

Quantum de Finetti theorem under fully-one-way adaptive measurements

Ke Li^{1,2,*} and Graeme Smith^{1,†}

¹IBM TJ Watson Research Center, Yorktown Heights, NY 10598, USA

²Center for Theoretic Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

(Dated: April 29, 2015)

We prove a version of the quantum de Finetti theorem: permutation-invariant quantum states are well approximated as a probabilistic mixture of multi-fold product states. The approximation is measured by distinguishability under fully one-way LOCC (local operations and classical communication) measurements. Our result strengthens Brandão and Harrow's de Finetti theorem where a kind of partially one-way LOCC measurements was used for measuring the approximation, with essentially the same error bound. As main applications, we show (i) a quasipolynomial-time algorithm which detects multipartite entanglement with amount larger than an arbitrarily small constant (measured with a variant of the relative entropy of entanglement), and (ii) a proof that in quantum Merlin-Arthur proof systems, polynomially many provers are not more powerful than a single prover when the verifier is restricted to one-way LOCC operations.

Consider random variables X_1, \dots, X_n representing the color of a sequence of balls drawn without replacement from a bag of 100 red balls and 100 blue balls. These variables are not independent, since the probability of withdrawing a red ball on the k th withdrawal depends on the number of balls of each color remaining. They are, however, *exchangeable*: the probability of removing a particular sequence of balls (x_1, \dots, x_n) is equal to the probability of removing any reordering of that sequence $(x_{\pi(1)}, \dots, x_{\pi(n)})$ for permutation π . Remarkably, the de Finetti theorem tells us that any such exchangeable random variables can be represented by independent and identically distributed ones [1, 2], yielding a profound result in probability theory and a powerful tool in statistics.

A series of works have established analogues of this theorem in the quantum domain [3–10], where a classical probability distribution is replaced by a quantum state and the situation is more complicated and interesting, due to entanglement and the existence of many different ways to distinguish states of multipartite systems. These quantum de Finetti theorems are appealing not only due to their own elegance on the characterization of symmetric states, but also because of the successful applications in many-body physics [5, 11, 12], quantum information [9, 13, 14], and computational complexity theory [10, 15, 16].

More precisely, a quantum de Finetti theorem concerns the structure of a *symmetric* state $\rho_{A_1 \dots A_n}$ that is invariant under any permutations over the subsystems [17]. It tells how the reduced state $\rho_{A_1 \dots A_k}$ on a smaller number $k < n$ of subsystems could be approximated by a mixture of k -fold product states, namely, *de Finetti states* of the form $\int \sigma^{\otimes k} d\mu(\sigma)$. Here μ is a probability measure over density matrices. Using the conventional distance measure, trace norm, Ref. [8] proved a standard de Finetti theorem with an essentially optimal error bound $2|A|^{2k}/n$ for the approximation ($|A|$ denotes the dimension of the subsystems). However, in many situations this bound is too large to be applicable. Luckily it is possible

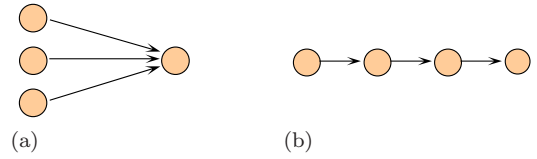


FIG. 1: Parallel vs. fully one-way LOCC. (a) $\text{LOCC}_1^{\parallel}$: Parallel one-way LOCC measurements used in [10]. Here the first $k-1$ parties make measurements in parallel and report their outcomes to the k th, who then makes a measurement that depends on the messages he receives. (b) LOCC_1 : Fully one-way LOCC measurements. We adopt a more complete generalization of one-way LOCC: all the parties measure their own systems sequentially, but in a fully adaptive way where each party chooses his own measurement setting depending on the outcomes of all the previous measurements performed by the other parties.

to circumvent this obstruction. For example, Renner's exponential de Finetti theorem employs the “almost” de Finetti states and has an error bound that decreases exponentially in $n-k$ [9], being very useful in dealing with cryptography or information theory problems [9, 13, 14].

