTIME}(\exp(\sqrt{n}\log^{O(1)}(n)))$. This assumption is implied by the plausible *Exponential Time Hypothesis* (**ETH**) of Impagliazzo and Paturi [29], which states that 3-SAT $\notin \mathsf{DTIME}(\exp(o(n)))$. We conclude this section by discussing three $\mathsf{QMA}_{\log}(2)$-complete problems, whose hardness is implied by **ETH**.

First observe that the acceptance probability in a $\mathsf{QMA}_m(2)$ protocol can be expressed as $\max_{\rho\in\mathrm{SEP}(2^m,2^m)} \mathrm{tr}\, M\rho$, where $0 \leq M \leq I$ is the measurement resulting from the verifier's quantum circuit and $\mathrm{SEP}(d_A, d_B)$ denotes the set of separable density matrices on $d_A \times d_B$ dimensions. In other words, $\mathsf{QMA}_m(2)$ is equivalent to optimising a linear objective function over the convex set $\mathrm{SEP}(2^m, 2^m)$.

In some cases, it may be useful to obtain an explicit description of M. This can be achieved up to error ϵ by running the verifier's circuit $\mathrm{poly}(2^m, 1/\epsilon)$ times and performing tomography. As a result, we trivially obtain that $\mathsf{QMA}_m(2)_{s,c} \subset \mathsf{NTIME}(\mathrm{poly}(2^m, n, 1/(c-s)))$. In particular, $\mathsf{QMA}(2) \subset \mathsf{NEXP}$. 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 $\mathsf{QMA}_{\log}(2) \subset \mathsf{NP}^{\mathsf{BQP}}$.

Application 1: Separability-testing. A folk theorem of convex optimisation [22] states that the problem of optimising a linear function over a convex set, such as SEP, is equivalent to determining membership in that set. Thus, we should be able to relate QMA(2) to the problem of determining membership in SEP. To make this precise, for any convex $K \subset \mathbb{R}^d$ we define $B(K, \epsilon)$ to be $\{x : \exists y \in K, \|x - y\| \leq \epsilon\}$ when $\epsilon > 0$ and $\{x : \nexists y \notin K, \|x - y\| \leq -\epsilon\}$ when $\epsilon < 0$. The weak membership problem for K, $\mathrm{WMEM}_\epsilon(K)$, is to determine whether a point x belongs to $B(K, \epsilon)$ or $B(K, -\epsilon)$ given the promise that one of these is the case. The weak optimisation problem for K, $\mathrm{WOPT}_\epsilon(K)$ is to maximize a linear objective function over any set L satisfying $B(K,-\epsilon) \subset L \subset B(K,\epsilon)$. Given some mild conditions on K, we can reduce $\mathrm{WOPT}_\epsilon(K)$ to $\mathrm{WMEM}_{\epsilon/\mathrm{poly}(d)}(K)$ in polynomial time [22]. This fact has been used to show the NP-hardness of $\mathrm{WMEM}_{1/\mathrm{poly}}(\mathrm{SEP})$ in Refs. [35, 19, 8] and, previously, of $\mathrm{WMEM}_{1/\exp}(\mathrm{SEP})$ by Gurvits [25] (although the connection to $\mathsf{QMA}_{\log}(2)$ was only observed by [8]).

Unfortunately, many of these techniques break down in the setting of constant error. We believe that it should not be possible to approximate $\mathrm{SEP}(d,d)$ to within a (sufficiently small) constant accuracy in time $\mathrm{poly}(d)$. 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.

Corollary 5. *Let K be a convex subset of the space of $d^2 \times d^2$ Hermitian matrices such that $K \subset B(\mathrm{SEP}(d,d),\delta)$ and $\mathrm{SEP}(d,d) \subset B(K,\delta)$, where $B(\cdot,\cdot)$ is defined relative to the trace norm and $\delta > 0$ is a universal constant. Then, assuming the Exponential Time Hypothesis,* $\mathrm{WMEM}_{1/\mathrm{poly}}(K)$ *cannot be decided in time* $\mathrm{poly}(d)$.

