Quantum de Finetti Theorems Under Local Measurements with Applications

Fernando G. S. L. Brandão¹, Aram W. Harrow²

¹ Department of Physics, California Institute of Technology, Pasadena, CA, USA.
E-mail: fgslbrandao@gmail.com
² Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA.
E-mail: aram@mit.edu

Received: 21 July 2014 / Accepted: 4 March 2017
Published online: 19 April 2017 – © Springer-Verlag Berlin Heidelberg 2017

Abstract: Quantum de Finetti theorems are a useful tool in the study of correlations in quantum multipartite states. In this paper we prove two new quantum de Finetti theorems, both showing that under tests formed by local measurements in each of the subsystems one can get an exponential improvement in the error dependence on the dimension of the subsystems. We also obtain similar results for non-signaling probability distributions. We give several applications of the results to quantum complexity theory, polynomial optimization, and quantum information theory. The proofs of the new quantum de Finetti theorems are based on information theory, in particular on the chain rule of mutual information. The results constitute improvements and generalizations of a recent de Finetti theorem due to Brandão, Christandl and Yard.

1. Background

A central problem in quantum information theory, quantum computation, and physics in general is to understand entanglement: quantum correlations with no counterpart in classical probability theory. An important technique in the study of entanglement are quantum versions of the de Finetti theorem. The latter states that the marginal probability distribution $p_{X_1\ldots X_l}$ on $l$ subsystems of a permutation-invariant probability distribution $p_{X_1\ldots X_k}$ on $k \geq l$ subsystems is close (within $l(l-1)/k$ in variational distance) to a convex combination of many copies of an unknown probability distribution [37]. This is a powerful result as it allows us to infer a very particular form for $p_{X_1\ldots X_l}$ merely based on a symmetry assumption on $p_{X_1\ldots X_k}$. Note we can always make sure this assumption holds true by merely forgetting the order of the $k$ subsystems. Quantum versions of the de Finetti theorem state that an $l$-partite quantum state $\rho_{A_1\ldots A_l}$ that is a reduced state of a permutation-symmetric state $\rho_{A_1\ldots A_l}$ that is a reduced state of a permutation-symmetric state on $k \geq l$ subsystems is close (for $k \gg l$) to a convex combination of i.i.d. quantum states, i.e. $\rho_{A_1\ldots A_l} \approx \int \mu(\sigma)\sigma^\otimes l$ for a probability measure $\mu$ on quantum states.
The quantum version \cite{20,27,30,40,44,59,68,74,77,79,82} appears very similar to the original de Finetti theorem, but it is much more remarkable. Not only does it say that the correlations are arranged in an organized fashion (as a convex combination of i.i.d. states) but also that the state of $l$ subsystems is close to a *separable*, non-entangled, state. A well-known property of entanglement is that it is monogamous: A quantum system cannot be very much entangled with a large number of other systems. The quantum de Finetti theorems provide a quantitative statement for the monogamy of entanglement; in a symmetric state all the subsystems are equally correlated with all the others and so each of them can only be slightly entangled with a few of the others.

We now know several possible quantum versions of the de Finetti theorem \cite{20,27,30,40,44,59,68,74,77,79,82}. A natural way to quantify the closeness to convex combinations of i.i.d. states is by the trace norm \cite{30}. In this case Christandl, König, Mitchison, and Renner \cite{30} proved an almost optimal quantum de Finetti theorem: $\rho^{A_1 \cdots A_l}$ is $(2d^2/l/k)$-close to a convex combination of i.i.d. states in trace norm, with $d$ the dimension of the subsystems, while there are examples where the error is $\Omega(d^2 l / k)$ \cite{30}. However in many applications the scaling with dimension makes this error too large to be useful. One possible way forward is therefore to consider other ways of quantifying the approximation rather than the trace norm.

There are two known quantum de Finetti theorems following this idea. The first is the exponential de Finetti theorem of Renner \cite{77}, that achieves an exponentially small error in $k - l$, but only shows that $\rho^{A_1 \cdots A_l}$ is close to a convex combination of “almost i.i.d.” states, a generalization of i.i.d. states having similar properties with respect to certain statistical tests. The second is the de Finetti theorem proved in Ref. \cite{24}, which works for $l = 2$ and has an error of $\sqrt{T_6 \ln(d) / k}$, an exponential improvement on the dimension dependence. The approximation is quantified by the one-way LOCC norm, \cite{2} a variant of the trace norm for bipartite systems in which only measurements implementable by local operations and one-directional classical communication are allowed. Both results have found interesting applications: The first to quantum key distribution \cite{76}, quantum hypothesis testing \cite{21}, and quantum state tomography \cite{77}; the second to entanglement testing, where it gives a quasipolynomial-time algorithm for determining if a quantum state is entangled or not \cite{20}, and to quantum complexity theory \cite{20}. These two results suggest that more quantum versions of the de Finetti theorem might exist. In this paper we show that this is indeed the case.

It has emerged that some of the properties of entanglement, such as its monogamous character, are shared by more general classes of correlations \cite{66}. A particular interesting example is the class of non-signaling distributions, which are a generalization of the correlations attainable by quantum mechanics. Versions of the de Finetti theorem for non-signaling distributions have also been derived \cite{12,31}, although here again the scaling of the error – linear in the number of possible settings of the non-signaling box – has limited the applicability of the results.

The main results of this paper are two new quantum versions of the de Finetti theorem, along with extensions to arbitrary non-signaling distributions. Both are based on a coarser notion of approximation to the target state than the trace norm, but as a pay-off their error scales exponentially better with dimension. The notion of approximation used is that two quantum states are close if they have the same statistics under any local measurements on

\footnote{The trace norm gives the maximum probability of distinguishing two quantum states by arbitrary measurements.}
\footnote{The name LOCC stands for local operations and classical communication. See Eq. \eqref{65} for a precise definition of one-way LOCC, stated for the general multipartite case.}
the subsystems. Our results thus extend the de Finetti bound of Ref. [20] to an arbitrary number of subsystems while improving on the error term to depend on the number of measurements instead of the local dimension and generalizing it to general non-signaling distributions.

Notation: Let $\mathcal{D}(\mathcal{H})$ be the set of quantum states on $\mathcal{H}$, i.e. positive semidefinite matrices of unit trace acting on the vector space $\mathcal{H}$. We denote the dimension of $\mathcal{H}$ by $|\mathcal{H}|$. We say $\rho^{AB} \in \mathcal{D}(A \otimes B)$ is a $k$-extendible state if there is a state $\tilde{\rho}^{AB_1...B_k} \in \mathcal{D}(A \otimes B^{\otimes k})$ such that $\tilde{\rho}^{AB_j} = \rho^{AB}$ for all $j \in [k]$. For a multipartite state such as $\rho^{XY}$, we use the convention that omitting superscripts corresponds to taking the partial trace over those systems; e.g. $\rho^{X} = tr_{Y} \rho^{XY}$ in the previous example. Let Sep$(A : B)$ denote the set of separable states in $\mathcal{D}(A \otimes B)$, which is defined to be the convex hull of the states of the form $\rho^{A} \otimes \rho^{B}$ (product states). Similarly Sep$(A^{\otimes l})$ is the convex hull of states of the form $\rho_{1} \otimes ... \otimes \rho_{l}$. We say $\rho^{A_{1}...A_{k}} \in \mathcal{D}(A^{\otimes k})$ is permutation symmetric if $\rho_{\pi(1)...A_{\pi(k)}} = \rho^{A_{1}...A_{k}}$ for any permutation $\pi \in S_{k}$ (with $S_{k}$ the symmetric group of order $k$).

A quantum measurement (also called a POVM or positive-operator valued measure) is given by a set of matrices $\{M_{k}\}$ such that $M_{k} \succeq 0$ and $\sum_{k} M_{k} = I$. We associate to any measurement a map $\Lambda(X) = \sum_{k} \text{tr}(M_{k}X) |k\rangle\langle k|$, with $\{|k\rangle\}$ an orthonormal basis. We denote the set of maps associated to measurements by $\mathcal{M}$. These are also called quantum-classical channels, since they map quantum states to probability distributions.

Let $p(x_{1}, ..., x_{k}|a_{1}, ..., a_{k})$ be a conditional probability distribution on $\mathcal{X}^{\times k} \times \mathcal{A}^{\times k}$. We say it is non-signaling if $p(x_{j}|a_{1}, ..., a_{j})$ is independent of $a_{i}$ for $i \neq j$. We say $p(x, y|a, b)$ is $k$-extendible if there is a non-signaling distribution $p(x, y_{1}, ..., y_{k}|a_{1}, b_{1}, ..., b_{k})$ which is permutation-symmetric in the $B$ systems, i.e. $p(x, \pi^{-1}(y_{1}), ..., \pi^{-1}(y_{k})|a, \pi^{-1}(b_{1}), ..., \pi^{-1}(b_{k})) = p(x, y_{1}, ..., y_{k}|a, b_{1}, ..., b_{k})$ for all permutations $\pi \in S_{k}$, and whose marginal is $p(x, y|a, b)$, i.e.

$$\sum_{y_{2}, ..., y_{k}} p(x, y_{1}, ..., y_{k}|a_{1}, b_{1}, ..., b_{k}) = p(x, y_{1}|a_{1}, b_{1}, ..., b_{k}) = p(x, y_{1}|a, b_{1}). \quad (1)$$

We call LHV (local hidden variable) the set of conditional probability distributions of the form $p(x, y|a, b) = \sum_{l} \pi_{l} q_{l}(x|a) r_{l}(y|b)$ for a probability distribution $\pi$ and local distributions $q_{l}, r_{l}$.

2. Results

By monogamy of entanglement we expect that a $k$-extendible state $\rho^{AB}$ is close to a separable state, since the $A$ subsystem is equally correlated to $k$ systems. The next theorem gives a quantitative version of this fact both for entanglement and for non-signaling distributions.

Theorem 1.

1. Let $\rho^{AB} \in \mathcal{D}(A \otimes B)$ be a $k$-extendible state and $\mu(m)$ a distribution over quantum operations (completely positive trace-preserving maps) $\{\mathcal{E}_{m}^{A\rightarrow\tilde{A}}\}_{m}$, with $\mathcal{E}_{m}^{A\rightarrow\tilde{A}} : \mathcal{D}(A) \rightarrow \mathcal{D}(\tilde{A})$. Then

$$\min_{\sigma \in \text{Sep}(A:B)} \max_{\Lambda^{B} \in \mathcal{M} \sim \mu} \mathbb{E} \left\| \mathcal{E}_{m}^{A\rightarrow\tilde{A}} \otimes \Lambda^{B} \left( \rho^{AB} - \sigma^{AB} \right) \right\|_{1} \leq \sqrt{\frac{2 \ln |\tilde{A}|}{k}}. \quad (2)$$
2. Let $\rho^{AB} \in \mathcal{D}(A \otimes B)$ be a $k$-extendible state, $\mu(m)$ a distribution over quantum operations $\{\mathcal{E}^A_{m} \to \tilde{A}\}_m$ from $\mathcal{D}(A) \to \mathcal{D}(\tilde{A})$ and $\Lambda^B$ a measurement on $\mathcal{D}(B)$. Then in time $\text{poly}(|A|, |B|^k, \log(1/\epsilon))$ a classical computer can compute $\sigma \in \text{Sep}(A : B)$ such that

$$\mathbb{E}_{m \sim \mu} \left\| \mathcal{E}^A_{m} \otimes \Lambda^B \left( \rho^{AB} - \sigma^{AB} \right) \right\|_1 \leq \sqrt{\frac{2 \ln |\tilde{A}|}{k}} + \epsilon. \quad (3)$$

3. Let $p(x, y|a, b) \in \mathcal{X}$ be a $k$-extendible non-signaling conditional probability distribution over $\mathcal{X} \times \mathcal{Y} \times A \times B$, and let $\mu$ be a distribution over $A$. Then

$$\min_{q \in \text{LHV}} \max_{b \in B} \mathbb{E} \left[ p(x, y|a, b) - q(x, y|a, b) \right]_1 \leq \sqrt{\frac{2 \ln |X|}{k}}. \quad (4)$$

The de Finetti bound from Ref. [20] can be recovered (with an improved constant) as a special case of part 1 of Theorem 1 by choosing the singleton distribution composed of the identity channel on $A$, since

$$\| \rho^{AB} - \sigma^{AB} \|_{\text{LOCC}^\leftarrow} = \max_{\Lambda \in \mathcal{M}} \| (\text{id} \otimes \Lambda) (\rho^{AB} - \sigma^{AB}) \|_1. \quad (5)$$

The next theorem gives a generalization of the result of [20] to an arbitrary number of subsystems, as well as to non-signaling distributions.

**Theorem 2.**

1. Let $\rho^{A_1 \cdots A_k} \in \mathcal{D}(A^{\otimes k})$ be a permutation-symmetric state. Then for every $0 < l < k$ there is a measure $\nu$ on $\mathcal{D}(A)$ such that

$$\max_{\Lambda_2, \ldots, \Lambda_l \in \mathcal{M}} \left\| (\text{id} \otimes \Lambda_2 \otimes \ldots \otimes \Lambda_l) \left( \rho^{A_1 \cdots A_l} - \int \nu(d\sigma) \sigma^{\otimes l} \right) \right\|_1 \leq \sqrt{\frac{2l^2 \ln |A|}{k - l}}. \quad (6)$$

2. Let $p(x_1 \cdots x_k|a_1 \cdots a_k)$ be a permutation-symmetric non-signaling conditional probability distribution (i.e. $p$ is invariant under simultaneous permutation of the $X$ and $A$ systems). Fix a product distribution $\mu = \mu_1 \otimes \cdots \otimes \mu_k$ on $A_1 \times \cdots \times A_k$. Then for every $0 < l < k$ there is a measure $\nu$ on single-system conditional probability distributions such that

$$\mathbb{E}_{a_1, \ldots, a_l \sim \mu} \left\| p(x_1 \cdots x_l|a_1, \ldots, a_l) - \mathbb{E}_{q \sim \nu} q(x_1|a_1) \otimes \cdots \otimes q(x_l|a_l) \right\|_1 \leq \sqrt{\frac{2l^2 \ln |X|}{k - l}}. \quad (7)$$

---

3 The bipartite one-way LOCC norm is defined as $\|X\|_{\text{LOCC}^\leftarrow} = \max_{0 \leq M \leq I} \text{tr}(XM)$, with the maximization over all POVMs $\{M, I - M\}$ that can be realized by local operations and one-way classical communication from $B$ to $A$. 
2.1. Discussion and comparison with previous work. Theorem 1 improves on Ref. [20] in several ways. First and most importantly, the error term is independent of the subsystem dimensions of $\rho^{AB}$, and only depends on the output dimension of the family of quantum operations $\{\mathcal{E}_{m}^{A \rightarrow \tilde{A}}\}_{m}$, thus yielding nontrivial bounds even if $A$ is infinite-dimensional. Likewise, for non-signaling distributions the bound in part 3 is independent of the number of measurement settings of $p(x, y|a, b)$. Second, if we think of $k$-extendable states as a relaxation of Sep, then part 2 provides an “explicit rounding” (i.e., actually produces the separable approximation), which did not exist in Ref. [20], although we note the caveat that $\sigma$ depends on the measurement $\Lambda^B$. Third, part 3 generalizes the result to non-signaling distributions. Note that in part 1, taking system $\tilde{A}$ to be classical would yield a special case of part 3, but in the more general case where $\tilde{A}$ is a quantum system, parts 1 and 3 are incomparable.

We remark that (2) (in the case where $\mathcal{E}_m = \text{id}$) also follows from the work of Yang [84] using the fact that the entanglement of formation [17] is upper-bounded by the log of either of the local dimensions, together with a variant of the Pinsker inequality adapted to LOCC$^{-}\text{adm}$ [72]. It also follows from the recent work of Li and Winter [64].

The proof of Theorem 1 (found in Sect. 3) is more direct and general than the proofs in [20,64,84], in particular not making use of entanglement measures in any explicit way. This enables us to obtain parts 2 and 3 of the theorem (but see the discussion of Conjecture 8 for an example of how the generality of Theorem 1 limits our ability to further improve it).

We remark that the explicit rounding in part 2 was mostly known only for the variants requiring $k \geq |B|$ [30,39,68,77], and the previous de Finetti theorems for non-signaling boxes [12,31] similarly required $k$ to scale as some power of the local dimension. This means that the resulting algorithms for approximating Sep would take time exponential in the dimension. At first glance it would appear that our (2) is incomparable to the previous results; for example, Thm 2 of [68] implies that if $\rho$ is $k$-extendable then there exists $\sigma \in \text{Sep}(A : B)$ with $\|\rho - \sigma\|_1 \leq 2/(1 + k/|B|)$. In Proposition 4 (found in Sect. 3), we show how similar bounds can be obtained with our information-theoretic methods.

One other previous work to find $k$ scaling logarithmically with dimension is [11, Section 3], which achieves a similar but incomparable bound for measurements with nonnegative matrix elements, together with an efficient rounding scheme. An early version of Ref. [11] was also an important source of inspiration for the current work.