In a beautiful work [10] Brandão and Harrow recently proved an LOCC (local operations and classical communication) de Finetti theorem, generalizing a similar result for the case $k=2$ [16]. Both [10] and [16] have overcome the limitation of the standard de Finetti theorem regarding the dimension dependence. The basic idea is to relax the measure of approximation by replacing the trace norm with a kind of one-way LOCC norm. This gives an error bound $\sqrt{\frac{2k^2 \ln |A|}{n-k}}$ [18], scaling polynomially in $\ln |A|$ instead of polynomially in $|A|$ as in earlier de Finetti results, which is crucial to the complexity-theoretic applications.

While [10] showed approximation in the *parallel* one-way LOCC norm associated with the measurement class $\text{LOCC}_1^{\parallel}$, here we prove a de Finetti theorem where the approximation is measured with the *fully* one-way LOCC

norm (or relative entropy) associated with LOCC_1 (cf. Fig. 1). The error bound remains essentially the same as that of [10]. This improves Brandão and Harrow's LOCC de Finetti theorem considerably: it is conceptually more complete and when applied to the problems considered in [10, 16, 19] gives new and improved results. For the problem of entanglement detection, so central to quantum information theory and experiment, we present strong guarantees for the effectiveness of the well-known hierarchy of entanglement tests of [20]. We also consider the power of multiple-prover quantum Merlin Arthur games, which bears directly on the problems of pure-state vs mixed-state N -representability [21] as well as the entanglement properties of sparse hamiltonian's ground states [22].

Operational norms as distance measures. We identify every positive operator-valued measure $\{M_x\}_x$ with a measurement operation \mathcal{M} : for any state ω , $\mathcal{M}(\omega) := \sum_x |x\rangle\langle x| \text{Tr}(\omega M_x)$ with $\{|x\rangle\}_x$ an orthonormal basis. For simplicity we call them both quantum measurement. Given a class of measurements \mathbf{M} , the operational norm is defined as [23]

$$\|\rho - \sigma\|_{\mathbf{M}} = \max_{\mathcal{M} \in \mathbf{M}} \|\mathcal{M}(\rho) - \mathcal{M}(\sigma)\|_1.$$

It measures the distinguishability of two quantum states under restricted classes of measurements. We will be particularly interested in $\|\cdot\|_{\text{LOCC}_1}$ and $\|\cdot\|_{\text{LOCC}_1^1}$. Obviously the former is lower bounded by the latter, since $\text{LOCC}_1^1 \subset \text{LOCC}_1$. In fact, these two norms can differ substantially: using a recent result obtained in [24], we can show for all d there are constant C and $d \times d \times 2$ states ρ_{ABC} and σ_{ABC} such that $\|\rho_{ABC} - \sigma_{ABC}\|_{\text{LOCC}_1} = 2$ but $\|\rho_{ABC} - \sigma_{ABC}\|_{\text{LOCC}_1^1} \leq C/\sqrt{d}$ (see Appendix).

Improved LOCC de Finetti theorem. Our main result is the following Theorem 1. Besides the improvement with the fully one-way LOCC norm, for the first time we employ relative entropy $D(\rho\|\sigma) = \text{Tr} \rho(\log \rho - \log \sigma)$ to measure the approximation, defining $D_{\text{LOCC}_1}(\rho\|\sigma) := \max_{\Lambda \in \text{LOCC}_1} D(\Lambda(\rho)\|\Lambda(\sigma))$.

In the proof, we will use information-theoretic methods similar to [10], along with some new ideas. In particular, Lemma 2 presented below is a crucial new technical tool, which may be of independent interest. We employ and manipulate entropic quantities to derive the final result: apart from relative entropy, the mutual information of a state ω_{AB} is defined as $I(A; B) := D(\omega_{AB}\|\omega_A \otimes \omega_B)$, and the conditional mutual information of a state ω_{ABC} is defined as $I(A; B|C) := I(A; BC) - I(A; C)$.

Theorem 1 *Let $\rho_{A_1 \dots A_n}$ be a permutation-invariant state on $\mathcal{H}_A^{\otimes n}$. Then for integer $0 \leq k \leq n$ there exists a probability measure μ on density matrices on \mathcal{H}_A*

such that

$$D_{\text{LOCC}_1}(\rho_{A_1 \dots A_k} \|\int \sigma^{\otimes k} d\mu(\sigma)) \leq \frac{(k-1)^2 \log |A|}{n-k}, \quad (1)$$

$$\left\| \rho_{A_1 \dots A_k} - \int \sigma^{\otimes k} d\mu(\sigma) \right\|_{\text{LOCC}_1} \leq \sqrt{\frac{2(k-1)^2 \ln |A|}{n-k}}. \quad (2)$$

Proof. Eq. (2) follows from Eq. (1) immediately by using the Pinsker's inequality [25], $D(\rho\|\sigma) \geq \frac{1}{2 \ln 2} \|\rho - \sigma\|_1^2$. So it suffices to prove Eq. (1).