(As with the other hardness results in this section, the precise value of δ is determined by the protocol in Ref. [2].)

Proof. Solving $\mathrm{WMEM}_{1/\mathrm{poly}(d)}(K)$ in polynomial time would allow us to solve $\mathrm{WOPT}_{1/\mathrm{poly}(d)}(K)$ in poly time, which in turn would give a poly-time algorithm for $\mathrm{WOPT}_{0.01+1/\mathrm{poly}(d)}(\mathrm{SEP})$. This last claim, together with Corollary 4, would contradict the Exponential Time Hypothesis. □

Application 2: Minimum output entropy of quantum channels. Our results also have implications for quantum information theory. Let $\mathcal{N}$ denote a quantum channel with d-dimensional input and output. Define the minimum output Rényi α-entropy of $\mathcal{N}$ to be $S_\alpha^{\min}(\mathcal{N}) = \min_\rho S_\alpha(\mathcal{N}(\rho))$, where $S_\alpha(\sigma) = \frac{1}{1-\alpha}\log \operatorname{tr}\sigma^\alpha$ and the minimum is taken over all quantum states ρ. Note that $S_\alpha^{\min}(\mathcal{N})$ is also equivalent to $\frac{\alpha}{1-\alpha}\log\|\mathcal{N}\|_{1\to\alpha}$, where $\|\cdot\|_{1\to\alpha}$ (also called ν_α in e.g. [5]) is the $\ell_1 \to \ell_p$ norm. When $\alpha = 0, 1, \infty$, we define $S_\alpha(\sigma)$ by letting α approach these values, obtaining $S_0(\sigma) = \log\operatorname{rank}\sigma$, $S_1(\sigma) = -\operatorname{tr}\sigma\log\sigma$ and $S_\infty(\sigma) = -\log\|\sigma\|_\infty$. Additivity of $S_1^{\min}(\mathcal{N})$, the minimum output entropy, is intimately connected to additivity of the Holevo capacity [42, 9].

It was observed by Matsumoto [36] (citing a personal communication from Watrous) that the maximum acceptance probability of a $\mathsf{QMA}_m(2)$ protocol is precisely $\|\mathcal{N}\|_{1\to\infty}$ for some quantum channel $\mathcal{N}$ acting on $d = 2^m$ dimensions. For completeness, we give a proof of this in Appendix E. This implies that determining whether $S_\infty^{\min}(\mathcal{N})$ is $\geq \log(1/s)$ or $\leq \log(1/c)$ is a complete problem for $\mathsf{QMA}_{\log(d)}(2)_{s,c}$ under BQP reductions.

The $\mathsf{QMA}(2)$-completeness of estimating $S_\infty^{\min}$ implies that other information-theoretic quantities that are close to $S_\infty^{\min}$ are also hard to approximate. For example, 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, our hardness result for approximating $S_\infty^{\min}$ immediately translates to a hardness result for approximating $S_\alpha^{\min}$.

Corollary 6. *There exists a universal constant $\delta > 0$ such that for any $\alpha \geq 0$, if the Exponential Time Hypothesis holds then it is impossible to determine whether $S_\alpha^{\min}(\mathcal{N}) = 0$ or $S_\alpha^{\min}(\mathcal{N}) \geq \delta$ in time* $\mathrm{poly}(d)$.

Beigi and Shor previously showed that it is NP-hard to compute the minimum output entropy up to $1/\mathrm{poly}(d)$ accuracy [9]. Our result improves theirs, but under a more restrictive complexity assumption. Another major goal in information theory is to estimate the regularised minimum output entropies of quantum channels, which are defined to be

$$S_\alpha^{R,\min}(\mathcal{N}) := \liminf_{n\to\infty} \frac{1}{n} S_\alpha^{\min}(\mathcal{N}^{\otimes n}).$$

The $S_\alpha^{R,\min}(\mathcal{N})$ are relevant to determining the ultimate channel capacity, to proving strong converse theorems [33] and to cryptographic protocols [34].