Concerning Theorem 2, in Ref. [37] Diaconis and Freedman proved in 1980 that for a permutation-symmetric probability distribution $p_k$ on $k$ subsystems, its $l$-body marginal $p_l$ is $\frac{l(l-1)}{k}$-close (in variational distance) to a convex combination of i.i.d. probability distributions. Theorem 2 can be seen as an analogue of this result to quantum states and non-signaling probability distributions. However instead of having a bound that is independent of the dimension, we only have a bound that depends logarithmically on the dimension (and the notion of approximation is weaker than variational distance). It is an interesting question whether this can be improved. Note however that we give in Sect. 5.2 a computational complexity argument that the $k = \Omega(l^2)$ dependency is optimal.

Just as Theorem 1 yielded a stronger version of the BCY result [20] as a corollary, Theorem 2 leads to a multipartite version of the BCY bound (5). The main difference, apart from considering states on $l$ systems that have symmetric $k$-partite extensions, is that the bipartite LOCC$^{-}\text{adm}$ norm is replaced by one in which parties $2, \ldots, l$ measure their systems and communicate the outcomes to party 1, who can then choose a measurement
adaptively based on these messages. This leads to a norm on states that can be thought of as a multipartite generalization of the LOCC$\leftarrow$ norm. We note that [22] also derived a multipartite generalization of [20] but with an exponentially worse scaling of $k$ with $l$. After the initial arXiv version of this paper, part 1 of our Theorem 1 was improved to handle a multipartite LOCC$\leftarrow$ norm in which $l$ sends a message to $l−1$ who measures and sends a message to $l−2$, etc. [63].

We note our approach here is in a sense a throwback to the original quantum de Finetti theorem of Caves, Fuchs and Schack [27]. They applied an informationally complete measurement to the quantum systems and applied a classical de Finetti theorem to the resulting outcomes. However, they could only prove that the approximation tends to zero as $k \to \infty$ with no control on the rate, while our error bounds are optimal or nearly-optimal in many settings. There are a few differences in our approaches. Ref. [27] does full tomography on the quantum state, while we only consider a specific measurement $A^B$, which is what enables us to have error scaling with the log of the dimension, but also which limits us only bounding error in the 1-LOCC norm. More importantly, Ref. [27] treats the classical de Finetti theorem as a black box, which means that approximation errors from de Finetti can become uncontrolled when inverting state tomography. Our approach essentially reproves the de Finetti theorem (as we discuss further in Section 4 of our follow-up paper [23]) and by keeping track of the states along the way, we avoid having to carry out such a state-reconstruction task. In a way, this was also the idea behind [59], which gave the first finite quantum de Finetti theorem. Our approach differs from [27,59] and most of the other previous work by avoiding state reconstruction and informationally complete measurements altogether; instead our “rounding” step is achieved by replacing a state $\rho^{AB}$ with $\rho^A \otimes \rho^B$ in a situation where this is guaranteed to cause little error.

2.2. Applications. Among the applications of the new quantum de Finetti theorems we address two problems in quantum complexity theory. Below we give a brief description of these applications. A more extended presentation is given in Sect. 5 of the paper

**Multiple Unentangled Proofs:** The first application concerns a protocol due to Chen and Drucker [28] in which a prover sends to a verifier $\sqrt{n}$ polylog($n$) unentangled quantum states, each composed of $O(\log(n))$ qubits, as a proof of the satisfiability of a 3-SAT instance with $n$ variables and $O(n)$ clauses. The quantum verifier then checks the validity of the proof by performing local quantum measurements on each of the proofs and post-processing the outcomes. This result (building on [2]), is surprising since one can convince a verifier of the satisfiability of a 3-SAT instance by sending only $\sqrt{n}$ polylog($n$) qubits! It is a natural question whether the total number of qubits could be decreased even further. As a direct application of one of the new quantum de Finetti theorems we give strong evidence against any further decrease: We show that any similar protocol with $O(n^{1/2-\varepsilon})$ qubits, for any $\varepsilon > 0$, would imply a $\exp(n^{1-2\varepsilon}\text{polylog}(n))$-time algorithm for 3-SAT. This proves the optimality of the protocol under the plausible assumption that there are no subexponential-time algorithms for SAT [46].

A related, but harder, problem is whether QMA(2) protocols can give at most a quadratic reduction in proof size with respect to QMA.$^4$ We believe the result we

---

$^4$ QMA is the quantum version of NP. QMA(2), in turn, is a version of QMA in which one is given two proofs, with the promise they are not entangled with each other; see Sect. 5.2.

$^5$ By Ref. [42] we know QMA(2) with constant soundness gives at least a quadratic reduction in proof size relative to QMA, under plausible computational complexity assumptions; see Sect. 5.2.
obtain gives evidence that this might be the case and a suitable quantum version of the de Finetti theorem (improving LOCC→ SEP measurements) might be the right tool to show it.\footnote{See \cite{20,42} for more evidence this might be the case, along with obstacles to prove it.}

**Non-local Games:** The second application concerns the computational complexity of non-local games. We give two results in this direction. The first is algorithmic and concerns the class of free games, defined as games in which the questions to each prover are chosen independently. We show that the maximum winning probability of such games can be approximated within additive error \(\epsilon\) by a linear program in time

\[
\exp\left(\Omega\left(\log |Q| + \log^2 |A|/\epsilon^2\right)\right),
\]

with \(|Q|\) and \(|A|\) the question and answer alphabet sizes, respectively. The run-time matches the performance of a different algorithm for the problem due to Aaronson, Impagliazzo and Moshkovitz \cite{3}.\footnote{This algorithm was communicated to us already in 2010, although the result has appeared publicly only in \cite{3}.} Although this is a purely classical result, we establish it by exploring a connection to non-local games: We show that for any two-player one-round free game, one can find another game on \(m\) players such that the maximum winning probability under non-signaling strategies, which can be computed by a linear program \cite{48}, gives a \(\sqrt{\ln |A|/2m}\)-additive approximation to the maximum winning probability of the original game. Since the non-signaling value of a game is always larger than the entangled value, the same result holds also for games in which the players share entanglement.

Using the observation above for entangled strategies, together with a hardness result for free games from \cite{3}, we also show that 3-SAT on \(n\) variables can be reduced to obtaining a constant error approximation of the maximum winning probability under entangled strategies of \(O(\sqrt{n})\)-player one-round non-local games, in which the players communicate \(O(\sqrt{n})\) bits all together. Finally, we show how one would be able to establish NP-hardness of approximating the maximum winning probability under entangled strategies of a 4-player one-round game if one could strengthen appropriately one of the new quantum de Finetti theorems of this paper. This gives a new approach to this problem, which was only recently resolved \cite{81}.

**Polynomial Optimization:** We consider the connection \cite{10,38,39} between quantum de Finetti theorems and the optimization over separable states, on one hand, and polynomial optimization and the Sum-of-Squares (Parrilo/Lasserre) hierarchy, on the other hand, and prove that the optimization of certain degree-\(d\) polynomials over the \(n\)-dimensional hypersphere can be approximated to additive error \(\epsilon\) in quasipolynomial-time in the number of variables by considering \(O(\log(n)d^3\epsilon^{-2})\) rounds of the Sum-of-Squares hierarchy of semidefinite programs. This result can be considered as an extension to the hypersphere of similar results for the simplex \cite{73}. Moreover, employing the result of Chen and Drucker \cite{28}, we show that \(\Omega(d^2)\) rounds are necessary to obtain even a constant error-approximation, unless there are subexponential-time algorithms for SAT.

**Separability Testing:** Another application is to give an algorithm for deciding separability of multipartite states which is quasi-polynomial in the local dimensions of the subsystems. Given a multipartite state \(\rho_{A_1\ldots A_l}\), we prove one can decide whether it is fully separable or \(\epsilon\)-away from separable in time

\[
\exp\left(\Omega\left(\left(\sum_k \ln |A_k|\right)^2 l^2 \epsilon^{-2}\right)\right),
\]

with distance measured either by the one-way LOCC norm \cite{22} or by a multipartite version
of the Frobenius norm introduced in [60]. This generalizes the findings of [20] from bipartite states to general multipartite states, and vastly improves on the bound of [22].

**Efficient State Tomography:** A final application of the new de Finetti theorems is to quantum state tomography. The starting point is a result due to Aaronson [1], based on computational learning theory, showing that given an unknown $n$-qubit state one can perform tomography that allows us to compute to good accuracy the statistics of most observables (with respect to an arbitrary a priori distribution over observables) by measuring only $O(n)$ i.i.d. copies of the state. The new de Finetti theorem we prove allow us to relax the assumption of having i.i.d. copies of the state (which can never be fully certified), showing that essentially the same conclusion holds true for arbitrary quantum states as long as one can select a few of its subsystems at random and perform the original scheme on them. The cost of this flexibility is that we increase the number of subsystems needed from $O(n)$ to $\text{poly}(n)$, of which only $O(n)$ are measured and the rest discarded.

3. Proof of Theorem 1

We will prove Theorem 1 by information-theoretic techniques, inspired by [11] and Lemma 4.5 of [75]. We will also state and prove Proposition 4, establishing an alternate bound that depends only on the dimension of the $B$ system.

Given two quantum states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, we define the quantum relative entropy (or quantum Kullback-Leibler divergence) as

$$S(\rho || \sigma) := \text{tr}(\rho (\ln(\rho) - \ln(\sigma))).$$  

(8)

Given a bipartite state $\rho^{AB} \in \mathcal{D}(A \otimes B)$ we define the mutual information as

$$I(A : B)_\rho := S(\rho^{AB} || \rho^A \otimes \rho^B).$$  

(9)

Given a tripartite quantum state of the form $\rho^{ABK} := \sum_k p_k \rho_k^{AB} \otimes |k\rangle \langle k|^{K}$ we define the conditional mutual information as

$$I(A : B|X)_\rho := \sum_k p_k I(A : B)_{\rho_k}. $$  

(10)

The mutual information satisfies the following properties that will be useful in the proof:

**Lemma 3.**

1. **Chain Rule:**

$$I(A : BX) = I(A : X) + I(A : B|X) $$  

(11)

2. **Monotonicity under Local Operations:** Let $\pi^{AB} = \text{id} \otimes \Lambda(\rho^{AB})$, for a quantum operation $\Lambda$. Then

$$I(A : B)_\pi \leq I(A : B)_\rho $$  

(12)

3. **Pinsker’s Inequality:**

$$I(A : B)_\rho \geq \frac{1}{2} \| \rho^{AB} - \rho^A \otimes \rho^B \|^2_1. $$  

(13)
Proof. 1. Direct calculation. 2. This is essentially Holevo’s theorem and follows from strong subadditivity. See Chapter 11 of [83] or Chapter 12 of [69]. 3. See Section 10.8.1 of [83]. □

(The absence of the usual \( \ln(2) \) factor in (13) is because of our convention that entropies are measured in “nats,” i.e. with logs taken base \( e \).)

We are now ready to prove Theorem 1:

**Theorem 1.** (restatement).

1. Let \( \rho^{AB} \in \mathcal{D}(A \otimes B) \) be a k-extendible state and \( \mu(m) \) a distribution over quantum operations (completely positive trace-preserving maps) \( \{ \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \}_{m} \), with \( \mathcal{E}_{m}^{A \rightarrow \tilde{A}} : \mathcal{D}(A) \rightarrow \mathcal{D}(\tilde{A}) \). Then

\[
\min_{\sigma \in \text{Sep}(A:B)} \max_{\Lambda^{B} \in \mathcal{M}} \mathbb{E} \left[ \left\| \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \otimes \Lambda^{B} \left( \rho^{AB} - \sigma^{AB} \right) \right\|_{1} \right] \leq \sqrt{\frac{2 \ln | \tilde{A} |}{k}}.
\]  

(14)

2. Let \( \rho^{AB} \in \mathcal{D}(A \otimes B) \) be a k-extendible state, \( \mu(m) \) a distribution over quantum operations \( \{ \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \}_{m} \) from \( \mathcal{D}(A) \rightarrow \mathcal{D}(\tilde{A}) \) and \( \Lambda^{B} \) a measurement on \( \mathcal{D}(B) \). Then in time \( \text{poly}(|A|, |B|^{k}, \log(1/\varepsilon)) \) a classical computer can compute \( \sigma \in \text{Sep}(A:B) \) such that

\[
\mathbb{E} \left[ \left\| \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \otimes \Lambda^{B} \left( \rho^{AB} - \sigma^{AB} \right) \right\|_{1} \right] \leq \sqrt{\frac{2 \ln | \tilde{A} |}{k}} + \varepsilon.
\]  

(15)

3. Let \( p(x, y|a, b) \in \mathcal{X} \) be a k-extendible non-signaling conditional probability distribution over \( \mathcal{X} \times \mathcal{Y} \times \tilde{A} \times \mathcal{B} \), and let \( \mu \) be a distribution over \( \tilde{A} \). Then

\[
\min_{q \in \text{LHV}} \max_{b \in \mathcal{B}} \mathbb{E} \left[ \| p(x, y|a, b) - q(x, y|a, b) \|_{1} \right] \leq \sqrt{\frac{2 \ln | \mathcal{X} |}{k}}.
\]  

(16)

Proof. The three parts of the theorem have similar proofs.

**Part 1:**

Define the states

\[
\pi_{m}^{\tilde{A}B_{1} \ldots B_{k}} := \left( \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \otimes \Lambda_{1}^{B_{1}} \otimes \ldots \otimes \Lambda_{k}^{B_{k}} \right) (\rho^{AB_{1} \ldots B_{k}}),
\]  

(17)

\[
\pi^{\tilde{A}B_{1} \ldots B_{k}M} := \mathbb{E}_{m \sim \mu} \pi_{m} \otimes |m\rangle \langle m|^{M}
\]  

(18)

with \( \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \) quantum operations from \( A \) to \( \tilde{A} \), \( \Lambda_{i} \) quantum-classical channels, and \( |m\rangle \) a classical label for which quantum operation \( \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \) was applied. Repeatedly applying the chain rule (11), we find

\[
I(\tilde{A} : B_{1} \ldots B_{k} | M)_{\pi} = I(\tilde{A} : B_{1} | M)_{\pi} + I(\tilde{A} : B_{2} | MB_{1})_{\pi} + \ldots + I(\tilde{A} : B_{k} | MB_{1} \ldots B_{k-1})_{\pi}.
\]  

(19)
Now we maximize over measurements and obtain
\[
\max_{\Lambda_1, \ldots, \Lambda_k \in \mathcal{M}} I(\tilde{A} : B_1 \ldots B_k | M)_{\pi} \\
= \max_{\Lambda_1, \ldots, \Lambda_{k-1} \in \mathcal{M}} \left( I(\tilde{A} : B_1 | M)_{\pi} + \ldots + I(\tilde{A} : B_{k-1} | MB_1 \ldots B_{k-2})_{\pi} + \max_{\Lambda_k \in \mathcal{M}} I(A : B_k | MB_1 \ldots B_{k-1})_{\pi} \right). \tag{20}
\]

Now
\[
I(A : B_k | MB_1 \ldots B_{k-1})_{\pi} = \mathbb{E}_{m \sim \mu} I(A : B_k | B_1 \ldots B_{k-1})_{\pi_m}. \tag{21}
\]

Since the $B_1 \ldots B_{k-1}$ systems of $\pi_m$ are classical, we can write the state of $\rho^{AB_k}$ as an average over them, namely
\[
\rho^{AB} = \rho^{AB_k} = \sum_i q_i \rho^{AB_k}_i, \tag{22}
\]
where $\{q_i, \rho_i\}$ depend on $\Lambda_1, \ldots, \Lambda_{k-1}$ but not on $\mathcal{E}_m$ and $\Lambda_k$. Then define
\[
\pi_{AB_k}^{m,i} := \left( \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda^{B_k} \right) (\rho^{AB_k}_i), \tag{23}
\]
so that $\pi_{AB_k}^{m,i} = \sum_i q_i \pi_{AB_k}^{m,i}$ and
\[
I(A : B_k | B_1 \ldots B_{k-1})_{\pi_m} = \sum_i q_i I(A : B_k)_{\pi_{m,i}} \tag{24}
\]

By Pinsker’s inequality
\[
I(A : B_k | B_1 \ldots B_{k-1})_{\pi_m} \geq \frac{1}{2} \sum_i q_i \left\| \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda^{B_k} (\rho_i - \rho_i^A \otimes \rho_i^{B_k}) \right\|_1^2, \tag{25}
\]
where $\rho_i^A$ and $\rho_i^{B_k}$ are the $A$ and $B_k$ reduced states of $\rho_i$.