Group the n subsystems as shown in Fig. 2: except for one subsystem, the others are divided into groups of $k-1$ subsystems each (we discard the possibly remaining qubits, of which there will be fewer than $k-1$). So, we have $m = \lfloor \frac{n-1}{k-1} \rfloor \geq \frac{n-k}{k-1}$ groups. Label the groups as bigger subsystems B_1, B_2, \dots, B_m and the isolated system as A . Let the $k-1$ subsystems in B_1 be A_1, A_2, \dots, A_{k-1} and the system A is also identified with A_k .

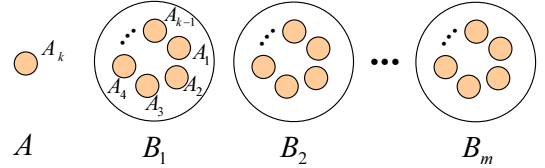


FIG. 2: Grouping and relabeling the n subsystems.

Obviously the total state is invariant under permutations over B_1, B_2, \dots, B_m . So Lemma 3 applies. Thus there exists a measurement $\mathcal{Q}^* : B_2 \dots B_m \rightarrow X$, such that for any measurement $\mathcal{P} : B_1 \rightarrow Y$ we have

$$I(A; Y|X) \leq \frac{\log |A|}{m} \leq \frac{(k-1) \log |A|}{n-k}. \quad (3)$$

\mathcal{Q}^* effectively decomposes the state on AB_1 into an ensemble. Specifically, we have $\rho_{AB_1} = \sum_x p_x \rho_{A_1 \dots A_k}^x$, where p_x is the probability of obtaining the measurement outcome x and $\rho_{A_1 \dots A_k}^x$ is the resulting state on $A_1 \dots A_k$. Note that since $\rho_{A_1 \dots A_n}$ is permutation-invariant, the post-measurement states $\rho_{A_1 \dots A_k}^x$ are also permutation-invariant. Now we rewrite Eq. (3) in terms of the relative entropy: for any measurement \mathcal{P} on $A_1 \dots A_{k-1}$,

$$\begin{aligned} & \sum_x p_x D(\mathcal{P} \otimes \text{id}^{A_k}(\rho_{A_1 \dots A_k}^x) \|\mathcal{P}(\rho_{A_1 \dots A_{k-1}}^x) \otimes \rho_{A_k}^x) \\ & \leq \frac{(k-1) \log |A|}{n-k}. \end{aligned} \quad (4)$$

Pick a one-way LOCC measurement Λ^k acting on systems A_1, \dots, A_k and denote its reduced measurement on the first ℓ systems as Λ^ℓ . Now we apply Lemma 2 to each

state $\rho_{A_1 \dots A_k}^x$ and get

$$\begin{aligned} & D(\Lambda^k(\rho_{A_1 \dots A_k}^x) \| \Lambda^k(\rho_{A_1}^x \otimes \dots \otimes \rho_{A_k}^x)) \\ & \leq \sum_{\ell=2}^k D(\Lambda^{\ell-1} \otimes \text{id}(\rho_{A_1 \dots A_{\ell}}^x) \| \Lambda^{\ell-1}(\rho_{A_1 \dots A_{\ell-1}}^x) \otimes \rho_{A_{\ell}}^x) \\ & \leq (k-1) D(\Lambda^{k-1} \otimes \text{id}(\rho_{A_1 \dots A_k}^x) \| \Lambda^{k-1}(\rho_{A_1 \dots A_{k-1}}^x) \otimes \rho_{A_k}^x), \end{aligned} \quad (5)$$

where for the first inequality we have also applied the monotonicity of relative entropy [26] and for the second inequality we used the monotonicity of relative entropy again as well as the symmetry of the state $\rho_{A_1 \dots A_k}^x$. Combining Eq. (4) and Eq. (5) we arrive at