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 $\alpha \geq 1$ [26, 27] and for α near 0 [14]. 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 [44, 48, 15, 49], 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.

Application 3: mean-field approximation. Finally, we discuss an application from condensed-matter physics. Consider a system of n d-dimensional quantum systems arranged in a lattice with identical nearest-neighbour pairwise interactions. The mean-field approximation replaces the true nearest-neighbour graph with the complete graph. When the number of spatial dimensions is 3 (or more), this is often a reasonable approximation. If K is a fixed two-qudit Hamiltonian and $K_{i,j}$ denotes the action of K on systems i, j and the identity on the other systems, then the total Hamiltonian is $H = \sum_{i \neq j} K_{i,j}$. To set the overall scale of the problem, assume that $0 \leq K \leq I$. One of the more important physical questions about H is to determine its ground-state energy; that is, its smallest eigenvalue.

In Ref. [17], 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 this case, finding the ground-state energy of H is equivalent to minimising $\operatorname{tr} \rho K$ over all $\rho \in \mathrm{SEP}(d, d)$. Again applying Corollary 3, we obtain:

Corollary 7. *Assuming the Exponential Time Hypothesis, and with H defined as above, it is impossible to estimate* $\min\{\operatorname{tr} \rho K : \rho \in \mathrm{SEP}(d, d)\}$ *in time* $\operatorname{poly}(d)$ *to within* $o(1)$ *error. Equivalently, it is impossible to estimate the ground-state energy of H to within additive error* $o(n^2)$.

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. In terms of approximation errors, 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 $\Omega(\exp(\log^2(d)))$ 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 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, $|\langle U, V \rangle| \leq 1$ for unitary operators U, V. 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 $\mathcal{B} = (|1\rangle, \ldots, |d\rangle)$. 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 $M, N \in M(d^n)$, $\langle M, N\rangle = \langle v(M)|v(N)\rangle = \operatorname{tr} M^\dagger N/d^n$.

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 $f : \{0,1\}^n \to \{0,1\}$ [11]; the classical special case is obtained by restricting U to be diagonal in the computational basis and to have diagonal entries all equal to ± 1.

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.

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 [38].

Theorem 8. *Given $U \in U(d^n)$, let*

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

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.

6 Conclusion

Our main result can be seen as a "stability" theorem for the output purity of the depolarising channel. 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. As a more modest open question, 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. Finally, we hope that a suitably strengthened version of our result can be used to improve the $\mathsf{QMA}(k)$ amplification results in Theorem 2 so that soundness can be made exponentially small.

Acknowledgements

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". We would like to thank Salman Beigi, Toby Cubitt, Julia Kempe, Thomas Vidick and Andreas Winter for inspiring discussions.

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 *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 417–426, New York, NY, USA, 2009. ACM. arXiv:0811.3412.

[4] G. Amosov, A. Holevo, and R. Werner. On some additivity problems in quantum information theory, 2000. math-ph/0003002.

[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. arXiv: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] 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.

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

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

[10] 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.

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

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

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

[14] 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`.

[15] 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`.

[16] 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`.

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

[18] 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.

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

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

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

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

[23] 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`.

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

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

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

[27] 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.

[28] 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. arXiv:quant-ph/0203003.

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

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

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

[32] 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`.

[33] R. König and S. Wehner. A strong converse for classical channel coding using entangled inputs. *Phys. Rev. Lett.*, 103:070504, 2009. `arXiv:0903.2838`.

[34] R. König, S. Wehner, and J. Wullschleger. Unconditional security from noisy quantum storage, 2009. `arXiv:0906.1030`.

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

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

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

[38] A. Montanaro and T. Osborne. Quantum boolean functions, 2008. `arXiv:0810.2435`.

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

[40] 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`.