By convexity of $x^2$ and the trace norm
\[
I(A : B_k | B_1, \ldots, B_{k-1})_{\pi_m} \geq \frac{1}{2} \left\| \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda^{B_k} \left( \rho^{AB} - \sum_i q_i \rho_i^A \otimes \rho_i^{B_k} \right) \right\|_1^2. \tag{26}
\]

Using Eq. (21)
\[
\max_{\Lambda_k \in \mathcal{M}} I(A : B_k | MB_1 \ldots B_{k-1})_{\pi} \geq \frac{1}{2} \max_{\Lambda_k \in \mathcal{M}} \mathbb{E}_{m \sim \mu} \left\| \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda^{B_k} \left( \rho^{AB_k} - \sum_i q_i \rho_i^A \otimes \rho_i^{B_k} \right) \right\|_1^2 \\
\geq \frac{1}{2} \min_{\sigma \in \text{SEP}(A:B_k)} \max_{\Lambda_k \in \mathcal{M}} \mathbb{E}_{m \sim \mu} \left\| \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda^{B_k} (\rho - \sigma) \right\|_1^2. \tag{27}
\]
Quantum de Finetti Theorems Under Local Measurements with Applications

| Part 1         | Part 3                      |
|----------------|-----------------------------|
| quantum states \(\rho^{AB_1\ldots B_k}\) | non-signaling distributions \(p(x, y_1, \ldots, y_k|a, b_1, \ldots, b_k)\) |
| quantum mutual information | classical mutual information maximized over choices of measurements \(a, b_1, \ldots, b_k\) |
| partial trace | taking marginals (allowed by no-signaling condition) |

Note that the second line is independent of \(\Lambda_1, \ldots, \Lambda_{k-1}\), since only the ensemble \(\{q_i, \rho_i\}\) depended on them.

From (20) and (27),

\[
\max_{\Lambda_1, \ldots, \Lambda_{k-1} \in \mathcal{M}} I(A : B_1 \ldots B_k | M)_{\pi} \geq \max_{\Lambda_1, \ldots, \Lambda_{k-1} \in \mathcal{M}} \sum_{j=1}^{k-1} I(A : B_j | MB_1 \ldots B_{j-1})_{\pi} + \frac{1}{2} \min_{\sigma \in \text{SEP}(A:B_k)} \max_{\Lambda_k \in \mathcal{M}} m_{\sim \mu} \| \mathcal{E}_{m}^{A \rightarrow \tilde{A}} \otimes \Lambda_k (\rho^{AB} - \sigma^{AB}) \|_1^2.
\]

Applying the same argument sequentially to all the remaining conditional mutual informations we find

\[
\frac{k}{2} \min_{\sigma \in \text{SEP}(A:B)} \max_{\Lambda \in \mathcal{M}} m_{\sim \mu} \| \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda (\rho^{AB} - \sigma^{AB}) \|_1^2 \leq \max_{\Lambda_1, \ldots, \Lambda_{k}} I(A : B_1 \ldots B_k | M)_{\pi} \leq \ln |\tilde{A}|,
\]

where we used that \(\pi \tilde{A} = \mathbb{E}_{m \sim \mu} (\mathcal{E}_m^{A \rightarrow \tilde{A}}(\rho^A)) \in \mathcal{D}(\tilde{A})\). Finally by convexity of \(\lambda^2\),

\[
\left( \min_{\sigma \in \text{SEP}(A:B)} \max_{\Lambda \in \mathcal{M}} m_{\sim \mu} \| \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda (\rho^{AB} - \sigma^{AB}) \|_1^2 \right)^{\frac{1}{k}} \leq \frac{2 \ln |\tilde{A}|}{k},
\]

and we are done with the proof of part 1.

**Part 2**: The proof of part 2 is mostly the same as that of part 1, and so we only give a brief outline of the changes. The main change is to omit the maximizations over \(\Lambda_1, \ldots, \Lambda_k\), instead using only the fixed measurement \(\Lambda\). We also set \(\sigma = \sum_i q_i \rho_i^A \otimes \rho_i^{B_k}\) rather than performing a minimization. The algorithm is as follows: Perform \(\Lambda\) on systems \(B_1, \ldots, B_j\) for \(j\) uniformly chosen between 0 and \(k-1\). For each measurement outcome and each \(j\), let \(\omega_{AB} \) denote the resulting state on \(AB_j\). Let \(\sigma\) denote the resulting average over \(\omega^A \otimes \omega^B\). These calculations require only time polynomial in the dimensions of the relevant states. Observe in particular that the algorithm is independent of \(\mathcal{E}_m\) and so does not depend on the number of different \(m\) or other features of this distribution. Taking issues of numerical precision into account adds a further error of \(\epsilon\) and multiplies the run-time by \(\text{poly log}(1/\epsilon)\).

**Part 3**: The proof of part 3 is similar to that of part 1, except that we need to make the following replacements. For brevity we will use the abbreviations \(b_k^k := (b_1, \ldots, b_k)\), \(b_k^{k-1} = (b_1, \ldots, b_{k-1})\) and so on. In more detail, the analogue of (18) is to define the non-signaling distribution \(\pi\) from \(\mathcal{B}_k^k \rightarrow \mathcal{X} \times \mathcal{Y}^k:\n
\pi(x, y^k, a|b^k) = \mu(a) p(x, y^k|a, b^k)
\]
We can also define
\[
\pi(x, y^{k-1}, a|b^{k-1}) = \mu(a) p(x, y^{k-1}|a, b^k),
\]
and, thanks to the no-signaling property of \(p\), this is well-defined, since the RHS does not depend on \(b_k\).

Again the chain rule gives us an analogue of (20).

\[
\max_{b^k \in \mathcal{B}^k} I(X : Y^k | A)_{\pi(\cdot|b^k)} = \max_{b^{k-1} \in \mathcal{B}^{k-1}} \left( \sum_{j=1}^{k-1} I(X : Y_j | AY^{j-1})_{\pi(\cdot|b^{k-1})} + \max_{b_k \in \mathcal{B}} I(X : Y_k | AY^{k-1})_{\pi(\cdot|b^k)} \right)
\]

Again we focus on the last term of Eq. (33). Define \(i := (a, b^k, y^{k-1})\), and compute

\[
\max_{b_k \in \mathcal{B}} I(X : Y_k | AY^{k-1})_{\pi(\cdot|b^k)} = \max_{b_k \in \mathcal{B}} \mathbb{E}_{a \sim \mu} I(X : Y_k)_{p(\cdot|a, b^k)}
\]

\[
\geq \frac{1}{2} \max_{b_k \in \mathcal{B}} \mathbb{E}_{a \sim \mu} \sum_{y^{k-1}} p(y^{k-1}|b^{k-1}) \left( \sum_{x \in \mathcal{X}} \sum_{y_k \in \mathcal{Y}} |p(x, y_k|i) - p(x|i) p(y_k|i)| \right)^2
\]

\[
\text{Pinsker}
\]

\[
\geq \frac{1}{2} \max_{b_k \in \mathcal{B}} \mathbb{E}_{a \sim \mu} \left( \sum_{y^{k-1}} p(y^{k-1}|b^{k-1}) \sum_{x \in \mathcal{X}} \sum_{y_k \in \mathcal{Y}} |p(x, y_k|i) - p(x|i) p(y_k|i)| \right)^2
\]

\[
\text{convexity of } x \mapsto x^2
\]

\[
\geq \frac{1}{2} \max_{b_k \in \mathcal{B}} \mathbb{E}_{a \sim \mu} \left( \sum_{x \in \mathcal{X}} \sum_{y_k \in \mathcal{Y}} \left| p(x, y_k|a, b^k) - \sum_{y^{k-1}} p(y^{k-1}|b^{k-1}) p(x|i) p(y_k|i) \right| \right)^2
\]

\[
\text{convexity of } \| \cdot \|_1
\]

\[
\geq \frac{1}{2} \min_{q \in \text{LHV}} \max_{b_k \in \mathcal{B}} \mathbb{E} \left\| p(X, Y_k|a, b) - q(X, Y_k|a, b) \right\|^2_1
\]

As with part 1, we can repeatedly apply this inequality to (33) in order to prove the theorem. \(\square\)

We conclude this section by comparing our work with previous results on the distance of \(k\)-extendable states to separable states. Our bound, like that of [20], requires \(k \gtrsim \log |\mathcal{A}|\) to achieve constant error in the 1-LOCC norm (ignoring for simplicity the possibility that \(|\tilde{\mathcal{A}}| \ll |\mathcal{A}|\)). By contrast, previous work [30,39,68,77] generally required that \(k \geq \text{poly}(|\mathcal{B}|)\) to achieve constant error in the trace norm. It is possible to convert
from the 1-LOCC norm to the trace norm by giving up a factor of dimension [67], but at first it might seem that dependence on $|A|$ is incomparable to dependence on $|B|$. However, it turns out that our information-theoretic methods can indeed also work in the case where $k \geq \text{poly}(|B|)$, although with polynomially worse scaling than [68].

**Proposition 4.** Suppose that $\rho^{AB}$ has a symmetric extension $\tilde{\rho}^{AB_1\ldots B_k}$. That is, $\rho^{AB} = \tilde{\rho}^{AB_i}$ for each $i$ and $\text{supp}\tilde{\rho}^{B_1\ldots B_k}$ is in the symmetric subspace of $B^{\otimes k}$. Then there exists $\sigma \in \text{Sep}(A : B)$ with

$$
\|\rho - \sigma\|_1 \leq 6|B|^2\sqrt{\frac{\ln(k+1)}{k}}. \tag{34}
$$

It is known [76, Lemma 4.2.2] that if $\rho^{AB}$ is $k$-extendable then we can find a purification of the extension $|\psi^{AA'B_1'B_1\ldots B_k'B_k}\rangle$ that lies in the symmetric subspace of $(B \otimes B')^{\otimes k}$ (i.e. is a $+1$ eigenstate of all $k!$ permutations of $B_1 B'_1, \ldots, B_k B'_k$). This means that whenever a state has an extension we can construct a symmetric extension of this form at the cost of squaring the local dimensions (i.e. replacing $A$ with $AA'$ and $B$ with $BB'$).

**Proof.** Since the dimension of the symmetric subspace is $(|B|+k-1)\choose k \leq (k+1)|B|\ln(k+1)$. Let $\Lambda$ be an informationally complete measurement on $B$. This is due to [60], but we will use the formulation of Lemma 16 of [23]. It states that

$$
\|\Lambda(X)\|_1 \geq \frac{\|X\|_1}{\sqrt{18|B|}} \tag{35}
$$

for any matrix $X$. We will in fact need a variant of this bound:

$$
\|\text{id}_A \otimes \Lambda_B(X)\|_1 \geq \frac{\|X\|_1}{\sqrt{18|B|^3}}. \tag{36}
$$

To prove this, define $\Lambda^{-1}$ to be the inverse of $\Lambda$ restricted to the range of $\Lambda$. Then (35) is equivalent to the statement that $\|\Lambda^{-1}\|_1 \to 1 \leq \sqrt{18|B|}$ and (36) is equivalent to $\|\text{id}_A \otimes \Lambda^{-1}_B\|_1 \to 1 \leq \sqrt{18|B|^3}$. Here $\|T\|_1 := \sup_{X \neq 0} \frac{\|T(X)\|_1}{\|X\|_1}$ for $T$ a linear map on $d \times d$ matrices. From Lemma 23 of [43] and Theorem 11.1 of [56] we have that for any $d'$, $\|\text{id}_{d'} \otimes T\|_1 \to 1 \leq \text{min}(d, d')\|T\|_1$. Setting $d' = |B|$, (36) follows.

The rest of the proof proceeds in a way similar to that of Theorem 1. Let $\pi := (\text{id} \otimes \Lambda^{\otimes k})(\tilde{\rho})$. Using first data processing and then the chain rule we find that

$$
|B| \ln(k) \geq I(A : B_1 \ldots B_k)_{\tilde{\rho}} \geq I(A : B_1 \ldots B_k)_{\pi} = \sum_{j=1}^k I(A : B_j | B_{<j})_{\pi}. \tag{37}
$$

Thus there exists $j$ for which

$$
I(A : B_j | B_{<j})_{\pi} \leq |B| \frac{\ln(k)}{k}. \tag{38}
$$

Now consider the state that would result from measuring only $B_1, \ldots, B_{j-1}$ using $\Lambda$. Call the outcome of this measurement $i$, the post-measurement state $\rho_i^{AB_j}$, and the probability of this outcome $p_i$. Then $\rho^{AB} = \sum_i p_i \rho_i^{AB_j}$, and $I(A : B_j | B_{<j})_{\pi}$.
\[ B_j | B_{<j} \rangle_\pi = \sum_i p_i I(A : B_j)(\mathrm{id} \otimes \Lambda)(\rho) \]. Now using Pinsker’s inequality and convexity we find that
\[
\frac{1}{2} \left\| (\mathrm{id} \otimes \Lambda)(\rho^{AB} - \sigma^{AB}) \right\|_1^2 \leq |B| \frac{\ln(k)}{k},
\] (39)
with \( \sigma := \sum_i p_i \rho_i^A \otimes \rho_i^B \). Applying (36) to (39) completes the proof. \( \square \)

4. Proof of Theorem 2

For a state \( \rho^{A_1 \ldots B_k} \) we define the multipartite mutual information
\[
I(A_1 : \ldots : A_k) := S(\rho^{A_1 \ldots A_k} || \rho^{A_1} \otimes \ldots \otimes \rho^{A_k}) = S(A_1) + \ldots + S(A_k) - S(A_1 \ldots A_k).
\] (40)
For a quantum-classical state \( \rho^{A_1 \ldots A_k R} = \sum_i p_i \rho_i^{A_1 \ldots A_k} \otimes |i\rangle \langle i|^R \) we define the conditional multipartite mutual information as follows
\[
I(A_1 : \ldots : A_k | R)_{\rho} := \sum_i p_i I(A_1 : \ldots : A_k)_{\rho_i}.
\] (41)

The multipartite mutual information satisfies the following properties:

Lemma 5.

1. Multipartite-to-Bipartite [85]:
\[
I(A_1 : \ldots : A_k | R) = I(A_1 : A_2 | R) + I(A_1 A_2 : A_3 | R) + \ldots + I(A_1 \ldots A_{k-1} : A_k | R).
\] (42)
2. Monotonicity under Local Operations: Let \( \Lambda^{A_1 \ldots A_k} = \Lambda^{A_1} \otimes \mathrm{id}^{A_2 \ldots A_k} (\rho^{A_1 \ldots A_k}) \), then
\[
I(A_1 : \ldots : A_k)_{\Lambda} \leq I(A_1 : \ldots : A_k)_{\rho}
\] (43)
3. Pinsker’s Inequality:
\[
I(A_1 : \ldots : A_k)_{\rho} \geq \frac{1}{2} \left\| \rho^{A_1 \ldots A_k} - \rho^{A_1} \otimes \ldots \otimes \rho^{A_k} \right\|_1^2.
\] (44)

The proof of parts 2 and 3 is essentially the same as that of Lemma 3.

Theorem 2. (restatement).

1. Let \( \rho^{A_1 \ldots A_k} \in \mathcal{D}(A^\otimes k) \) be a permutation-symmetric state. Then for every \( 0 < l < k \) there is a measure \( \nu \) on \( \mathcal{D}(A) \) such that
\[
\max_{\Lambda_2, \ldots, \Lambda_l \in \mathcal{M}} \left\| (\mathrm{id} \otimes \Lambda_2 \otimes \ldots \otimes \Lambda_l) \left( \rho^{A_1 \ldots A_l} - \int \nu(d\sigma)\sigma^\otimes l \right) \right\|_1
\leq \sqrt{\frac{2l^2 \ln|A|}{k - l}}.
\] (45)
2. Let \( p(x_1 \cdots x_k | a_1 \cdots a_k) \) be a permutation-symmetric non-signaling conditional probability distribution (i.e. \( p \) is invariant under simultaneous permutation of the \( X \) and \( A \) systems). Fix a product distribution \( \mu = \mu_1 \otimes \cdots \otimes \mu_k \) on \( A_1 \times \cdots \times A_k \). Then for every \( 0 < l < k \) there is a measure \( \nu \) on single-system conditional probability distributions such that

\[
\mathbb{E}_{a_1, \ldots, a_l \sim \mu} \left\| p(x_1 \cdots x_l | a_1, \ldots, a_l) - \mathbb{E}_{q \sim \nu} q(x_1 | a_1) \otimes \cdots \otimes q(x_l | a_l) \right\|_1 \leq \sqrt{\frac{2l^2 \ln |X|}{k-l}}.
\]

(46)

**Proof. Part 1:**

Let

\[
\pi^{A_1 \cdots A_l R} := (\text{id}_{A_1} \otimes \Lambda_2 \otimes \cdots \otimes \Lambda_l \otimes \mathcal{E}^{A_{l+1} \cdots A_k})(\rho^{A_1 \cdots A_k}),
\]

(47)

with \( \Lambda_j : \mathcal{D}(A) \to \mathcal{D}(X) \) and \( \mathcal{E} : \mathcal{D}(A^{\otimes k-l}) \to \mathcal{D}(R) \) quantum-classical channels. Then from Eq. (42) of Lemma 5,

\[
\min_{\mathcal{E}} \max_{\Lambda_2, \ldots, \Lambda_l} I(A_1 : \ldots : A_l | R)_\pi^{A_1 \cdots A_l R} = \min_{\mathcal{E}} \max_{\Lambda_2, \ldots, \Lambda_l} \sum_{j=2}^l I(A_1 \ldots A_{j-1} : A_j | R)_\pi^{A_1 \cdots A_l R} \leq \min_{\mathcal{E}} \max_{\Lambda_2, \ldots, \Lambda_l} \sum_{j=2}^l I(A_1 \ldots A_{j-1} : A_j | R)_\pi_j
\]

(48)

with

\[
\pi_j := (\text{id}^{A_1 \cdots A_{j-1}} \otimes \Lambda_j \otimes \text{id}^{A_{j+1} \cdots A_l} \otimes \mathcal{E}^{A_{l+1} \cdots A_k})(\rho^{A_1 \cdots A_k}).
\]

(49)

The last inequality in Eq. (48) follows by the monotonicity of the mutual information under local operations (Eq. (43) of Lemma 5). Then

\[
\min_{\mathcal{E}} \max_{\Lambda_2, \ldots, \Lambda_l} I(A_1 : \ldots : A_l | R)_\pi^{A_1 \cdots A_l R} \leq \min_{\mathcal{E}} \max_{\Lambda_2, \ldots, \Lambda_l} \sum_{j=2}^l I(A_1 \ldots A_{j-1} : A_j | R)_\pi_j = \min_{\mathcal{E}} \sum_{j=2}^l \max_{\Lambda_j} I(A_1 \ldots A_{j-1} : A_j | R)_\pi_j \leq \min_{\Lambda_l} [(l-1) \max_{\Lambda_l} I(A_1 \ldots A_{l-1} : A_l | R)]_{\pi_l},
\]

(50)

where the last inequality follows from the monotonicity of mutual information under tracing out and the permutation invariance of the state \( \rho^{A_1 \cdots A_k} \).