$$\begin{aligned} & D(\Lambda^k(\rho_{A_1 \dots A_k}) \| \Lambda^k(\sum_x p_x \rho_{A_1}^x \otimes \dots \otimes \rho_{A_k}^x)) \\ & \leq \sum_x p_x D(\Lambda^k(\rho_{A_1 \dots A_k}^x) \| \Lambda^k(\rho_{A_1}^x \otimes \dots \otimes \rho_{A_k}^x)) \\ & \leq \frac{(k-1)^2 \log |A|}{n-k}, \end{aligned} \quad (6)$$

where the first inequality is due to the joint convexity of relative entropy. At this point we are able to conclude Eq. (1) from Eq. (6), noticing that $\Lambda^k \in \text{LOCC}_1$ is picked arbitrarily and $\sum_x p_x \rho_{A_1}^x \otimes \dots \otimes \rho_{A_k}^x$ is a de Finetti state of the form $\sum_x p_x (\rho_A^x)^{\otimes k}$ due to the symmetry of $\rho_{A_1 \dots A_k}^x$. \square

Lemma 2 *Let Λ^k be a one-way LOCC measurement on quantum systems A_1, \dots, A_k . Denote its reduced measurement corresponding to the first ℓ steps on A_1, \dots, A_{ℓ} as Λ^{ℓ} . Then for any state $\rho_{A_1 \dots A_k}$ we have*

$$\begin{aligned} & D(\Lambda^k(\rho_{A_1 \dots A_k}) \| \Lambda^k(\rho_{A_1} \otimes \dots \otimes \rho_{A_k})) \\ & = \sum_{\ell=2}^k D(\Lambda^{\ell}(\rho_{A_1 \dots A_{\ell}}) \| \Lambda^{\ell}(\rho_{A_1 \dots A_{\ell-1}} \otimes \rho_{A_{\ell}})). \end{aligned}$$

Proof. It suffices to show

$$\begin{aligned} & D(\Lambda^k(\rho_{A_1 \dots A_k}) \| \Lambda^k(\rho_{A_1} \otimes \dots \otimes \rho_{A_k})) \\ & = D(\Lambda^{k-1}(\rho_{A_1 \dots A_{k-1}}) \| \Lambda^{k-1}(\rho_{A_1} \otimes \dots \otimes \rho_{A_{k-1}})) \\ & \quad + D(\Lambda^k(\rho_{A_1 \dots A_k}) \| \Lambda^k(\rho_{A_1 \dots A_{k-1}} \otimes \rho_{A_k})), \end{aligned} \quad (7)$$

because applying this relation recursively allows us to obtain the equation claimed in Lemma 2. Write $\Lambda^{k-1}(\rho_{A_1 \dots A_{k-1}}) = \sum_x p_x |x\rangle\langle x|$ and $\Lambda^{k-1}(\rho_{A_1} \otimes \dots \otimes \rho_{A_{k-1}}) = \sum_x q_x |x\rangle\langle x|$. Let Λ^k be realized as follows. We first apply Λ^{k-1} on A_1, \dots, A_{k-1} . Then depending on the measurement outcome x we apply a measurement \mathcal{M}_x on A_k . Thus we can write

$$\begin{aligned} \Lambda^k(\rho_{A_1 \dots A_k}) & = \sum_x p_x |x\rangle\langle x| \otimes \mathcal{M}_x(\rho_{A_k}^x), \\ \Lambda^k(\rho_{A_1 \dots A_{k-1}} \otimes \rho_{A_k}) & = \sum_x p_x |x\rangle\langle x| \otimes \mathcal{M}_x(\rho_{A_k}), \\ \Lambda^k(\rho_{A_1} \otimes \dots \otimes \rho_{A_k}) & = \sum_x q_x |x\rangle\langle x| \otimes \mathcal{M}_x(\rho_{A_k}), \end{aligned}$$

where $\rho_{A_k}^x$ is the state of A_k when Λ^{k-1} is applied on $\rho_{A_1 \dots A_k}$ and outcome x is obtained. With these, we can confirm by direct computation that

$$\begin{aligned} & D(\Lambda^k(\rho_{A_1 \dots A_k}) \| \Lambda^k(\rho_{A_1} \otimes \dots \otimes \rho_{A_k})) \\ & = D(\Lambda^{k-1}(\rho_{A_1 \dots A_{k-1}}) \| \Lambda^{k-1}(\rho_{A_1} \otimes \dots \otimes \rho_{A_{k-1}})) \\ & \quad + \sum_x p_x D(\mathcal{M}_x(\rho_{A_k}^x) \| \mathcal{M}_x(\rho_{A_k})) \end{aligned} \quad (8)$$

and

$$\begin{aligned} & D(\Lambda^k(\rho_{A_1 \dots A_k}) \| \Lambda^k(\rho_{A_1 \dots A_{k-1}} \otimes \rho_{A_k})) \\ & = \sum_x p_x D(\mathcal{M}_x(\rho_{A_k}^x) \| \mathcal{M}_x(\rho_{A_k})). \end{aligned} \quad (9)$$