[41] T. Ogawa and H. Nagaoka. A new proof of the channel coding theorem via hypothesis testing in quantum information theory. In *Proc. 2002 IEEE International Symposium on Information Theory*, page 73, 2002. `quant-ph/0208139`.

[42] P. W. Shor. Equivalence of additivity questions in quantum information theory. *Comm. Math. Phys.*, 246(3):453–472, 2004. `quant-ph/0305035`.

[43] W. F. Stinespring. Positive functions on c*-algebras. *Proc. Amer. Math. Soc.*, 6:211–216, 1955.

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

[45] 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.

[46] 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.

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

[48] 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.

[49] 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.

A The depolarising channel

Let $\mathcal{D}_\delta$ be the qudit depolarising channel as defined in equation (1). 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 9. *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,

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

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,

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

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 $\mathrm{tr}\,\rho_S^2 \le 1$ 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 10. *Given* $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, *let*

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

Then

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

In particular,

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

Proof. Without loss of generality assume that one of the states achieving the maximum in Eq. (2) 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 9,

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

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),

$$\sum_{S \subseteq [n]} \gamma^{|S|} \operatorname{tr} |0\rangle\langle 0|_S |\phi\rangle\langle\phi|_S = \sum_{k=0}^{n-|x|} \gamma^k \binom{n-|x|}{k} = (1+\gamma)^n \sum_{x \neq 0} |\alpha_x|^2 \left(\frac{1}{1+\gamma}\right)^{|x|}.$$

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:

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

Combining these terms, we have

$$\sum_{S\subseteq[n]} \gamma^{|S|} \operatorname{tr}\psi_S^2 \leq (1+\gamma)^n((1-\epsilon)^2 + 2\epsilon(1-\epsilon)\sum_{x\neq 0}|\alpha_x|^2(1+\gamma)^{-|x|}(\gamma^{|x|}+1) + \epsilon^2 +$$

$$4\epsilon^{3/2}\sqrt{1-\epsilon}\left(\sum_x \left(\frac{1+\gamma^2}{(1+\gamma)^2}\right)^{|x|}|\alpha_x|^2\right)^{1/2}).$$

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. (3), this implies that only the $|x| \geq 4$ terms contribute (since $x = y \vee z$ and $y \wedge z = \emptyset$). Therefore, $|\phi\rangle$ having no weight 1 components would imply that

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

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

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

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 [38] expresses the probability of passing in terms of the partial traces of the input states; we include a proof for completeness.

Lemma 11. *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 9. □

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 [46], is essentially immediate from Lemma 11.

Lemma 12. *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 13. *Given $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, let*

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

Then

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

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 11 and Theorem 10. The statement of Theorem 10 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 13 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 14. *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, $P^P_{test}(|\psi\rangle\langle\psi|) \leq P_{test}(|\psi\rangle\langle\psi|)$.*

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

Lemma 15. *Let $|\psi\rangle$, $|\phi\rangle$ be pure states such that $|\langle\psi|\phi\rangle|^2 = 1 - \epsilon$, and let P be a projector. Then $|\langle\psi|P|\psi\rangle - \langle\phi|P|\phi\rangle| \leq \sqrt{\epsilon}$.*

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 [39, Chapter 9]). ☐

Theorem 16. *Given $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, let*

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

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

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: $P_{\text{test}}(|\psi\rangle\langle\psi|) \leq 1 - \frac{3}{4}\epsilon + 2\epsilon^2$. For a contradiction, assume that $P_{\text{test}}(|\psi\rangle\langle\psi|) > p := 501/512$, while $\epsilon \geq 11/32$.