We claim that

\[
\min_{\mathcal{E}} \max_{\Lambda_l} I(A_1 \ldots A_{l-1} : A_l | R)_\pi_l \leq \frac{(l-1) \ln |A|}{k-l+1}.
\]

(51)
Indeed, defining $\nu^{A_1 \ldots A_k} := (\text{id}^{A_1 \ldots A_{l-1}} \otimes \Lambda_l \otimes \cdots \otimes \Lambda_k)(\rho^{A_1 \ldots A_k})$, for quantum-classical channels $\Lambda_j$, we have

$$\max_{\Lambda_{l+1}, \ldots, \Lambda_k} I(A_1 \ldots A_{l-1} : A_l \ldots A_k)_{\nu}$$

$$= \max_{\Lambda_{l+1}, \ldots, \Lambda_k} \sum_{j=l}^{k} I(A_1 \ldots A_{l-1} : A_j | A_{j+1} \ldots A_k)_{\nu}$$

$$= \max_{\Lambda_{l+1}, \ldots, \Lambda_k} \left( \sum_{j=l+1}^{k} I(A_1 \ldots A_{l-1} : A_j | A_{j+1} \ldots A_k)_{\nu} + \max_{\Lambda_l} I(A_1 \ldots A_{l-1} : A_l | A_{l+1} \ldots A_k)_{\nu} \right)$$

$$\geq \max_{\Lambda_{l+1}, \ldots, \Lambda_k} \left( \sum_{j=l+1}^{k} I(A_1 \ldots A_{l-1} : A_j | A_{j+1} \ldots A_k)_{\nu} + \min_{E} \max_{\Lambda_l} I(A_1 \ldots A_{l-1} : A_l | R)_{\pi_l} \right), \tag{52}$$

where the last inequality comes from replacing the specific measurement $\Lambda_{l+1} \otimes \cdots \otimes \Lambda_k$ with the minimum over all measurements $E$ on systems $A_{l+1} \ldots A_k$. Iterating the argument and exploiting permutation invariance we find

$$\max_{\Lambda_{l+1}, \ldots, \Lambda_k} I(A_1 \ldots A_{l-1} : A_l \ldots A_k)_{\nu} \geq (k - l + 1) \min_{E} \max_{\Lambda_l} I(A_1 \ldots A_{l-1} : A_l | R)_{\pi_l}, \tag{53}$$

and obtain Eq. (51) from the bound $(l - 1) \ln |A| \geq I(A_1 \ldots A_{l-1} : A_l \ldots A_k)_{\nu}$. Combining it with Eq. (50) we get

$$\min_{E} \max_{\Lambda_{2}, \ldots, \Lambda_l} I(A_1 : \ldots : A_l | R)_{\pi} \leq \frac{(l - 1)^2 \ln |A|}{k - l + 1}. \tag{54}$$

We now show how to combine this bound with a few properties of the measure $I(A_1 : \ldots : A_l | R)$ to complete the proof. We have

$$I(A_1 : \ldots : A_l | R)_{\pi} = \sum_{i} p_i I(A_1 : \ldots : A_l)_{\pi_i}, \tag{55}$$

with $\pi_i := (\text{id}^{A_1} \otimes \Lambda_2 \otimes \cdots \otimes \Lambda_l)(\rho_i)$, for an ensemble $\{p_i, \rho_i\}$ such that each $\rho_i \in \mathcal{D}(A^{\otimes l})$ is permutation-symmetric and $\sum_i p_i \rho_i = \rho^{A_1 : \ldots : A_l}$. Then, by Pinsker’s inequality (Eq. (44)) and the convexity of $x^2$:

$$\min_{E} \max_{\Lambda_{2}, \ldots, \Lambda_l} I(A_1 : \ldots : A_l | R)_{\pi} \geq \frac{1}{2} \left\| (\text{id}^{A_1} \otimes \Lambda_2 \otimes \cdots \otimes \Lambda_l) \left( \rho^{A_1 : \ldots : A_l} - \sum_i p_i \rho_i^{A_1} \otimes \cdots \otimes \rho_i^{A_l} \right) \right\|_1^2$$

Part 1 of the theorem follows from Eq. (54).
Part 2: Let \( p(x_1, \ldots, x_k|a_1, \ldots, a_k) \) be a permutation-symmetric non-signaling distribution and \( \mu = \mu_1 \times \cdots \times \mu_k \) a product distribution on \( A_1 \times \cdots \times A_k \). We will use the abbreviations \( X_{<l} := X_1 \ldots X_{l-1} \) and \( X_{>l} := X_{l+1} \ldots X_k \).

\[
\min_{a_1, \ldots, a_k} \mathbb{E}_{a_1, \ldots, a_k} I(X_1 : \cdots : X_l | X_{>l})_p
\]

\[
= \min_{a_1, \ldots, a_k} \mathbb{E}_{a_1, \ldots, a_k} \sum_{j=2}^{l} I(X_{<j} : X_j | X_{>l})_p
\]  

(56a)

\[
= \min_{a_1, \ldots, a_l} \sum_{j=2}^{l} \mathbb{E}_{a_1, \ldots, a_l} I(X_{<j} : X_j | X_{>l})_p
\]

(56b)

\[
\leq (l-1) \min_{a_{l+1}, \ldots, a_k} \mathbb{E}_{a_1, \ldots, a_l} I(X_{<l} : X_l | X_{>l})_p
\]  

(56c)

To derive the last inequality, observe that \( I(X_{<j} : X_j | X_{>l}) = I(X_{<j} : X_l | X_{>l}) \leq I(X_{<l} : X_l | X_{>l}) \), where the equality is from the symmetry of \( p \) and the inequality is from the monotonicity of mutual information under tracing out systems.

Next, \( (l-1) \ln |X| \geq \min_{a_1, \ldots, a_l} \sum_{j=l+1}^{k} \mathbb{E}_{a_1, \ldots, a_l} I(X_{<l} : X_j | X_{>l})_p \) \quad (57a)

\[
= \min_{a_1, \ldots, a_k} \left( \sum_{j=l+1}^{k} \mathbb{E}_{a_1, \ldots, a_{l-1}} I(X_{<l} : X_j | X_{>l})_p \right)
\]

(57b)

\[
+ \mathbb{E}_{a_1, \ldots, a_l} I(X_{<l} : X_l | X_{>l})_p
\]

\[
\geq \min_{a_1, \ldots, a_k} \sum_{j=l+1}^{k} \mathbb{E}_{a_1, \ldots, a_{l-1}} I(X_{<l} : X_j | X_{>l})_p
\]

(57c)

Iterating, we find that

\[
\min_{a_1, \ldots, a_k} \mathbb{E}_{a_1, \ldots, a_l} I(X_{<l} : X_l | X_{>l})_p \leq \frac{(l-1) \ln |X|}{k-l+1}
\]  

(58)

\[
\min_{a_1, \ldots, a_k} \mathbb{E}_{a_1, \ldots, a_l} I(X_1 : \cdots : X_l | X_{>l})_p \leq \frac{(l-1)^2 \ln |X|}{k-l+1} \quad \text{using (56)}
\]  

(59)

Fix \( a_{l+1}, \ldots, a_k \) achieving the minimum in Eq. (59). Using the non-signaling property, we can decompose

\[
p(X_{<l}|A_{<l}) = \sum_{X_{>l}} p(x_{>l}|a_{>l}) p(X_{<l}|A_{<l}, a_{>l}, x_{>l}).
\]  

(60)

The astute reader will realize that it is now time to deploy Pinsker’s inequality (Eq. (44)). Along with Eq. (59) and the convexity of \( x^2 \), this concludes the proof of the theorem. 

\( \square \)
5. Applications

Our de Finetti theorems provide approximation algorithms for separable states and LHV distributions. The common feature of these sets is that they are convex combinations of tensor products of simpler objects. These have many applications in quantum information and classical optimization, which we discuss in this section.

5.1. Non-local games: algorithms and hardness results. One application of Theorem 1 is to the computational complexity of non-local games. A multiprover, or non-local, game is played between a set of cooperative players/provers, who are not allowed to communicate with each other, and a referee/verifier who interrogates the provers to decide if they win the game. In a one-round game, for example, the verifier chooses questions to each prover at random (from a prior distribution) and checks the answers obtained from the provers in order to decide whether to accept the answer or not. Even though the provers cannot communicate with each other, they can agree on a common strategy in order to win the game with the maximum probability possible.

Multiprover games have had a central role in computational complexity theory. In a seminal paper Babai, Fortnow, and Lund proved $\text{NEXP} = \text{MIP}$ [9], with MIP the class of languages having multi-prover interactive proof with a polynomial number of provers, rounds, and bits exchanged between the provers and the verifier in each round. Building on [9], it was then proven in [6, 7] that it is NP-hard to approximate to constant error the maximum winning probability of a two-player one-round game (with the input size given by the total number of questions to the players and their answers). This hardness result is equivalent to the PCP theorem [6, 7], which has a pivotal role in hardness of approximation results (see e.g. [4]).

What if the players share correlations? Can this increase their winning probability? While it is easy to see that shared randomness is of no help, it has been known since the work of Bell [14] that entanglement can sometimes increase this winning probability. One can even consider stronger correlations than the ones allowed by quantum mechanics, such as arbitrary non-signaling correlations. Upper bounds on the maximum winning probability of a one-round game under classical strategies are known as Bell inequalities, and non-local strategies that beat these bounds are known as Bell inequality violations. Such violations of Bell inequalities are central in the foundations of quantum mechanics as they can provide experimental verification to show that nature cannot be described by a local hidden variable theory [8].

We consider here the complexity of computing the entangled value of the game $\omega_e$, defined as the maximum probability of winning the game using entanglement, or the non-signaling value of the game $\omega_{ns}$, defined as the optimal probability under non-signaling strategies. By contrast, the maximum winning probability under classical strategies is called the classical value of the game $\omega_c$.

Although a priori computing the entangled value of a game requires optimizing over a large set, in some cases this can be easier. Indeed, for unique games, the best known algorithms for the classical value [5] run in time $\exp(n^\varepsilon)$ (with $0 < \varepsilon < 1$ depending on the desired degree of approximation), whereas the entangled value of the game can be approximated within a constant factor (technically it is $1 - \omega_e$ that is approximated) in polynomial time using semidefinite programming [54] (or exactly calculated for the special case of XOR games [33]). These two classes of games could be taken as evidence that the estimation of the entangled value is generally easier than of the classical value. However, if one is interested in a high-accuracy estimation this turns
out not to be true. Kempe, Kobayashi, Matsumoto, Toner, and Vidick proved that it is \( \text{NP} \)-hard to approximate to an inverse polynomial (in the size of the game) the entangled value of one-round 3-prover games [53] (see also [32, 49, 50]). Recently in a beautiful development Ito and Vidick [51] proved that it is \( \text{NP} \)-hard, under quasi-polynomial reductions (improved to a polynomial-time reduction in Ref. [81]), to approximate the entangled value of 3-prover games with polynomially many rounds even to constant error. The result [51] has a more elegant formulation in terms of interactive proof systems: It shows that \( \text{NEXP} \subseteq \text{MIP}^* \), with \( \text{MIP}^* \) the analogue of \( \text{MIP} \) in which the provers share entanglement [57]. The maximum probability of non-signaling strategies, in turn, can always be computed efficiently by linear programming [48].

Probably the biggest open question in this area is to determine the computational complexity of approximating the entangled value of two-player one-round games to constant accuracy, since recent work of Vidick [81] has now resolved this for three or more players. There are two reasons why this is a particularly interesting setting. The first is the fact that the PCP theorem can be stated as the \( \text{NP} \)-hardness of approximating the classical value of one-round games to constant accuracy. Thus an analogous result for the entangled value could be interpreted as a version of the PCP theorem in the presence of entanglement. One of our goals here is to propose a new approach to address this problem.

A particular class of games that we will consider are the so-called free games, defined as games in which the questions to each of the players are chosen independently from the questions to the other players [26]. A famous example from physics is the CHSH game. The fact that the verifier cannot coordinate questions suggests that the computation of the maximum winning probability of such games should not be as hard as for general games. And indeed Bellare, Feige and Killian proved that the analogue of \( \text{MIP} \) for poly-round free games is equal to \( \text{PSPACE} \) [15], while Aaronson, Impagliazzo and Moshkovitz [3] proved that the classical value of one-round free games with questions to the two provers in \( Q \times Q \) and answers in \( A_1 \times A_2 \) can be simulated to within error \( \epsilon \) by AM (Arthur-Merlin) proofs with an \( O(\log |Q|+\log(|A_1|\cdot|A_2|)/\epsilon) \)-bit message from Arthur to Merlin and an \( O(\log |A_1|\log |A_2|/\epsilon^2) \)-bit message from Merlin to Arthur. As a result, the value of such games can be estimated in time \( \text{poly}(\log |Q|) \cdot \exp(\log |A_1|\log |A_2|/\epsilon^2) \). They also gave a matching hardness of approximation result for free games, showing that one can reduce 3-SAT on \( n \) binary variables to computing the classical value of the game \( \omega_c(G) \) to within constant additive error for 2-player one-round free games with \( \exp(O(\sqrt{n})) \)-sized answer alphabet.\(^8\)

As a corollary of Theorem 1 we will prove that the classical value of free games can be computed efficiently by linear programming, matching the run-time of the algorithm of [3]. Moreover, we will also derive a non-trivial hardness of approximation result for the entangled value of free games by importing to the case of entangled strategies the hardness of approximation result for the classical value of free games from [3]. Finally we will show how a conjectured strengthening of Theorem 1 would yield an alternate proof of the \( \text{NP} \)-hardness of obtaining a constant error approximation to the entangled value of the game \( \omega_e \) for four-player one-round games.

Before we turn to the precise statement of the main result of this section let us give a more formal definition of non-local games.

**Definition 6.** We define a \( m \)-prover game \( G(m, \pi, V) \) by two parameters \( \pi \) and \( V \):

---

\(^8\) There are suggestive similarities between this result and results about \( \text{QMA}(2) \) and variants thereof; see Sect. 5.2 and [42].
1. \( \pi \) is a probability distribution on \( Q_1 \times \ldots \times Q_m \) for finite sets \( Q_1, \ldots, Q_m \).
2. \( V \) is a predicate on \( Q_1 \times \ldots \times Q_m \times A_1 \times \ldots \times A_m \) for finite sets \( A_1, \ldots, A_m \).

The sets \( Q_i \) and \( A_i \) consist of the possible questions and answers, respectively, for player \( i \). The function \( 0 \leq V(a_1, \ldots, a_m | q_1, \ldots, q_m) \leq 1 \) is the payoff of the answer \( (a_1, \ldots, a_m) \) given the question \( (q_1, \ldots, q_m) \).

The classical value of the game \( G \) is given by
\[
\omega_c(G(m, \pi, V)) := \max_{a_1, \ldots, a_m} \sum_{q_1, \ldots, q_m} \pi(q_1, \ldots, q_m) V(a_1(q_1), \ldots, a_m(q_m)),
\]
where the maximum is over all functions \( a_j : Q_j \to A_j \).