Eq. (8) and Eq. (9) together lead to Eq. (7) and this concludes the proof. \square

Remark. The quantity $D(\rho_{A_1 \dots A_k} \| \rho_{A_1} \otimes \dots \otimes \rho_{A_k})$ is sometimes denoted as $I(A_1; A_2; \dots; A_k)_{\rho}$ and called the multipartite mutual information. It is easy to see that $I(A_1; \dots; A_k) = I(A_1 \dots A_{\ell}; A_{\ell+1} \dots A_k) + I(A_1; \dots; A_{\ell}) + I(A_{\ell+1}; \dots; A_k)$. Using this repeatedly we can write the multipartite mutual information as a sum of bipartite mutual information quantities. This decomposition can be done in many different ways depending on how we split the subsystems. Lemma 2 is a similar result. However, with the one-way LOCC measurement Λ^k , the decomposition only works for our special choice of splitting.

The following lemma, which is a statement of the monogamy of entanglement, is adapted from [10]. For completeness we give a proof in the Appendix.

Lemma 3 *Let $\rho_{AB_1 \dots B_m}$ be a state that is invariant under any permutation over B_1, B_2, \dots, B_m . Let $\mathcal{P}^{B_1 \rightarrow Y}$ and $\mathcal{Q}^{B_2 \dots B_m \rightarrow X}$ be measurement operations performed on systems B_1 and $B_2 \dots B_m$, respectively. We have*

$$\min_{\mathcal{Q}} \max_{\mathcal{P}} I(A; Y|X)_{\text{id}^A \otimes \mathcal{P} \otimes \mathcal{Q}(\rho_{AB_1 \dots B_m})} \leq \frac{\log |A|}{m}.$$

Applications. By replacing the $\text{LOCC}_1^{\parallel}$ (or Bell) measurements in [10] with measurements from LOCC_1 , we obtain a couple of interesting results as follows, for which technical proofs are given in the Appendix.

Detecting multipartite entanglement. Deciding whether a density matrix is entangled or separable is one of the most basic problem in quantum information theory, with both theoretical and practical significance [27]. Despite the existence of many entanglement criteria, up to date the only complete ones that detect all entangled states are infinite hierarchies [27]. Among them searching for symmetric extensions is probably the most useful [20]. This is exactly the scenario where quantum de Finetti theorems could be expected to be useful.

We consider the situation where a small error ϵ is permitted, meaning that we must detect all the entangled states except for those very weak ones that are ϵ -close to separable (at the same time all the separable states should be detected correctly). This is equivalently formulated as the Weak Membership Problem for separability: given a state $\rho_{A_1 A_2 \dots A_k}$ that is either separable or ϵ -away from any separable state, we want to decide which is the case. It has been shown that this problem is NP-hard when ϵ is of the order no larger than inverse polynomial of local dimensions (in trace norm) [28–30]. Surprisingly, Brandão, Christandl and Yard found a quasipolynomial-time algorithm for constant ϵ in one-way LOCC norm for bipartite states [16]. This algorithm was generalized to multipartite states in [19], then in [10] using a stronger method. These algorithms are all based on the searching for symmetric extensions of [20]. Along these lines, we present the following result, which is obtained by applying Theorem 1 to bound the distance between properly extendible states and separable states.

Corollary 4 *Testing multipartite entanglement of a state $\rho_{A_1 A_2 \dots A_k}$ with error ϵ can be done via searching for symmetric extensions in time*

$$\exp \left(c \left(\sum_{i=1}^k \log |A_i| \right)^2 k^2 f(\epsilon) \right), \quad (10)$$

where $f(\epsilon) = \epsilon^{-2}$ if the error is measured by the norm $\|\cdot\|_{\text{LOCC}_1}$ and $f(\epsilon) = \epsilon^{-1}$ if it is measured by the relative entropy D_{LOCC_1} .