For any partition P of $[n]$ into $1 \leq k \leq n$ 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

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

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 $P^P_{\text{test}}(|\psi\rangle\langle\psi|) \leq p$, implying by Lemma 14 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 the 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

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

and by the arguments at the end of Theorem 10, the i^{th} marginal of $|\xi\rangle$ has support orthogonal to $|\phi_i\rangle$. Assume without loss of generality that $|\phi_1\rangle$ is a state of two or more qudits. Now we know that

$$\max_{|\phi'_{1,1}\rangle, |\phi'_{1,2}\rangle} |\langle\phi_1|\phi'_{1,1}\rangle|\phi'_{1,2}\rangle|^2(1-\epsilon') < 1 - h, \tag{3}$$

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 $|\xi\rangle$ 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. (3) implies that

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

Using Lemma 12, we find that $P_{\text{test}}(|\psi_1\rangle\langle\psi_1|) \leq 1-\delta+\delta^2 < 751/961$. Next we use Lemma 15 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 12 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 13 and 16, we obtain Theorem 1 and thus have proven correctness of the product test. The constants in Theorem 16 have not been optimised as far as possible and could be improved somewhat.

C Optimality of the product test

Our test has perfect completeness in the sense that if $|\psi\rangle$ is exactly a product state then it will always pass the product test. It is hard to precisely define soundness, since no state is orthogonal to all product states: however, 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 [7]. 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 $\mathrm{Sym}^k\,\mathbb{C}^d$ is given by $\frac{1}{k!}\sum_{\pi\in\mathcal{S}_k} P(\pi)$, where

$$P(\pi) = \sum_{i_1,\dots,i_k\in[d]} |i_1,\dots,i_k\rangle\langle i_{\pi(1)},\dots,i_{\pi(k)}|.$$

For $k = 1$, $\mathrm{Sym}^1\,\mathbb{C}^d$ 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 $k = 2$, $\mathrm{Sym}^2\,\mathbb{C}^d$ 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 $\{|\psi\rangle^{\otimes k} : |\psi\rangle \in S\}$. However, we will not explore these possibilities further in this paper.

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

In this section, we prove correctness of Protocol 2. It is obvious that this protocol achieves completeness c: if the Merlins follow the protocol, the product test passes with certainty, and hence Arthur accepts with the same probability that $\mathcal{A}$ accepts, which is at least c. Showing soundness is somewhat more complicated.

We first show that we can assume that the two states Arthur receives are identical. Imagine that this does not hold, and Arthur receives different states $|\phi\rangle, |\varphi\rangle$. Then the probability that the product test accepts is

$$\begin{aligned}
\frac{1}{2^k}\sum_{S\subseteq[k]} \operatorname{tr}\phi_S\varphi_S &\le \frac{1}{2^k}\sum_{S\subseteq[k]} \sqrt{\operatorname{tr}\phi_S^2}\sqrt{\operatorname{tr}\varphi_S^2}\\
&\le \frac{1}{2^k}\sum_{S\subseteq[k]} \frac{\operatorname{tr}\phi_S^2 + \operatorname{tr}\varphi_S^2}{2}\\
&= \frac{1}{2}\left(P_{\text{test}}(\phi) + P_{\text{test}}(\varphi)\right),
\end{aligned}$$

where the first inequality is Cauchy-Schwarz and the second is the AM-GM inequality. As we run $\mathcal{A}$ on a random choice of the two states in the second stage, the probability that the whole algorithm accepts is also upper bounded by the average probability of it accepting when run on $|\phi\rangle$ and $|\varphi\rangle$. So, to achieve maximal probability of accepting, the two states might as well be identical.

To prove the remainder of the theorem, we will need the following "gentle measurement" lemma.