The entangled value of the game, in turn, is given by
\[
\omega_e(G(m, \pi, V)) := \sup_{\psi} \sum_{q_1, \ldots, q_m} \pi(q_1, \ldots, q_m) \sum_{a_1, \ldots, a_m} V(a_1, \ldots, a_m | q_1, \ldots, q_m) \langle \psi | M^1_{a_1 | q_1} \otimes \ldots \otimes M^m_{a_m | q_m} | \psi \rangle,
\]
where the supremum is over states \( |\psi\rangle \) of arbitrary dimension and arbitrary POVMs
\[
\{M^1_{a_1 | q_1} | a_1 \in A_1, \ldots, \{M^m_{a_m | q_m} | a_m \in A_m, \}
\]
with \( \sum_{a_k \in A_k} M^k_{a_k | q_k} = I \) for every \( q_k \in Q_k \) and \( k \in [m] \).

Finally, the non-signaling value of the game \( G \) is defined as
\[
\omega_{ns}(G(m, \pi, V)) := \max_{\rho} \sum_{q_1, \ldots, q_m} \pi(q_1, \ldots, q_m) \sum_{a_1, \ldots, a_m} V(a_1, \ldots, a_m | q_1, \ldots, q_m) \rho(a_1, \ldots, a_m | q_1, \ldots, q_m),
\]
where the maximum is over all non-signaling probability distributions \( \rho(a_1, \ldots, a_m | q_1, \ldots, q_m) \).

Corollary 7.

1. Let \( G(2, \pi, V) \) be a two-player one-round non-local free game with \( \pi \) a product probability distribution on \( R \times Q \) and \( V \) a predicate on \( R \times Q \times A \times B \). Then there is a \((m + 1)\)-player one-round non-local game \( \widetilde{G}(m + 1, \pi, \widetilde{V}) \) with \( \pi \) a probability distribution on \( R \times Q_1 \times \ldots \times Q_m \), with \( |Q_k| = |Q| \) for \( k \in [m] \), and \( \widetilde{V} \) a predicate on \( R \times Q_1 \times \ldots \times Q_m \times A \times B_1 \times \ldots \times B_m \), with \( |B_k| = |B| \) for \( k \in [m] \), such that
\[
\omega_c(G) = \omega_e(\widetilde{G}) \leq \omega_{ns}(\widetilde{G}) \leq \omega_c(G) + \sqrt{\frac{\ln |A|}{2m}}.
\]

2. For a free game \( G(2, \pi, V) \) there is a linear-programming relaxation of size \( \ln |A| \left( |Q| \left| B \right| \right)^{\frac{1}{2} \omega} \) for computing \( \omega_c(G) \) to within additive error \( \epsilon \).
3. One can reduce 3-SAT on n variables to computing $\omega_e(G)$ to within constant additive error for $O(\sqrt{n})$-player one-round non-local games with answer alphabet size of $\exp(O(\sqrt{n}))$ in which only two players are asked questions. The reduction runs in time polynomial in the output size, namely $\exp(O(\sqrt{n}))$.

We note that it is trivial to prove either a version of part 3 of Corollary 7 in which the answer alphabet size is $2^n$ (in which case even one prover is clearly enough), or one in which the answer alphabet size is constant but one has $n$ provers, or one with $\sqrt{n}$ provers and alphabet size $2\sqrt{n}$ in which all provers respond. However, in our result, the total number of bits sent is $O(\sqrt{n})$.

Part 2 of Corollary 7 follows directly from part 1 and the fact that $\omega_{ns}$ can be computed by linear programming. This gives a new algorithm matching the performance of the algorithm due to Aaronson, Impagliazzo and Moshkovitz [3]. Part 3 of Corollary 7 follows from part 1 and the hardness of approximation result of Ref. [3] for free games.

Part 1 in turn gives a generic relation between the classical value of a free game, on one hand, and the quantum and non-signaling values of a modified game with more players, on the other hand. The idea of adding more players is to try to immunize the original game from entanglement (or general non-signaling correlations) by adding extra consistency tests that forces the entanglement between the players to have a specific form. Indeed the new game with $m+1$ players consists of playing the original game with player one and one of the remaining $m$ players chosen at random. This essentially allows us to consider a two-player game where the provers can only share an $m$-extendible state (or $m$-extendible non-signaling conditional distribution). Then by Theorem 1 we obtain that this $m$-extendible state cannot be much better than a separable state or a local hidden variable distribution, which themselves are no better than just having shared randomness. The crucial aspect of Theorem 1 used here is that the error term only depends on the number of outcomes (which is given by the number of possible answers of the non-local game in question), and not on the dimension of the entangled state or on the number of different POVMs in the family in the quantum case (or the number of measurement settings in the non-signaling case). The idea of immunizing entanglement by introducing more players is not new and was used before by Kempe et al [53] to prove the hardness of estimating the entangled value within error inverse polynomial in the size of the game.

More generally, it was observed by Terhal, Doherty, and Schwab [80] that $m$-extendible states cannot violate any Bell inequality with fewer than $m$ measurements for Bob (and an arbitrary number of measurements for Alice). In contrast Theorem 1 shows that a non-signaling $m$-extendible conditional distribution can violate a Bell inequality associated to a free game (an example of which is the CHSH inequality) with an arbitrary number of measurements, each with $M$ possible outcomes, by at most $\epsilon$-close to a separable state in trace norm (thus having similar statistics under general quantum measurements) one must consider $m$-extendible states with $m = \Omega(|B|/\epsilon)$, with $|B|$ the dimension of the $B$ subsystem [30]. The monogamy of non-locality we find here, in comparison, has a bound that is independent of the dimension of the state.

Finally let us mention a conjecture whose validity would imply the NP-hardness of estimating $\omega_e$ to within constant error for 4-player one-round games. The conjecture is the following strengthening of Theorem 1.
Conjecture 8. Let $\rho^{AB} \in \mathcal{D}(A \otimes B)$ be a $k$-extendible state and $\mu(m)$ a distribution over quantum operations $\{\mathcal{E}_m^{A \rightarrow \tilde{A}}\}_m$, with $\mathcal{E}_m^{A \rightarrow \tilde{A}} : \mathcal{D}(A) \rightarrow \mathcal{D}(\tilde{A})$. Then

$$\min_{\sigma \in \text{Sep}(A:B)} \mathbb{E}_{m \sim \mu} \max_{\Lambda^B \in \mathcal{M}} \left\| \mathcal{E}_m^{A \rightarrow \tilde{A}} \otimes \Lambda^B \left( \rho^{AB} - \sigma^{AB} \right) \right\|_1 \leq \sqrt{\frac{2 \ln |\tilde{A}|}{k}}. \quad (66)$$

The difference with Theorem 1 is that the order of the expectation over $\mu$ and the maximization over measurements $\Lambda^B$ is reversed. It is easy to check that one would be able to carry through the proof of part 1 of Corollary 7 for general games (of course only for the relation of $\omega_e$ and $\omega_c$). The fact that we would be able to prove NP-hardness for 4-player games would then follow from the combination of this stronger version of Eq. (65) with a recent version of the PCP theorem due to Khot and Safra, in the language of two-prover one-round games [55].

We have written (66) in a way that is meant to parallel (2) from Theorem 1, with a consequence that systems $A$ and $B$ are treated very differently. However, the conjecture could equivalently be restated in a more symmetric form. If we explicitly include the maximization over $\mu$, then the LHS (66) becomes $\sup_{\mu} \min_{\sigma} [\mathbb{E}_{m \sim \mu} \max_{\Lambda^B} \| \cdots \|_1]$. Observe that the term inside the $[\cdots]$ is linear in $\mu$ and convex in $\sigma$; indeed, it is a seminorm of $\sigma$. Thus, we can use Sion’s minimax theorem [78] and reverse the order of the $\sup_{\mu}$ and $\min_{\sigma}$. At this point the $\sup_{\mu} \mathbb{E}_{m \sim \mu}$ become superfluous, and we can replace the pair with simply a maximum over maps $\mathcal{E}^{A \rightarrow \tilde{A}}$. Thus, Conjecture 8 could equivalently be stated as

$$\min_{\sigma \in \text{Sep}(A:B)} \max_{\mathcal{E}^{A \rightarrow \tilde{A}} } \max_{\Lambda^B \in \mathcal{M}} \left\| \mathcal{E}^{A \rightarrow \tilde{A}} \otimes \Lambda^B \left( \rho^{AB} - \sigma^{AB} \right) \right\|_1 \leq \sqrt{\frac{2 \ln |\tilde{A}|}{k}}. \quad (67)$$

Although the conjecture is consistent with all the examples of states we are aware of, we note that a proof would have to follow a very different approach to the one used in Theorem 1, as it cannot apply to non-signaling distributions. The reason is that the quantum version of the conjecture would imply that $\text{NEXP} \subset \text{MIP}^e(4, 1)$, and the no-signaling version would imply that $\text{NEXP} \subset \text{MIP}^{ns}(4, 1)$, but this latter class is contained in $\text{EXP}$. A scaled down version of this argument shows that the no-signaling version of Conjecture 8 would imply that $P = \text{NP}$. It is an interesting open question to find a more direct counter-argument, such as an example of a $k$-extendable no-signaling distribution whose difference from LHV distributions can be detected by correlated measurements.

Thus, despite the superficial similarity of (the quantum version of) Conjecture 8 with our Theorem 1, any proof will need to find features of quantum states that are not shared by no-signaling distributions. In this respect the hypothesis-testing approach of Refs. [20, 64] might be a promising route.

5.1.1. Proof of Corollary 7. Let us turn to the proof of the corollary. The first lemma is an adaptation of a similar result of Kempe et al. [53]. It shows that by symmetrizing the questions and answers of a subset $S$ of the players one can without loss of generality

9 The proof that $\text{MIP}^{ns}(\text{poly, poly}) \subseteq \text{EXP}$ follows by linear programming: the no-signaling constraints are linear constraints on an exponential-sized prover strategy. Ref. [53] attributes it to a personal communication from Daniel Preda, and Ref. [48] builds on this approach to show that $\text{MIP}^{ns}(2, 1) \subseteq \text{PSPACE}$ by finding a way to parallelize the LP in the 2-prove 1-round case.
Lemma 9. Let $G(N, \pi, V)$ be a non-signaling-prover game such that $\pi(i_1, \ldots, i_N)$ is symmetric in $i_1, \ldots, i_m$ and $V$ is symmetric under simultaneous permutation of registers $1, \ldots, m$ of the questions $q_i_1, \ldots, q_i_N$ and of the answers $a_i_1, \ldots, a_i_N$ for $m \leq N$. Then given any strategy given by a non-signaling strategy that wins with probability $p$, there exists a symmetric strategy with respect to provers $1, \ldots, m$, meaning that it is symmetric under simultaneous permutation of questions and answers.

The next lemma gives a hardness of approximation result for approximating the classical value of free games.

Lemma 10. (Aaronson-Impagliazzo-Moshkovitz [3]). 3-SAT with $n$ variables can be reduced to the problem of obtaining a constant error approximation to $\omega_c(G)$ for two-player one-round free games with $2^{O(\sqrt{n})}$-sized output alphabet.

Corollary 7.

Part 1: Define a game $\overline{G}$ in which the verifier chooses a pair $(r, q)$ from the distribution $\pi(r, q)$ and sends $r$ to the first prover (let us call it Alice) and the $q$ to one of the other $m$ provers chosen at random (let us call them Bob $1$ to Bob $m$). The verifier does not send a question and does not expect an answer from the remaining Bobs. Then the verifier uses the answers obtained from Alice and the chosen Bob to compute $V(a, b|r, q)$. Applying Lemma 9 to the case of non-signaling games we can restrict the parties to use non-signaling distributions which are symmetric on the Bobs. Thus

$$
\omega_{ns}(\overline{G}) = \sup_p \sum_{q, r} \pi(r, q) \sum_{a, b_1, \ldots, b_m} \left( \frac{1}{m} \sum_{k=1}^m V(a, b_k|r, q_k) \right) \\
\times p(a, b_1, \ldots, b_m|r, q_1, \ldots, q_m)
$$

$$
= \sup_{p \in m-\Ext} \sum_{q, r} \pi(r, q) \sum_{a, b} V(a, b|r, q) p(a, b|r, q),
$$

(68)

where the supremum in the last line is taken over all $m$-extendible non-signaling distributions $p$. Then by Theorem 1

$$
\sup_{p \in m-\Ext} \sum_{q, r} \pi(r, q) \sum_{a, b} V(a, b|r, q) p(a, b|r, q) \\
\leq \sup_{s \in \LHV} \sum_{q, r} \pi(r, q) \sum_{a, b} V(a, b|r, q) s(a, b|r, q) + \frac{1}{2} \sqrt{\frac{2 \ln |A|}{m}}.
$$

(69)

In more detail, since the game is free we have that $\pi(r, q) = \pi_1(r)\pi_2(q)$. Then

$$
\left| \sum_{q, r} \pi_1(r)\pi_2(q) \sum_{a, b} V(a, b|r, q) (p(a, b|r, q) - s(a, b|r, q)) \right| \\
\leq \mathbb{E}_{\pi_1(r)} \mathbb{E}_{\pi_2(q)} \| p(a, b|r, q) - s(a, b|r, q) \|_1 \\
\leq \mathbb{E}_{\pi_1(r)} \max_{q \in Q} \| p(a, b|r, q) - s(a, b|r, q) \|_1.
$$

(70)
From theorem 1

\[
\min_{s \in \text{LHV}} \mathbb{E}_{\pi_1(r)} \max_{q \in Q} \| p(a, b|r, q) - s(a, b|r, q) \|_1 \leq \frac{1}{2} \sqrt{\frac{2 \ln |A|}{m}}. \tag{71}
\]

**Part 2:** Follows from part 1 and the fact that \( \omega_{\text{ns}} \) can be computed by a linear program [48].

**Part 3:** Follows from part 1 of this Lemma and part 1 of Lemma 10. \( \square \)

5.2. Optimal of Chen and Drucker’s multiple-proof protocol for 3-SAT. One first application of Theorem 2 is to proof systems with multiple unentangled provers.

Given a 3-SAT formula with \( n \) variables and \( O(n) \) clauses, what is the minimum proof that can convince a verifier the formula is satisfiable? Under the exponential time hypothesis [45] – which says 3-SAT cannot be solved in subexponential time – \( \Omega(n) \) bits are required, i.e. it is believed one cannot do anything substantially better than just write down a \( n \)-bit satisfying assignment. What if we can send a quantum state as a proof to a verifier who has a quantum computer to check its validity? Perhaps we could pack more information into the quantum state so that \( o(n) \) qubits would be enough to convince the verifier? It turns out that assuming a quantum version of the exponential time hypothesis – namely that to solve 3-SAT takes exponential time even on a quantum computer (see e.g. [16] for the oracle version of this claim) – \( \Omega(n) \) qubits are required [65].

Quantum mechanics allows us to add a new twist to this question. What if we want to convince a quantum verifier by sending a quantum state to her, but with the promise that parts of the quantum state are not entangled with each other? In this case the argument of Ref. [65] does not apply anymore and at least we do not have any implausible consequence for having a sublinear proof. And indeed Aaronson, Beigi, Drucker, Fefferman, and Shor [2] (building on [18]) proved that \( \sqrt{n} \) polylog \( n \) unentangled quantum states, each of \( \log(n) \) qubits, are enough to convince a quantum verifier that a 3-SAT instance with \( n \) variables and \( O(n) \) clauses is satisfiable.

The result of [2] was strengthened in two directions: First Harrow and Montanaro [41] proved that two unentangled proofs, each of \( \sqrt{n} \) polylog \( n \) qubits, are sufficient. Second Chen and Drucker [28] showed that \( \sqrt{n} \) polylog \( n \) identical unentangled quantum proofs of \( O(\log(n)) \) qubits each are sufficient to convince even a verifier who measures each of the proofs separately and postprocesses the classical outcomes in order to decide whether to accept or not.

To state the main result of this section we define a few quantum complexity classes (see Sect. 5.2.1 for formal definitions). The first is a natural quantum analogue of \( \text{NP} \) (more precisely of \( \text{MA} \)). Let \( \text{QMA}_n(c, s) \) be the class of problems such that: (i) for "yes" instances there is a quantum proof composed of \( n \) qubits that makes the verifier, who has access to polynomial quantum computation, to accept with probability at least \( c \); and (ii) for "no" instances every proof is accepted with probability at most \( c \). Let \( \text{QMA}_n(m, c, s) \) be the analogue of \( \text{QMA} \) in which instead of one quantum proof the verifier receives \( m \) quantum proofs, each of \( n \) qubits, with the promise that they are not entangled with each other [58].

Further let \( \text{BellQMA}_n(m, c, s) \) be an analogue of \( \text{QMA}_n(m, c, s) \) in which the verification procedure is restricted to applying independent measurements to each of the \( m \) proofs and then post-processing the outcomes classically [2]. The name of the class
comes from the fact that the verifier is basically constrained to apply a Bell test as his verification procedure. Finally let BellSymQMA_n(m, c, s) be the analogue of BellQMA_n(m, c, s) in which all the m proofs are promised to be identical (possibly mixed) states; this corresponds to the set SepSym from [42]. See Sect. 5.2.1 for formal definitions of these classes.