It is worth mentioning that the run time in Eq. (10) is quasipolynomial, for constant particle number k and constant error ϵ . The algorithm in [19] using LOCC_1 -norm behaves exponentially slower than ours with respect to the number of particles k , while the algorithm of [10] has the same run-time as ours but works only for $\text{LOCC}_1^{\parallel}$ -norm rather than our LOCC_1 -norm approximation. Thus our result has bridged the gap between these two works. Furthermore, here for the first time we catch the importance of the *amount of entanglement* in this problem. The quantity $E_r^{\text{LOCC}_1}(\rho) := \min\{D_{\text{LOCC}_1}(\rho\|\sigma) : \sigma \text{ being separable}\}$, introduced in [31], is asymptotically normalized since $E_r^{\text{LOCC}_1}(\Phi_d) = \log(d+1) - 1$ for maximally entangled state Φ_d of local dimension d [32]. Corollary 4 shows that, detecting all the k -partite entangled states ρ such that $E_r^{\text{LOCC}_1}(\rho) \geq \epsilon$ can be done in quasipolynomial time in local dimensions. This is a stronger statement than using LOCC_1 -norm as the error measure. We point out that for the bipartite case this result can also be obtained by combining the algorithm of [16] with the “commensurate lower bound” for squashed entanglement of [32].

QMA proof system with multiple proofs. QMA, the quantum analogue of the complexity class NP, is the set of

decision problems whose solutions can be efficiently verified on a quantum computer, provided with a polynomial-size quantum proof [33]. In recent years there have been significant advances on the structure of QMA systems, where multiple *unentangled* proofs and possibly locally restricted measurements in the verification were considered [10, 16, 34–36]. It has been proven that many natural problems in quantum physics are characterized by QMA proof systems (see, e.g., [21, 22, 37, 38]).

To solve a problem, the verifier performs a quantum algorithm on the input $x \in \{0, 1\}^n$ along with the quantum proofs. The algorithm then returns “yes” or “no” as the answer to the instance x . This procedure of verification can be effectively described as a set of two-outcome measurements $\{(M_x, \mathbb{1} - M_x)\}_x$ on the proofs. In the definition below, a problem is formally identified with a “language”.

Definition 5 *A language L is in $\text{QMA}^M(k)_{m,c,s}$ if there exists a polynomial-time implementable verification $\{(M_x, \mathbb{1} - M_x)\}_x$ with each measurement from the class M such that*

- *Completeness:* If $x \in L$, there exist k states as proofs $\omega_1, \dots, \omega_k$, each of size m qubits, such that

$$\text{Tr}(M_x(\omega_1 \otimes \dots \otimes \omega_k)) \geq c.$$

- *Soundness:* If $x \notin L$, then for any $\omega_1, \dots, \omega_k$,

$$\text{Tr}(M_x(\omega_1 \otimes \dots \otimes \omega_k)) \leq s.$$

We are also interested in QMA systems with multiple symmetric proofs. $\text{SymQMA}^M(k)_{m,c,s}$ is defined in a similar way but here we replace independent proofs $\omega_1, \dots, \omega_k$ with identical ones $\omega^{\otimes k}$ in both completeness and soundness parts. As a convention, we set M to be ALL (the class of all measurements), $m = \text{poly}(n)$, $k = 1$, $c = 2/3$ and $s = 1/3$ as defaults [40]. We can now state our application of Theorem 1 to these complexity classes.

Corollary 6 *We have*

$$\text{QMA} = \text{QMA}^{\text{LOCC}_1}(\text{poly}) = \text{SymQMA}^{\text{LOCC}_1}(\text{poly}). \quad (11)$$

In particular,

$$\text{SymQMA}^{\text{LOCC}_1}(k)_{m,c,s} \subseteq \text{QMA}_{0.6m^2k^2\epsilon^{-2}, c, s+\epsilon}, \quad (12)$$

$$\text{QMA}^{\text{LOCC}_1}(k)_{m,c,s} \subseteq \text{QMA}_{0.6m^2k^4\epsilon^{-2}, c, s+\epsilon} \quad (13)$$

It has been proven in [16] that $\text{QMA} = \text{QMA}^{\text{LOCC}_1}(k)$ for constant k . Our result generalizes this statement to a polynomial number of proofs. It is also a generalization of the results in [10, 41] which prove the reduction of $\text{QMA}^{\text{LO}}(k)$ to QMA (LO denotes local measurements). On the other hand, Ref. [10] proved that, assuming ETH (exponential time hypothesis for 3-SAT) [42], any multi-prover QMA protocol with symmetric proofs and Bell verification for 3-SAT, can not bring better than the

square-root reduction of [43] to the proof size. Eq. (12) implies that, this is still true even if *adaptively* local verification (one-way LOCC measurement) is permitted.