Lemma 17 (Gentle measurement lemma [47, 41])**.** *Let ρ be a density operator, and let $0 \le X \le I$ be a projector such that $\operatorname{tr}\rho X \ge 1-\delta$. Then $\frac{1}{2}\|\rho - X\rho X\|_1 \le \sqrt{\delta}$.*

Assume that the maximum overlap of $|\psi\rangle$ with a product state is $1-\epsilon$ and that the product test accepts $|\psi\rangle^{\otimes 2}$ with probability $1-\delta$. Recall from Theorem 1 that $\delta \geq \frac{11}{512}\epsilon$. Next, by Lemma 15, $\mathcal{A}$ accepts $|\psi\rangle$ with probability $\leq s+\sqrt{\epsilon}$. Thus, Lemma 17 implies that, conditioned on passing the product test, we are left with a state that has probability $\leq s+\sqrt{\epsilon}+\sqrt{\delta}$ of being accepted by $\mathcal{A}$. Combining the two tests, we find that the acceptance probability is

$$s' \leq \max_{\epsilon \leq \frac{512}{11}\delta} \min(1-\delta, s+\sqrt{\epsilon}+\sqrt{\delta}) \leq 1-\frac{(1-s)^2}{27}. \tag{4}$$

The last inequality is the result of a lengthy, but boring, calculation.

As a result of Eq. (4), 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.

To complete the proof, we observe that $\mathsf{QMA}(k)$ protocols can be amplified if the verifier is given many unentangled copies of the proof. Such amplification can be handled with standard methods. The only subtlety is that we treat the case of perfect completeness separately.

First, we consider $\mathsf{QMA}(p(n))_{1-1/q(n),1}$. By repeating the proof $q(n)$ times, we find that $\mathsf{QMA}(p(n))_{1-1/q(n),1} \subset \mathsf{QMA}(p(n)q(n))_{1/e,1}$. Next, using Protocol 2 and Eq. (4), we find that $\mathsf{QMA}(p(n)q(n))_{1/e,1} \subset \mathsf{QMA}(2)_{0.99,1}$. This completes the proof of the first claim of the theorem.

For the second claim, a standard Chernoff bound implies that

$$\mathsf{QMA}(p(n))_{s,c} \subset \mathsf{QMA}(2p(n)q(n)(c-s)^2)_{e^{-q(n)},1-e^{-q(n)}}.$$

Again using Protocol 2 and Eq. (4), we find that

$$\mathsf{QMA}(2p(n)q(n)(c-s)^2)_{e^{-q(n)},1-e^{-q(n)}} \subset \mathsf{QMA}(2)_{0.99,1-e^{-q(n)}}.$$

This completes the proof of Theorem 2. □

E QMA(2) protocols and the maximum output ∞-norm

In this section, we elaborate on the connection between the maximum acceptance probability of a $\mathsf{QMA}(2)$ protocol and the quantity $S_\infty^{\min}(\mathcal{N})$ (or equivalently $\|\mathcal{N}\|_{1\to\infty}$) of a quantum channel $\mathcal{N}$. This connection has been known for some time as folklore and has appeared before in Ref. [36].

The idea is seen most simply when the verifier performs a projective measurement M on the witness $\rho \in \mathrm{SEP}(d_1, d_2)$. In this case, we can define $\mathcal{N}$ to be the channel from $\mathrm{supp}\, M$ to $\mathcal{B}(d_1)$ (the space of $d_1 \times d_1$ density matrices) that is obtained by embedding $\mathrm{supp}\, M$ into $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ and then tracing out the second system. The operator M is related to the Stinespring embedding [43] of $\mathcal{N}$. Now $\|\mathcal{N}\|_{1\to\infty}$ is the largest eigenvalue of the least mixed state in the output of $\mathcal{N}$, which is equivalently the least entangled state in $\mathrm{supp}\, M$, which in turn is the maximum acceptance possibility of the protocol. In case M isn't a projector, then the corresponding $\mathcal{N}$ is now trace non-increasing instead of trace preserving, but the same argument still goes through. We formalise the correspondence as follows:

Proposition 18. *If $\mathcal{N}$ is a trace non-increasing channel with output space $\mathcal{B}(d_1)$ and d_2 Kraus operators, then there exists an operator $M \in M(d_1, d_2)$ satisfying $0 \leq M \leq I$ such that*

$$\|\mathcal{N}\|_{1\to\infty} = \max\{\mathrm{tr}\, \rho M : \rho \in \mathrm{SEP}(d_1, d_2)\} \tag{5}$$

Similarly, for any such operator M, there exists a corresponding channel $\mathcal{N}$ satisfying (5).