With this notation the Chen-Drucker result can be stated as showing the containment of 3-SAT with n variables and \( O(n) \) clauses in BellSymQMA_{\log(n)}(\sqrt{n} \text{ polylog}(n), 1 - 2^{-O(\sqrt{n})}, 1/\text{poly}(n)) [28]. (A similar, but incomparable, result holds for BellQMA also follows from [28].) A corollary of Theorem 2 is that this is essentially optimal, i.e. the square-root improvement found for the total proof size is all there is if we restrict ourselves to BellSymQMA protocols. It is an open question whether an analogous optimal result holds for BellQMA.

**Corollary 11.**

1. \( \text{BellSymQMA}_n(m, c, s) \subseteq \text{QMA}_{10n^2m^2/\epsilon^2}(c, s + \epsilon) \).
2. For every \( \epsilon > 0 \) and \( c - s = \Omega(1) \), there is no BellSymQMA_{\Omega(\log(n))}(n^{1/2-\epsilon}, c, s) protocol for 3-SAT with \( n \) variables and \( O(n) \) clauses, unless 3-SAT can be solved in \( \exp(n^{1-2\epsilon} \text{ polylog}(n)) \) time.
3. \( \text{BellQMA}_n(m, c, s) \subseteq \text{QMA}_{10n^2m^4/\epsilon^2}(c, s + \epsilon) \).
4. \( \text{QMA}_{\text{poly}(n)}(\frac{2}{3}, \frac{1}{3}) = \text{BellQMA}_{\text{poly}(n)}(\text{poly}(n), \frac{2}{3}, \frac{1}{3}) \).

See Sect. 5.2.1 for the proof.

In [19, 20] it was shown that BellQMA(m) is contained in QMA for a constant number of provers \( m \). Corollary 11 strengthens the containment to even to a polynomial number of provers. This gives a new characterization of the class QMA and shows that the only advantage (in the regime where \( c - s \geq 1/\text{poly}(n) \)) that BellQMA protocols can offer is a polynomial reduction in the proof size, such as in the protocol of [28].

**Remark.** In fact we can prove something slightly stronger than Corollary 11. Instead of Bell measurements, where \( k \) parties individually measure their systems and send the results to a referee, we can handle a slightly larger class of measurements. Our proofs apply equally to the setting where \( k - 1 \) parties measure their systems and send classical messages to the last party, who can choose a measurement adaptively based on these messages. This will follow from the fact that in part 1 of Theorem 2, we can leave one subsystem unmeasured. To keep the exposition simple, we will not formally state this improved version of Corollary 11.

5.2.1. Proof of Corollary 11. Let us turn to the proof of Corollary 11.

We begin with a definition of analogues of QMA with multiple unentangled proofs.

**Definition 12.** A language \( L \) is in M-QMA_n(m, s, c) if there exists a polynomial-time implementable two-outcome measurement \( \{M_x, I - M_x\} \) from the class \( M \) such that

1. Completeness: If \( x \in L \), there exist \( m \) proofs \( \sigma_1, \ldots, \sigma_m \), each of \( n \) qubits, such that
   \[
   \text{tr} (M_x (\sigma_1 \otimes \cdots \otimes \sigma_m)) \geq c. \tag{72}
   \]
2. Soundness: If \( x \notin L \), then for any states \( \sigma_1, \ldots, \sigma_m \),
   \[
   \text{tr} (M_x (\sigma_1 \otimes \cdots \otimes \sigma_m)) \leq s. \tag{73}
   \]
If $M$ is the class of all polynomial-time implementable two-outcome measurements we denote the complexity class simply by $\text{QMA}_n(m, s, c)$.

Some examples of classes of measurements that we consider in this paper are:

1. **Bell** is composed of measurements $0 \leq M \leq I$ of the form
   \[
   M = \sum_{(i_1, \ldots, i_m) \in S} M_{1,i_1} \otimes \ldots \otimes M_{m,i_m}
   \]  
   (74)
   where $\sum_i M_{j,i} = I$ for all $j \in [m]$, and $S$ is a set of $m$-tuples of indices. In words the $m$ subsystems are measured locally giving outcomes $(i_1, \ldots, i_m)$ and the verifier accepts if $(i_1, \ldots, i_m) \in S$.

2. **LOCC** is composed of measurements of the form
   \[
   M = \sum_i M_{1,i} \otimes \ldots \otimes M_{m,i}
   \]  
   (75)
   such that $0 \leq M_{1,i} \leq I$ for all $i$, and $0 \leq \sum_i M_{k,i} \leq I$ for every $k \in \{2, \ldots, m\}$.

3. **SEP** is composed of measurements $0 \leq M \leq I$ such that
   \[
   M = \sum_i M_{1,i} \otimes \ldots \otimes M_{m,i},
   \]  
   (76)
   for positive semi-definite matrices $M_{j,i}$.

See [42] for more examples of classes of measurements as well as relations between them.

We will also make use of $\text{QMA}$ with multiple identical proofs:

**Definition 13.** A language $L$ is in $\text{M-SymQMA}_n(m, s, c)$ if there exists a polynomial-time implementable two-outcome measurement $\{M_x, I - M_x\}$ from the class $M$ such that

1. Completeness: If $x \in L$, there exists a proof $\sigma$ of $n$ qubits such that
   \[
   \text{tr} \left( M_x \sigma^{\otimes m} \right) \geq c.
   \]  
   (77)

2. Soundness: If $x \notin L$, then for any state $\sigma$,
   \[
   \text{tr} \left( M_x \sigma^{\otimes m} \right) \leq s.
   \]  
   (78)

**Proof of Corollary 11.**

*Part 1:* To simulate a $\text{BellSymQMA}_n(m, c, s)$ protocol in $\text{QMA}_{10n^2m^2/\epsilon^2}(c, s + \epsilon)$ the verifier receives the proof of $10n^2m^2/\epsilon^2$ qubits from the prover and consider it as $10nm^2/\epsilon^2$ blocks of $n$ qubits. Then he symmetrizes all the blocks, traces out all of them except the first $m$ blocks and runs the original $\text{BellQMA}$ protocol on them. It is clear that completeness is not changed. To analyze soundness we use part 1 of Theorem 2.

*Part 2:* Follows easily from the previous part.

*Part 3:* To simulate a $\text{BellQMA}_n(m, c, s)$ protocol in $\text{QMA}_{10n^2m^4/\epsilon^2}(c, s + \epsilon)$ the verifier receives the proof of $10n^2m^4/\epsilon^2$ qubits from the prover and consider it as $10nm^3/\epsilon^2$ blocks of $nm$ qubits. Then he symmetrizes all the blocks, traces out all of them except the first $m$ blocks. Then he divides each of these $m$ blocks into $m$ sub-blocks of $n$ qubits. Let us denote the $i$-th sub-block of the $j$-th block by $X_{i,j}$. Then the verifier runs the original $\text{BellQMA}$ protocol using the state in subsystems $X_{1,1}, X_{2,2}, \ldots, X_{m,m}$ as a proof. It is clear that completeness is not changed. To analyze soundness we use part 1 of Theorem 2.

*Part 4:* Follows easily from the previous part. □
5.3. Polynomial optimization and sum-of-squares proofs. Another application of our main theorems is to classical algorithms for maximizing polynomials over $\mathbb{C}^n$. The concepts of $k$-extendable and separable states turn out to correspond naturally to semidefinite programming (SDP) hierarchies for polynomial optimization, and thus we are able to prove convergence of these hierarchies for polynomials that correspond to LOCC measurements. This connection was first established by Doherty, Parrilo and Spedalieri [38], and was more recently made quantitative for general polynomials over the unit sphere in $\mathbb{R}^n$ by Doherty and Wehner [Personal communication (2009), [39]].

In this section, we consider the problem of maximizing real-valued polynomial functions over the complex unit sphere $S^{2n-1} \subseteq \mathbb{C}^n$. More precisely, we consider polynomials of $z_1, \ldots, z_n, \bar{z}_1, \ldots, \bar{z}_n$ that are bihomogenous of degree $d$, $d$, i.e. homogenous of degree $d$ in the $z_1, \ldots, z_n$ and homogenous of degree $d$ in the $\bar{z}_1, \ldots, \bar{z}_n$. This problem is closely related [35] to optimization over the real unit sphere, though not always identical [34]. For constant $d \geq 3$, maximizing a homogenous polynomial of degree $d$ over the unit sphere in $\mathbb{R}^n$ is NP-hard; see [36]. A promising general-purpose approximation scheme is to use an SDP hierarchy invented independently by Parrilo [71] and Lasserre [61] known as the sum-of-squares (SOS) hierarchy; see also [70] for a recent review of the complexity-theoretic properties of the SOS hierarchy. To define the SOS hierarchy, we introduce some notation. Let $\mathbb{C}[z, \bar{z}] := \mathbb{C}[z_1, \ldots, z_n, \bar{z}_1, \ldots, \bar{z}_n]$ denote complex polynomials in $n$ variables, let $\mathbb{C}[z, \bar{z}]_{d,d}$ denote the set of bihomogenous polynomials of degree $d$, $d$, and let $\mathbb{C}[z, \bar{z}]_{d,d}^k$ denote the set of Hermitian linear functionals from $\mathbb{C}[z, \bar{z}]_{d,d}$ to $\mathbb{R}$. Here we will consider only Hermitian linear functionals $L$, meaning that $L[\prod_{j=1}^n z_j^{a_j} \bar{z}_j^{b_j}] = L[\prod_{j=1}^n z_j^{a_j} \bar{z}_j^{b_j}]$ for any $a_1, \ldots, a_n, b_1, \ldots, b_n$.

If $p(z) \in \mathbb{C}[z, \bar{z}]_{d,d}$ and $k \geq d$, then we can upper bound $\max_{z \in S^{2n-1}} p(z)$ with the following SDP:

\begin{align}
\text{max } L(p) \text{ such that } & \quad (79a) \\
L \in \mathbb{C}[z, \bar{z}]_{k,k}^* & \quad (79b) \\
L(1) = 1 & \quad (79c) \\
L(q\bar{q}) \geq 0 & \quad \forall q \in \mathbb{C}[z, \bar{z}]_{k,0} \quad (79d) \\
L((z_1\bar{z}_1 + \ldots + z_n\bar{z}_n)q) = L(q) & \quad \forall q \in \mathbb{C}[z, \bar{z}]_{k-1,k-1} \quad (79e)
\end{align}

Here (79c) and (79d) are constraints that any collection of moments of a random variable should satisfy (with (79b) enforcing linearity), while (79e) expresses the $\sum_{i=1}^n |z_i|^2 = 1$ constraint (and can in general be replaced with any polynomial constraint; see [61,70,71]). To see that (79) is an SDP, observe that (79e) is a linear constraint and (79d) is equivalent to the constraint that the moment matrix $M(L) \geq 0$, where the entries of $M(L)$ are indexed by monomials in $\mathbb{C}[z, \bar{z}]_k$ and are defined by $M(L)_{\alpha,\beta} := L(z_\alpha \bar{z}_\beta)$.

We can interpret this SDP as replacing the maximum over $S^{2n-1}$ by a maximum over probability distributions over $S^{2n-1}$ (which of course changes nothing), and in turn approximating this by considering only the moments of order $\leq k$. The dual of (79) is

\begin{align}
\min \lambda \text{ such that } & \quad (80a) \\
\lambda - p = \left(\sum_{i=1}^n z_i\bar{z}_i - 1\right) & \quad (80b) \\
q_0 \in \mathbb{C}[z, \bar{z}]_{k-1,k-1} & \quad (80c)
\end{align}
\[ q_1, \ldots, q_m \in \mathbb{C}[\mathbf{z}, \bar{\mathbf{z}}]_{k,0} \]

This can be thought of as “proving” that \( p(x) \leq \lambda \) by using the fact that \( p(x) - \lambda \) is a sum of squares of polynomials; hence the “sum-of-squares” name for the hierarchy. More detail on this hierarchy, including proofs that these SDPs are bounded and satisfy strong duality, can be found in Section 6 of [62].

The parameter \( k \) is called the “level” of the SOS relaxation. Under reasonable assumptions, this SDP converges to \( \max_{z \in S^{2n-1}} p(z) \) as \( k \) grows [61, 71]. However, since the effort to compute (79) or (80) grows exponentially with \( k \), optimizations over the simplex, but less is known for the sphere [36].

As a result, we immediately obtain a bound on the ability of the sum-of-squares hierarchy to approximate certain polynomials over the complex hypersphere.

For matrices \( x, y \) we define \( \langle x, y \rangle := \text{tr} x \dagger y \). Let \( M \) be a one-way LOCC operator of the form

\[ M = \sum_{i_2, \ldots, i_l} P_{i_2, \ldots, i_l} \otimes Q_{i_2} \otimes \cdots \otimes Q_{i_l, i_l}, \]

with \( 0 \leq P_{i_2, \ldots, i_l} \leq I \) for each \( i_2, \ldots, i_l \) and \( 0 \leq \sum_{i_j} Q_{i_j, i_j} \leq I \) for each \( 2 \leq j \leq l \). Define \( \text{Sep}(A^{\otimes l}) \) to be the convex hull of \( \rho_1 \otimes \cdots \otimes \rho_l \) over \( \rho_1, \ldots, \rho_l \in \mathcal{D}(A) \). A variant of Theorem 2 implies that the SOS hierarchy can give a good approximation of

\[ h_{\text{Sep}(A^{\otimes l})}(M) \]

The quantity in (83) can be thought of as the solution to a polynomial optimization problem, or more precisely, a multiquadratic optimization problem. Given \( |z^{(1)}\rangle, \ldots, |z^{(l)}\rangle \in \mathbb{C}^n \), define

\[ p(z^{(1)}, \ldots, z^{(l)}) := \langle z^{(1)} | \otimes \cdots \otimes \langle z^{(l)} | M | z^{(1)} \rangle \otimes \cdots \otimes | z^{(l)} \rangle. \]

Then (83) is equal to the maximum of (84) over all collections of unit vectors \( z^{(1)}, \ldots, z^{(l)} \).

There is an alternate interpretation in terms of polynomial optimization, albeit of some not-very-natural polynomials. Define \( z \in \mathbb{C}^{nl} \) to be the vector obtained by concatenating \( z^{(1)}, \ldots, z^{(l)} \), so that \( p(z) \) is a degree-\( l \) polynomial in \( z \). Maximizing \( p(z) \) over unit vectors \( z \) can be seen (using a convexity argument) to yield (83) times the normalization factor \( l^{-l} \).
Corollary 14. Let \( p \) be a multiquadratic polynomial of the form in (84) with \( M \) described by (82). Then

\[
\max \|z^{(1)}\|_2 = \ldots = \|z^{(l)}\|_2 = 1 \quad p(z^{(1)}, \ldots, z^{(l)})
\]

(85)
can be computed to within additive error \( \epsilon \) by \( O(\log(n)t^4/\varepsilon^2) \) levels of the sum-of-squares hierarchy, in time \( \exp(O(t^5 \log^2(n)/\varepsilon^2)) \)

Proof of Corollary 14. Directly applying part 1 of Theorem 2 to this problem would yield an approximation of \( \max_{\sigma} \text{tr} M \sigma \otimes I \). To instead relate this to \( h_{\text{Sep}}(A^\otimes l)(M) \) we introduce systems \( A_i^j \) for \( i \in [k] \) and \( j \in [l] \), for \( k \) to be determined later. Now consider \( M \) to be an operator on systems \( A_1^1 A_2^2 \cdots A_l^l \), and define \( \tilde{M} \) to be the operator acting on \( \otimes_{i,j} A_i^j \) with \( M \) acting on \( A_1^1 \cdots A_l^l \) and \( I \) acting on the other positions.

Define the \( n^l \)-dimensional block systems \( B_i := A_i^1 \cdots A_i^l \). Let \( \rho \) be a permutation-symmetric state on \( D(B_1 \otimes \cdots \otimes B_k) \). According to part 1 of Theorem 2, there exists a measure \( \nu \) over density matrices such that, if \( \omega = \int \nu(d\sigma)\sigma \otimes I \), then

\[
\sqrt{2l^2 \cdot l \ln(n)} / k - l \geq | \text{tr}(\tilde{M}(\rho B_1 \cdots B_l - \omega))| = | \text{tr}(M(\rho A_1^1 \cdots A_l^l - \omega A_1^1 \cdots A_l^l)) |. \quad (86)
\]

The maximum of \( \text{tr} \tilde{M} \omega \) over all \( \omega \) is given by \( h_{\text{Sep}}(A^\otimes l)(M) \). According to (86), and the fact that one possible solution for \( \rho \) is given by \( (\otimes_{i=1}^l |z^{(i)}\rangle \langle z^{(i)}|)^{\otimes m} \), we have

\[
h_{\text{Sep}}(A^\otimes l)(M) \leq \max_{\rho} \text{tr} \tilde{M} \rho \leq h_{\text{Sep}}(A^\otimes l)(M) + \sqrt{2l^3 \ln(n) / k - l}. \quad (87)
\]

Finally, optimizing over \( \rho \) is a strictly weaker relaxation of \( h_{\text{Sep}} \) than is provided by the SOS hierarchy. That hierarchy would also have constraints such as the PPT condition. See Lemma 9.10 of [10] for a similar comparison. Thus, the SOS hierarchy achieves bounds at least as tight as those of (87). We conclude by taking \( k = O(l^3 \ln(n)/\varepsilon^2) \). Since each \( B_i \) is a collection of \( l \) basic variables, this corresponds to level \( O(t^4 \ln(n)/\varepsilon^2) \) of the SOS hierarchy. The runtime bound follows from the fact that the total dimension of the resulting SDP is \( n^{kl} \). \( \square \)

Note that the result of Chen and Drucker [28] implies that \( \log(n)t^{2-o(1)} \) levels of the sum-of-squares hierarchy are not sufficient to compute even a constant-error approximation to (85), for general \( p \) of the form described in the corollary, unless there is a subexponential time algorithm for 3-SAT. It is an open question whether Corollary 14 can be improved to replace the \( l^4 \) with \( l^2 \) to match this bound. (Recall from Sect. 5.2 that the Chen-Drucker result cannot be substantially improved without contradicting the exponential-time hypothesis.)