Arguably the biggest open question in the study of QMA proof systems is whether $\text{QMA} = \text{QMA}(2)$ (note that Harrow and Montanaro have proved that $\text{QMA}(2) = \text{QMA}(k)$ for any polynomial $k > 2$ [36]). On the one hand, there are natural problems from quantum physics that are in $\text{QMA}(2)$ but not obviously in QMA [21, 22, 38]. On the other hand, Harrow and Montanaro showed that if the first equality in Eq. (11) holds for a kind of separable measurements (even only for the case of two proofs), then $\text{QMA} = \text{QMA}(2)$. Our result here, although does not touch this open question directly, is a step towards a larger measurement class compared to [10] and we hope it will stimulate future progress in solving this open question.

Polynomial optimization over hyperspheres. Theorem 1 also gives some improved results on the usefulness of a general relaxation method, called the Sum-of-Squares (SOS) hierarchy [44, 45], for polynomial optimization over hyperspheres (see, e.g., [10, 46]). The relevance in physics is that pure states of a quantum system form exactly a hypersphere and hence some computational problems in quantum physics are indeed to optimize a polynomial over hyperspheres. See Appendix for the details.

Discussions. The advantage of our method, inherited from [10], is that it tells us more information than that of [16, 32] about the valid de Finetti (separable) state that approximates the symmetric (extendible) state. As a result, we obtain a huge improvement over [19] on the particle-number dependence, and we are able to strengthen the relation $\text{QMA} = \text{QMA}^{\text{LOCC}_1}(k)$ from the constant k of [16] to polynomial k . We hope that the de Finetti theorem presented in this letter will find more applications in the future.

We ask whether Theorem 1 can be further improved, to work for two-way LOCC or even separable measurements. This would accordingly give stronger applications, and possibly, solve the QMA vs $\text{QMA}(2)$ puzzle due to the result of [36]. Another open question is, for a state supported on the symmetric subspace (aka Bose-symmetric state), whether its reduced states have pure-state approximations of the form $\int \varphi^{\otimes k} d\mu(\varphi)$ with φ *pure* that are not worse than the mixed-state approximations given by our theorem. We notice that this is indeed the case for the de Finetti theorem of [8] and a similar statement holds for [9]. However, our method, as well as that of [10] seems to require that the state φ must be generally mixed.

Acknowledgements. KL is supported by NSF Grant CCF-1110941 and CCF-1111382. GS acknowledges NSF Grant CCF-1110941. We thank Charles Bennett, Fernando Brandão, Aram Harrow and John Smolin for inter-

esting discussions, and the anonymous referees for helping improve the manuscript.