Proof. Given M, we define $\mathcal{N}$ as the channel from $\mathcal{B}(\operatorname{rank} M)$ to $\mathcal{B}(d_1)$ defined by $\mathcal{N}(X) = \operatorname{tr}_2(M^{1/2}XM^{1/2})$. In what follows, if V is a vector space, define $B(V)$ to be the set of unit vectors in V.

$$
\begin{aligned}
\|\mathcal{N}\|_{1\to\infty} &= \max_{|\psi\rangle\in B(\operatorname{supp} M)} \|N(\psi)\|_\infty \\
&= \max_{|\psi\rangle\in B(\operatorname{supp} M)} \|\operatorname{tr}_2 M^{1/2}\psi M^{1/2}\|_\infty \\
&= \max_{|\psi\rangle\in B(\operatorname{supp} M)} \max_{|\eta\rangle\in B(\mathbb{C}^{d_2})} \operatorname{tr}(I_{d_1}\otimes\eta)M^{1/2}\psi M^{1/2} \\
&= \max_{|\psi\rangle\in B(\operatorname{supp} M)} \max_{|\eta\rangle\in B(\mathbb{C}^{d_2})} \max_{|\beta\rangle\in B(\mathbb{C}^{d_1})} \operatorname{tr}(\beta\otimes\eta)M^{1/2}\psi M^{1/2} \quad [(I_{d_1}\otimes\eta)M^{1/2}\psi M^{1/2} \text{ has rank } 1] \\
&= \max_{|\eta\rangle\in B(\mathbb{C}^{d_2})} \max_{|\beta\rangle\in B(\mathbb{C}^{d_1})} \operatorname{tr} M^{1/2}(\beta\otimes\eta)M^{1/2} \qquad [M^{1/2}(\beta\otimes\eta)M^{1/2} \text{ has rank } 1] \\
&= \max_{|\eta\rangle\in B(\mathbb{C}^{d_2})} \max_{|\beta\rangle\in B(\mathbb{C}^{d_1})} \operatorname{tr} M(\beta\otimes\eta) = \max_{\rho\in \mathrm{SEP}(d_1,d_2)} \operatorname{tr} M\rho.
\end{aligned}
$$

These arguments can also be run backwards. In this case, we are given a channel $\mathcal{N}$ and need only to find an M such that $\mathcal{N}(X) = \operatorname{tr}_2(M^{1/2}XM^{1/2})$. If the Kraus form of $\mathcal{N}$ is $\mathcal{N}(X) = \sum_{i=1}^k N_i X N_i^\dagger$ (with $\sum_i N_i^\dagger N_i \le I$) then this is achieved by taking M to be $\sum_{i=1}^k N_i N_i^\dagger \otimes |i\rangle\langle i|$. $\square$

As a result of Proposition 18, the question of whether QMA(2) protocols can be amplified to exponentially small error is directly related to the question of multiplicativity of the maximum output infinity norm (equivalently, additivity of the minimum output min-entropy). Indeed, additivity violations for $S_\infty^{\min}$ (e.g. [28, 27, 23]) translate directly into QMA(2) protocols for which perfect parallel repetition fails[2]. Conversely, [36] observed that 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.

F Proof of correctness of the product unitary test

This appendix is devoted to the proof of Theorem 8. 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 19. *Given $U \in U(d^n)$, let*

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

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

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

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

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

This implies that we can expand

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

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.

Proof of Theorem 8. 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 19, 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 11) 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 20. *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

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

□

We note that, in this interpretation, our main result is reminiscent of the so-called inverse theorem for the second Gowers uniformity norm [20, 21], 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 $\|\hat{f}\|_\infty^4 \leq \|\hat{f}\|_4^4 \leq \|\hat{f}\|_\infty^2$, 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 [20].

Via the correspondence of Proposition 20, 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.