There is also more direct evidence that Corollary 14 cannot be improved to yield a PTAS for polynomial optimization over the unit sphere. Ref. [25] proved that for any \( n \), there exists a local measurement \( M \) (derived from a Bell inequality) on \( n \times n \) systems such that

\[
\frac{\text{tr}(M \Phi_n)}{h_{\text{Sep}}(M)} \geq \Omega \left( \frac{n}{\log^2(n)} \right),
\]

(88)

with \( \Phi_n \) the projector onto the \( n \)-dimensional maximally entangled state. Since \( \rho := \frac{1}{k} \Phi_n + (1 - \frac{1}{k}) \mathbb{I} \) can be shown to be \( k \)-extendable, it follows that the \( k \)-extendable
approximation can make multiplicative errors as large as $\Omega\left(\frac{n}{k \log^2(n)}\right)$. Intriguingly, the example of [25] is based on the unique games problem. This suggests that using de Finetti theorems to give algorithms for unique games, as suggested by [10], will need to take advantage of the PPT condition (that the partial transpose of the state with respect to any of the local vector spaces should still be positive semidefinite) in addition to merely the $k$-extendability property. The only previous evidence that using the PPT condition gives an asymptotic improvement over mere $k$-extendability was given by [68].

5.4. Testing multipartite separability. Another application of part 1 of Theorem 2, closely related to section 5.3, is to the quantum separability problem, a well-studied problem in quantum information theory of both theoretical and practical interest [47]. Given a multipartite state $\rho^{A_1 \cdots A_l}$ we say it is fully separable if

$$\rho^{A_1 \cdots A_l} = \sum_j p_j \sigma_j^{A_1} \otimes \cdots \otimes \sigma_j^{A_l},$$

for a probability distribution $\{p_j\}$ and quantum states $\sigma_j^{A_i}$.

The goal in the weak-membership problem for separability is to decide whether a given multipartite state $\rho^{A_1 \cdots A_l}$ is separable or if it is $\epsilon$-away from any separable state, given the promise that one of the two alternatives holds true. In fact one has a family of problems depending on which norm we choose to quantify the distance of quantum states. We consider two choices of norms. The first is the one-way LOCC norm, defined for a Hermitian matrix acting on a $l$-partite vector space as

$$\|X\|_{\text{LOCC}^-} := \max_{\Lambda_2, \ldots, \Lambda_l} \|\text{id} \otimes \Lambda_2 \otimes \cdots \otimes \Lambda_l(X)\|_1.$$  

The name comes from the interpretation of norm as $\max_M \text{tr}(MX)$, with $M$ any POVM element that can be implemented by parties $2, \ldots, l$ measuring their systems locally and communicating the outcome to party 1, who then performs a measurement dependent on the information received. Therefore we have one-directional communication from all the parties to party 1.

The second is a multipartite version of the Frobenius norm recently introduced by Lancien and Winter [60]:

$$\|X\|_{2/l} := \sqrt{\sum_{I \subseteq [l]} \text{tr} |\text{tr}_I X|^2},$$

with $\text{tr}_I$ the partial trace over the subsystems indexed by $l$.

**Corollary 15.** For some $c > 0$, the Sum-of-Squares hierarchy solves the weak membership problem for separability for the norm $\| \ast \|_{\text{LOCC}^-}$ in time

$$\exp \left( c \left( \sum_j \log |A_j| \right)^2 l^2 \epsilon^{-2} \right).$$

In turn, the Sum-of-Squares hierarchy solves the weak membership problem for separability for the norm $\| \ast \|_{2/l}$ in time
\[
\exp \left( c \left( \sum_j \log |A_j| \right) \right)^2 (18)^{l/2} 2^l \varepsilon^{-2} \right). \tag{93}
\]

We note this gives a generalization of the result of [20], which proved the same result for bipartite quantum states. A early generalization of [20] to multipartite states was given in [22]; however there only a bound of
\[
\exp \left( c \log |A_1| \cdots \log |A_l| 2^{2l-1} \varepsilon^{-2(l-1)} \right) \tag{94}
\]
was obtained for the running time of the algorithm.

**Proof of Corollary 15.** According to the promise of the weak membership problem, we are given a state \( \rho^{A_1, \ldots, A_l} \in D(A_1 \otimes \cdots \otimes A_l) \) and wish to determine whether it is separable or \( \varepsilon \)-far from separable in the LOCC\( ^{\leq} \) norm.

The idea of the proof is to approximate the set \( \text{Sep}(A_1 : \cdots : A_l) \) with its \( k \)-extendible relaxation, for \( k \) chosen to give a good approximation guarantee according to Theorem 2. This means introducing systems \( X^1, \ldots, X^k \), each of which is composed of \( l \) subsystems (i.e. \( X^j := X^j_1 \cdots X^j_k \) for each \( j \), with \( X^j_i \equiv A_i \) for each \( i, j \)), and searching for a state \( \sigma^{X^1 \cdots X^k} \) such that

1. \( \sigma \) is invariant under permutation of the \( X^1, \ldots, X^k \) subsystems;
2. and \( \sigma^{X^1_i \cdots X^k_i} = \rho^{A_1 A_2 \cdots A_l} \).

Given a separable \( \rho^{A_1 \cdots A_l} \), such an extension \( \sigma^{X^1 \cdots X^k} \) exists for every \( k \geq l \). We can determine whether such a \( \sigma \) exists using semidefinite programming in time polynomial in the overall dimension of \( \sigma \). If we choose \( k = l + 4l^2 \varepsilon^{-2} \sum_j \log |A_j| \), then this will yield the runtime claimed in (92). Moreover, by Theorem 2 we have that there exists a measure \( \nu \) on \( D(A_1 \otimes \cdots \otimes A_l) \) such that

\[
\max_{\Gamma_2, \ldots, \Gamma_l \in M} \left\| (id \otimes \Gamma_2 \otimes \cdots \otimes \Gamma_l) \left( \sigma^{X^1 \cdots X^l} - \int \nu(\omega) \omega^{\otimes l} \right) \right\|_1 \leq \varepsilon, \tag{95}
\]

where \( \Gamma_2, \ldots, \Gamma_l \) range over all measurements of \( X^2, \ldots, X^l \). Restricting to measurements on the \( X^2_i, \ldots, X^l_i \) subsystems (which we denote \( \Lambda_2, \ldots, \Lambda_l \)) and using the monotonicity of the trace norm under partial trace, we obtain

\[
\min_{\omega \in \text{Sep}(A_1 : \cdots : A_l)} \max_{\Lambda_2, \ldots, \Lambda_l \in M} \left\| (id \otimes \Lambda_2 \otimes \cdots \otimes \Lambda_l) \left( \sigma^{X^1_i \cdots X^l_i} - \omega \right) \right\|_1 \leq \varepsilon. \tag{96}
\]

Of course, \( \sigma^{X^1_i \cdots X^l_i} \) in (96) is equal to \( \rho^{A_1 \cdots A_l} \), and so the existence of the symmetric extension \( \sigma \) implies that \( \rho \) is no more than \( \varepsilon \)-far from separable in the one-way LOCC norm. Conversely, if \( \rho \) is more than \( \varepsilon \)-far from separable in the LOCC\( ^{\leq} \) norm, then such a \( \sigma \) will not exist, and thus our algorithm will be able to distinguish this case from the one where \( \rho \) is separable.

The bound for \( \| * \|_{2(l)} \), follows from the reasoning above and the following bound (given by Theorem 5 of [60]):

\[
\| X \|_{\text{LOCC}^{\leq}} \geq 18^{-l/2} \| X \|_{2(l)}. \tag{97}
\]

\( \square \)
5.5. **Pretty-good tomography in permutation-symmetric states.** A final application of part 1 of Theorem 2 is to quantum state tomography, in which one obtains a description of an unknown quantum system by making measurements on the system. In quantum state tomography one tries to obtain a classical description of an unknown quantum state by finding its density matrix. If one has access to many i.i.d. copies of an unknown quantum state, then one can perform measurements on those copies in order to learn the identity of the quantum state. Mathematically, we can model this situation as saying that the global quantum state is of the form

\[ \omega_n = \int \sigma^\otimes n \mu(d\sigma), \]  

(98)

for an unknown measure \( \mu \) on quantum states. However the assumption of having many i.i.d. copies of an unknown state cannot always be ensured, and in many situations it simply does not hold true. It is thus an important task to try to relax this requirement. It has long been realized \[27\] that quantum de Finetti theorems are exactly the right tool here. Instead of having to assume that \( \omega_n \) has the form given by Eq. (98), one can merely assume that \( \omega_n \) is the reduced state of a larger permutation-symmetric state \( \omega_n^{+k} \). Then for \( k \) sufficiently large \( \omega_n \) will be close to a convex combination of i.i.d. states. The point is that one can easily ensure the latter situation by selecting \( n \) subsystems at random from the \( n + k \) available ones. Our work will allow this i.i.d. assumption to be relaxed. Indeed this was one of the original motivations for quantum de Finetti theorems \[27\].

The state of affairs is more complicated once complexity is taken into account. Since a quantum state of \( l \) qubits has \( 4^l \) parameters, reconstructing it generally requires \( 2^{O(l)} \) different measurement settings. However, in many cases we may care about only a small subset of those parameters. For instance, in order to predict expectation values of single-qubit observables, then a linear number of parameters suffices. Is there a way to explore this intuition in order to construct more efficient tomographic schemes?

One beautiful result in this direction was obtained by Aaronson in Ref. \[1\], using tools from computational learning theory \[52\], and can be roughly stated as follows: Given an arbitrary distribution \( \mathcal{M} \) over measurements and an unknown quantum state on \( l \) qubits, \( O(l) \) measurements settings are sufficient to get a density matrix which, with high probability over the measurement choice from \( \mathcal{M} \), agrees with the expectation of the true quantum state up to small error. Thus a linear – in the number of qubits of the state – number of measurement settings are enough to get a density matrix which gives a good estimate to the statistics of the true state for almost all choices of measurements; one can perform a “pretty-good” tomography just with a linear number of measurement settings. The formal statement of Aaronson’s result is as follows, restated slightly in order to facilitate our later extension of the result.

**Lemma 16.** (Theorem 1.3 of \[1\]). Let \( \gamma, \epsilon, \delta > 0 \). Let \( \omega_{m+n} \in \mathcal{D}(\mathcal{H}^\otimes m+n) \) be a state of the form

\[ \omega_{m+n} = \int v(d\rho) \rho^\otimes m+n, \]

for a probability measure \( v \) on \( \mathcal{D}(\mathcal{H}) \). Let \( \mathcal{M} \) be a distribution over two-outcome measurements on \( \mathcal{H} \) and \( \mathcal{E} = (E_1, \ldots, E_m) \) a training set of independently sampled measurements from \( \mathcal{M} \). Suppose we measure the first \( m \) systems of \( \omega \) according to \( \mathcal{E} \) and obtain outcomes \( B = (b_1, \ldots, b_m) \in \{0,1\}^m \). For any outcome \( B \), we will choose a hypothesis state

\[ \sigma_B := \arg \min_\sigma \sum_{i=1}^m (\tr(E_i \sigma) - b_i)^2. \]

(99)
Then there exists a constant $K > 0$ such that if
\[
m \geq \frac{K}{\gamma^4 \varepsilon^2} \left( \log |\mathcal{H}| \log^2 \frac{1}{\gamma \varepsilon} + \log \frac{1}{\delta} \right),
\] (100)
then with probability at least $1 - \delta$ the post-measured state $\tilde{\omega}_n$ satisfies
\[
\tilde{\omega}_n = \int \rho \otimes^n \mu(d\rho),
\] (101)
where the measure $\mu$ only has non-zero support on states $\rho$ such that
\[
\Pr_{E \in \mathcal{M}} \left[ |\text{tr}(E\rho) - \text{tr}(E\sigma_B)| > \gamma \right] \leq \varepsilon.
\] (102)

A limitation of Aaronson’s result [1], common of other tomographic schemes as well, is the assumption that one is given several i.i.d. copies of the unknown quantum state. Here too one could try to apply the standard quantum de Finetti theorems [30,59,77] to find a way around this assumption. However since the error in those depend polynomially on the dimension of the state, one would obtain a non-trivial result only if one would select subsystems at random from a state of $2^{O(l)}$ subsystems, which is not a reasonable assumption. Theorem 2 allows us to circumvent this problem.

**Corollary 17.** Let $\omega_{m+n+k} \in \mathcal{D}(\mathcal{H}^{\otimes m+n+k})$ be a permutation-symmetric state, let $\mathcal{M}$ be a distribution over two-outcome measurements on $\mathcal{H}$, and let $\mathcal{E} = (E_1, \ldots, E_m)$ be a training set consisting of $m$ measurements drawn independently from $\mathcal{M}$. Suppose we discard the last $k$ systems, measure the first $m$ systems of $\omega$ according to $\mathcal{E}$ and obtain outcomes $B = (b_1, \ldots, b_m) \in \{0, 1\}^m$. For any outcome $B$, we will choose a hypothesis state
\[
\sigma_B := \arg \min_{\sigma} \sum_{i=1}^{m} (\text{tr}(E_i \sigma) - b_i)^2.
\] (103)
Fix error parameters $\varepsilon, \eta, \gamma, \nu > 0$. Suppose that (for some universal constant $K > 0$) we have
\[
m \geq \frac{K}{\gamma^4 \varepsilon^2} \left( \log |\mathcal{H}| \log^2 \frac{1}{\gamma \varepsilon} + \log \frac{1}{\delta} \right),
\] (104)
\[
k \geq \frac{4(m + n)^2 \ln |\mathcal{H}|}{\nu^2}.
\] (105)
Then with probability at least $1 - \delta$ the post-measured state $\tilde{\omega}_n$ satisfies
\[
\max_{\Lambda_1, \ldots, \Lambda_n} \left\| \Lambda_1 \otimes \ldots \otimes \Lambda_n \left( \tilde{\omega}_n - \int \rho \otimes^n \mu(d\rho) \right) \right\|_1 \leq \nu,
\] (106)
with the maximum over quantum-classical channels $\Lambda_1, \ldots, \Lambda_n$. Here the measure $\mu$ only has non-zero support on states $\rho$ such that
\[
\Pr_{E \in \mathcal{M}} \left[ |\text{tr}(E\rho) - \text{tr}(E\sigma_B)| > \gamma \right] \leq \varepsilon
\] (107)
The proof of Corollary 17 follows immediately from part 1 of Theorem 2 and Lemma 16.

Let us say a few words about the interpretation of the result. Suppose we had Eq. (106) with \( \nu = 0 \). Then

\[
\tilde{\omega}_n = \int \rho \otimes^n \mu(d\rho),
\]

(108)

with \( \mu \) a measure with non-zero support only on states \( \rho \) that, for most measurements from \( \mathcal{M} \), gives approximately the same statistics as any state \( \sigma_B \) compatible with the observed data (in the sense that it satisfies Eq. (103)). Therefore any state \( \sigma_B \) compatible with the measured data can be used correctly to infer the statistics of future measurements, with high probability over the choice of the observable. For non-zero \( \nu \) we have a similar situation. While the state \( \tilde{\omega}_n \) might be very far away from a convex combination of i.i.d. in trace norm, if we only consider the statistics of local measurements on the \( n \) subsystems, then, up to error \( \nu \), we have the same conclusions as in the case of \( \nu = 0 \).

The price we have to pay for being able to relax the assumption of having i.i.d. copies of the state is that instead of starting from \( O(\log |\mathcal{H}| + n) \) copies of the state, now we need a global state with \( O((n + \log |\mathcal{H}|)^2 \log |\mathcal{H}|) \) subsystems (of which we only measure \( O(\log |\mathcal{H}|) \) of them). The main point is that this is still polynomial in the number of qubits of the unknown state one wants to learn.

We note that while this approach gives an efficient alternative for tomography of states on a large number of qubits in what concerns the number of measurements needed, it says nothing about the computational complexity of finding the hypothesis state \( \sigma_B \). As noted in [1], it is an interesting problem to determine for which classes of states one can obtain \( \sigma_B \) efficiently.

6. Open Problems

It would be desirable to strengthen several of the results in this work:

1. Conjecture 8 is a proposed improvement of Theorem 1 that would imply that \( O(\log(k)) \)-extendable states cannot be distinguished from separable states by Bell measurements with \( k \) outcomes per party. As we discuss in Sect. 5.1 this would have a very interesting application to the complexity of non-local games.

2. We would also like to improve Theorem 1 to apply to separable measurements\(^{10} \) instead of merely 1-LOCC measurements. If this were true, it would imply by the results of [42], that \( \text{QMA}_n(m, c, s) \subseteq \text{QMA}_{O(mn^2/\epsilon)}(1, c, s + \epsilon) \). It would also yield quasipolynomial-time classical algorithms for separability testing and a large number of tensor optimization problems described in [42].

3. One of the few barriers to improving de Finetti theorems is the example of the maximally mixed state on the antisymmetric subspace of \( \mathbb{C}^d \otimes \mathbb{C}^d \) [30]. This so-called “universal counterexample” state is \( d \)-extendable, and yet is far from separable. However, this distinguishability cannot be achieved by a measurement whose “not separable” outcome is itself a separable measurement operator; aka a “SEP-YES” measurement. As mentioned in the previous open problem, proving a more efficient de Finetti theorem against such measurements would improve the algorithm for approximating \( h_{\text{Sep}}(M) \) for general measurements \( M \). Additionally, the antisymmetric state is not PPT, and such examples of highly-extendable far-from-separable states

\(^{10}\) Technically, we refer here to approximating \( h_{\text{Sep}}(M) \) for “SEP-YES” measurements, meaning that \( M \) is of the form \( \sum_i A_i \otimes B_i \) for p.s.d. \( A_i, B_i \), but without any such requirement for \( I - M \).
are not known to occur when we add the PPT constraint, as proposed by [38]. Intriguingly, the “worst” known example (i.e. most extendable while being far from separable) of a PPT state is only $O(\log d)$-extendable [24].\footnote{In more detail this follows by considering a variant of the example of [24] (page 6) in which the EPR pair is replaced by a constant-dimensional PPT entangled state. Note that by [13] for every $\varepsilon > 0$ there is a PPT state with trace distance $2 - \varepsilon$ from separable states, thus for every $\varepsilon > 0$ one can get a PPT $O(\log(d))$-extendible $d \times d$ state which is $(2 - \varepsilon)$-away from any separable state. The same holds true if we use the 1-LOCC version of the trace norm.} It would be of great interest either to prove a better bound on the combination of PPT and $k$-extendable constraints (see [68] or Section 9.3.2 of [10] for some progress), or to find better counterexample states.

4. It would also be interesting to use our information-theoretic techniques to examine the various extensions of the de Finetti theorem. For example, is there a version of the post-selection technique [29] where the dimension dependence is replaced by a dependence on the number of measurement outcomes? One difficulty here (highlighted by taking the local dimension to be infinite) is in choosing the right test state upon which the channels should act. Another question is whether our techniques can improve the exponential de Finetti theorem [76]. Unfortunately, this theorem is known not to have a classical analogue (due to unpublished work of Christandl and Toner), while our proofs use entropic properties of classical, or classical-quantum, states.

Acknowledgements. We are grateful to Kevin Milner, Thomas Vidick and Mark Wilde for many helpful comments on an early version of the paper, to Ashley Montanaro for explaining to us the remark at the end of Sect. 5.2, to Scott Aaronson for telling us about [3] in 2010, to Graeme Smith and Ke Li for catching a bug in an earlier version of Corollary 14 and especially to Boaz Barak, Jon Kelner and David Steurer for sharing with us an early version of [11]. We also benefited from interesting discussions with Matthias Christandl and Stephanie Wehner. Much of this work was done while FGSLB was working at the Institute for Theoretical Physics in ETH Zürich and AWH was working in the Department of Computer Science at the University of Washington. FGSLB acknowledges support from EPSRC through an Early Career Fellowship, the Polish Ministry of Science and Higher Education Grant No. IdP2011 000361, the Swiss National Science Foundation, via the National Centre of Competence in Research QSIT, the German Science Foundation (Grant CH 843/2-1), the Swiss National Science Foundation (Grants PP00P2_128455, 20CH21_138799 (CHIST-ERA project CQC)), the Swiss National Center of Competence in Research “Quantum Science and Technology (QSIT)”, and the Swiss State Secretariat for Education and Research supporting COST action MP1006. AWH was funded by NSF Grants 0916400, 0829937, 0803478, 1111382, 1452616 and 1629809, and ARO contract W911NF-12-1-0486.

References

1. Aaronson, S.: The learnability of quantum states. Proc. R. Soc. A \textbf{463}, 2088 \texttt{arXiv:quant-ph/0608142} (2007)
2. Aaronson, S., Beigi, S., Drucker, A., Fefferman, B., Shor, P.: The power of unentanglement. Theory Comput. \textbf{5}(1), 1–42. \texttt{arXiv:0804.0802} (2009)
3. Aaronson, S., Impagliazzo, R., Moshkovitz, D.: Am with multiple merlins. In: Proceedings of the 2014 IEEE 29th Conference on Computational Complexity, CCC ’14, pp. 44–55. \texttt{arXiv:1401.6848} (2014)
4. Arora, S., Barak, B.: Computational Complexity: A Modern Approach. Cambridge University Press, Cambridge (2009)
5. Arora, S., Barak, B., Steurer, D.: Subexponential algorithms for unique games and related problems. In: FOCS, pp. 563–572 (2010)
6. Arora, S., Lund, C., Motwani, R., Sudan, M., Szegedy, M.: Proof verification and the hardness of approximation problems. J. ACM \textbf{45}(3), 501–555 (1998)
7. Arora, S., Safra, S.: Probabilistic checking of proofs: a new characterization of NP. J. ACM \textbf{45}(1), 70–122 (1998)
8. Aspect, A., Dalibard, J., Roger, G.: Experimental test of Bell’s inequalities using time-varying analyzers. Phys. Rev. Lett. 49, 1804–1807 (1982)

9. Babai, L., Fortnow, L., Lund, C.: Non-deterministic exponential time has two-prover interactive protocols. Comput. Complex. 1, 3–40 (1991)

10. Barak, B., Brandão, F.G., Harrow, A.W., Kelner, J., Steurer, D., Zhou, Y.: Hypercontractivity, sum-of-squares proofs, and their applications. In: STOC ‘12, STOC ‘12, pp. 307–326. arXiv:1205.4484 (2012)

11. Barak, B., Kelner, J., Steurer, D.: Rounding sum-of-squares relaxations. In: STOC. arXiv:1312.6652 (2014)

12. Barrett, J., Leifer, M.: The de Finetti theorem for test spaces. New J. Phys. 11, 033024. arXiv:0712.2265 (2009)

13. Beigi, S., Shor, P.: Approximating the set of separable states using the positive partial transpose test. J. Math. Phys. 51, 042202. arXiv:0904.2705 (2010)

14. Bell, J.: On the Einstein–Podolsky–Rosen paradox. Physics 1, 195 (1964)

15. Bellare, M., Feige, U., Kilian, J.: On the role of shared randomness in two prover proof systems. In: ISTCS ‘95, pp. 199–208 (1995)

16. Bennett, C., Bernstein, E., Brassard, G., Vazirani, U.: The strengths and weaknesses of quantum computation. SIAM J. Comput. 26, 1510–1523. arXiv:quant-ph/9701001 (1997)

17. Bennett, C.H., DiVincenzo, D.P., Smolin, J.A., Wootters, W.K.: Mixed-state entanglement and quantum error correction. Phys. Rev. A 52, 3824–3851. arXiv:quant-ph/9604024 (1996)

18. Blier, H., Tapp, A.: All languages in NP have very short quantum proofs. In: First International Conference on Quantum, Nano, and Micro Technologies, pp. 34–37. IEEE Computer Society, Los Alamitos, CA, USA. arXiv:0709.0738 (2009)

19. Brandão, F.: Entanglement theory and the quantum simulation of many-body physics. Ph.D. thesis, Imperial College London arXiv:0810.0026 (2008)

20. Brandão, F.G., Christandl, M., Yard, J.: A quasipolynomial-time algorithm for the quantum separability problem. In: Proceedings of the 43rd Annual ACM Symposium on Theory of Computing, STOC ‘11, pp. 343–352. arXiv:1011.2751 (2011)

21. Brandão, F.G., Plenio, M.B.: A generalization of quantum Stein’s lemma. Commun. Math. Phys. 295, 791. arXiv:0904.0281 (2010)

22. Brandão, F.G.S.L., Christandl, M.: Detection of multiparticle entanglement: quantifying the search for symmetric extensions. Phys. Rev. Lett. 109, 160502. arXiv:1105.5720 (2012)

23. Brandão, F.G.S.L., Harrow, A.W.: Product-state approximations to quantum ground states. In: Proceedings of the 45th Annual ACM Symposium on Theory of Computing, STOC ‘13, pp. 871–880. arXiv:1310.0017 (2013)

24. Brandão, F.G.S.L., Christandl, M., Yard, J.: Faithful squashed entanglement. Commun. Math. Phys. 306(3), 805–830. arXiv:1010.1750 (2011)

25. Buhrman, H., Regev, O., Scarpa, G., de Wolf, R.: Near-optimal and explicit Bell inequality violations. In: CCC ’11, pp. 157–166 arXiv:1012.5043 (2011)

26. Cai, J.-Y., Condon, A., Lipton, R.: Playing games of incomplete information. In: STACS ’90, vol. 415, pp. 58–69 (1990)

27. Caves, C.M., Fuchs, C.A., Schack, R.: Unknown quantum states: the quantum de Finetti representation. J. Math. Phys. 43(9), 4537–4559. arXiv:quant-ph/0104088 (2002)

28. Chen, J., Drucker, A.: Short multi-prover quantum proofs for SAT without entangled measurements. arXiv:1011.0716 (2010)

29. Christandl, M., Koenig, R., Renner, R.: Post-selection technique for quantum channels with applications to quantum cryptography. Phys. Rev. Lett. 102, 020504. arXiv:0809.3019 (2009)

30. Christandl, M., König, R., Mitchison, G., Renner, R.: One-and-a-half quantum de Finetti theorems. Commun. Math. Phys. 273, 473–498. arXiv:quant-ph/0602130 (2007)

31. Cleve, R., Gavinsky, D., Jain, R.: Entanglement-resistant two-prover interactive proof systems and non-adaptive PIRs. Quantum Inf. Comput. 9(7), 648–656. arXiv:0707.1729 (2007)

32. Cleve, R., Hoyer, P., Toner, B., Watrous, J.: Consequences and limits of nonlocal strategies. In: CCC ’04, pp. 236–249. arXiv:quant-ph/0404076 (2004)

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

34. D’Angelo, J.P., Putinar, M.: Polynomial optimization on odd-dimensional spheres. In: Putinar, M., Sullivant, S. (eds.) Emerging Applications of Algebraic Geometry, Volume 149 of The IMA Volumes in Mathematics and Its Applications, pp. 1–15. Springer, New York (2009)

35. Klerk, E.de.: The complexity of optimizing over a simplex, hypercube or sphere: a short survey. Cent. Eur. J. Oper. Res. 16(2), 111–125 (2008)

36. Diaconis, P., Freedman, D.: Finite exchangeable sequences. Ann. Probab. 8, 745–764 (1980)
38. Doherty, A.C., Parrilo, P.A., Spedalieri, F.M.: Complete family of separability criteria. Phys. Rev. A 69, 022308. arXiv:quant-ph/0308032 (2004)

39. Doherty, A.C., Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hyper-sphere. arXiv:1210.5048 (2012)

40. Fannes, M., Lewis, J.T., Verbeure, A.: Symmetric states of composite systems. Lett. Math. Phys. 15, 255–260 (1988)

41. Harrow, A.W., Montanaro, A.: An efficient test for product states, with applications to quantum Merlin-Arthur games. In: FOCS ’10, pp. 633–642. arXiv:1001.0017 (2010)

42. Harrow, A.W., Montanaro, A.: Testing product states, quantum Merlin-Arthur games and tensor optimization. J. ACM 60(1), 3–1343. arXiv:1001.0017 (2013)

43. Hayden, P., Winter, A.: Weak decoupling duality and quantum identification. IEEE Trans. Inf. Theory 58(7), 4914–4929. arXiv:1003.4994 (2012)

44. Hudson, R.L., Moody, G.R.: Locally normal symmetric states and an analogue of de Finetti’s theorem. Z. Wahrschein. Verw. Geb. 33, 343–351 (1976)

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

46. Impagliazzo, R., Paturi, R., Zane, F.: Which problems have strongly exponential complexity? In: FOCS’98, pp. 653–662. IEEE (1998)

47. Ioannou, L.: Computational complexity of the quantum separability problem. Quantum Inf. Comput. 7(4), 335 (2007)

48. Kitaev, A.Y., Shen, A.H., Vyalyi, M.N.: Classical and Quantum Computation, Volume 47 of Graduate Studies in Mathematics. AMS, Providence (2002)

49. Koenig, R., Renner, R.: A de Finetti representation for finite symmetric quantum states. J. Math. Phys. 46(12), 122108. arXiv:quant-ph/0410229 (2005)

50. Lancien, C., Winter, A.: Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. Commun. Math. Phys. 323(2), 555–573. arXiv:1206.2884 (2013)

51. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Opt. 11(3), 796–815 (2001)

52. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. Emerg. Appl. Algebraic Geom. 149, 157–270 (2009)

53. Li, K., Smith, G.: Quantum de finetti theorem under fully-one-way adaptive measurements. Phys. Rev. Lett. 114, 160503. arXiv:1408.6829 (2015)

54. Marriott, C., Watrous, J.: Quantum Arthur-Merlin games. Comput. Complex. 14(2), 122–152. arXiv:cs/0506068 (2005)

55. Masanes, L., Acín, A., Gisin, N.: General properties of nonsignaling theories. Phys. Rev. A 73, 021112. arXiv:quant-ph/0508016 (2006)

56. Matthews, W., Wehner, S., Winter, A.: Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. Commun. Math. Phys. 291(3), 813–843. arXiv:0810.2327 (2009)
68. Navascues, M., Owari, M., Plenio, M.B.: The power of symmetric extensions for entanglement detection. Phys. Rev. A 80, 052306. arXiv:0906.2731 (2009)

69. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, New York (2000)

70. O’Donnell, R., Zhou, Y.: Approximability and proof complexity. In: SODA ’13, pp. 1537–1556. arXiv:1211.1958 (2013)

71. Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, MIT (2000)

72. Piani, M.: Relative entropy of entanglement and restricted measurements. Phys. Rev. Lett. 103, 160504. arXiv:0904.2705 (2009)

73. Powers, V., Reznick, B.: A new bound for Pólya’s theorem with applications to polynomials positive on polyhedra. J. Pure Appl. Algebra 164(1–2), 221–229 (2001)

74. Raggio, G.A., Werner, R.F.: Quantum statistical mechanics of general mean field systems. Helv. Phys. Acta 62, 980–1003 (1989)

75. Raghavendra, P, Tan, N.: Approximating CSPs with global cardinality constraints using SDP hierarchies. In: SODA ’12, pp. 373–387. arXiv:1110.1064 (2012)

76. Renner, R.: Security of quantum key distribution. Ph.D. thesis, ETHZ, Zurich. arXiv:quant-ph/0512258 (2005)

77. Renner, R.: Symmetry implies independence. Nat. Phys. 3, 645–649. arXiv:quant-ph/0703069 (2007)

78. Sion, M.: On general minimax theorems. Pac. J. Math. 8(1), 171–176 (1958)

79. Størmer, E.: Symmetric states of infinite tensor products of C*-algebras. J. Funct. Anal. 3, 48 (1969)

80. Terhal, B.M., Doherty, A.C., Schwab, D.: Symmetric extensions of quantum states and local hidden variable theories. Phys. Rev. Lett. 90, 157903. arXiv:quant-ph/0210053 (2003)

81. Vidick, T.: Three-player entangled XOR games are NP-hard to approximate. SIAM J. Comput. 45(3), 1007–1063. arXiv:1302.1242 (2016)

82. Werner, R.F.: An application of Bell’s inequalities to a quantum state extension problem. Lett. Math. Phys. 17, 359 (1989)

83. Wilde, M.M.: Quantum Information Theory. Cambridge University Press, Cambridge. (2013) arXiv:1106.1445

84. Yang, D.: A simple proof of monogamy of entanglement. Phys. Lett. A 360(2), 249–250. arXiv:quant-ph/0604168 (2006)

85. Yang, D., Horodecki, K., Horodecki, M., Horodecki, P., Oppenheim, J., Song, W.: Squashed entanglement for multipartite states and entanglement measures based on the mixed convex roof. IEEE Trans. Inf. Theory 55(7), 3375–3387. arXiv:0704.2236 (2009)

Communicated by M. M. Wolf