* Electronic address: carl.ke.lee@gmail.com

† Electronic address: gsbsmith@gmail.com

- [1] B. de Finetti, Ann. Inst. H. Poincaré **7**, 1 (1937).
- [2] P. Diaconis and D. Freedman, The Annals of Probability **8**, 745 (1980).
- [3] E. Størmer, J. Funct. Anal. **3**, 48 (1969).
- [4] R. L. Hudson and G. R. Moody, Z. Wahrsch. Verw. Geb. **33**, 343 (1976).
- [5] G. A. Raggio and R. F. Werner, Helv. Phys. Acta **62**, 980 (1989).
- [6] C. M. Caves, C. A. Fuchs and R. Schack, J. Math. Phys. **43**, 4537 (2002).
- [7] R. König and R. Renner, J. Math. Phys. **46**, 122108 (2005).
- [8] M. Christandl, R. König, G. Mitchison and R. Renner, Commun. Math. Phys. **273**, 473 (2007).
- [9] R. Renner, PhD thesis, ETHZ, Zurich (2005), arXiv:quant-ph/0512258; Nature Phys. **3**, 645 (2007).
- [10] F. G. S. L. Brandão and A. W. Harrow, in Proc. of the 45th ACM Symposium on theory of computing (STOC 2013), pp. 861-870 (2013), arXiv:1210.6367.
- [11] M. Fannes and C. Vandenplas, J. Phys. A **39**, 13843 (2006).
- [12] M. Lewin, P. T. Nam and N. Rougerie, arXiv:1303.0981.
- [13] F. G. S. L. Brandão and M. B. Plenio, Comm. Math. Phys. **295**, 791 (2010).
- [14] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).
- [15] S. Beigi, P. Shor and J. Watrous, Theory of Computing **7**, 101 (2011).
- [16] F. G. S. L. Brandão, M. Christandl and J. Yard, Comm. Math. Phys. **306**, 805 (2011); Proc. of the 43rd ACM Symposium on theory of computing (STOC 2011), pp. 343-352 (2011); arXiv:1010.1750.
- [17] The exchange of two systems A_i and A_j causes a unitary transformation $U_{ij}|\phi_{A_i}\rangle|\phi_{A_j}\rangle = |\phi_{A_i}\rangle|\phi_{A_j}\rangle$ on their state. We say $\rho_{A_1\dots A_n}$ is permutation-invariant if $U_{ij}\rho_{A_1\dots A_n}U_{ij}^\dagger = \rho_{A_1\dots A_n}$ for any $0 < i < j \leq n$.
- [18] In the present paper, \ln and \log are logarithms with base e and 2, respectively.
- [19] F. G. S. L. Brandão and M. Christandl, Phys. Rev. Lett. **109**, 160502 (2012).
- [20] A. C. Doherty, P. A. Parrilo and F. M. Spedalieri, Phys. Rev. Lett. **88**, 187904 (2002); Phys. Rev. A **69**, 022308 (2004); Phys. Rev. A **71**, 032333 (2005).
- [21] Y.-K. Liu, M. Christandl and F. Verstraete, Phys. Rev. Lett. **98**, 110503 (2007).
- [22] A. Chailloux and O. Sattath, in Proc. of IEEE 27th Annual Conference on Computational Complexity (CCC), pp 32 – 41 (2012).
- [23] W. Matthews, S. Wehner and A. Winter, Comm. Math. Phys. **291**, 813 (2009).
- [24] G. Aubrun and C. Lancien, arXiv:1406.1959.
- [25] C. A. Fuchs and J. van de Graaf, IEEE. Tran. Inf. Theory **45**, 1216 (1999).
- [26] G. Lindblad, Comm. Math. Phys. **40**, 147 (1975); A. Uhlmann, Comm. Math. Phys. **54**, 21 (1977).

- [27] R. Horodecki, P. Horodecki, M. Horodecki and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [28] L. Gurvits, in *Proc. of the 35th ACM Symposium on theory of computing (STOC 2003)*, pp. 10–19 (2003).
- [29] S. Gharibian, *Quantum Inf. Comput.* **10**, 343 (2010).
- [30] S. Beigi, *Quantum Inf. Comput.* **10**, 141 (2010).
- [31] M. Piani, *Phys. Rev. Lett.* **103**, 160504 (2009).
- [32] K. Li and A. Winter, *Comm. Math. Phys.* **326**, 63 (2014).
- [33] J. Watrous, in *Encyclopedia of Complexity and System Science* (Springer, 2009).
- [34] H. Kobayashi, K. Matsumoto and T. Yamakami, in *Proc. of the 14th Annual International Symposium on Algorithms and Computation*, pp. 189–198 (2003), arXiv:quant-ph/0306051.
- [35] S. Aaronson, R. Impagliazzo, D. Moshkovitz and P. Shor, *Theory of Computing* **5**, 1 (2009).
- [36] A. W. Harrow and A. Montanaro, *J. ACM* **60** (1), article 3 (2013); earlier version in *Proc. of the IEEE 51st Symp. on Found. of Comp. Sci. (FOCS 2010)*, pp. 633–642 (2010).
- [37] J. Kempe, A. Kitaev and O. Regev, *SIAM J. Comput.* **35**, 1070 (2006); N. Schuch and F. Verstraete, *Nature Phys.* **5**, 732 (2009); T.-C. Wei, M. Mosca and A. Nayak, *Phys. Rev. Lett.* **104**, 040501 (2010).
- [38] G. Gutoski, P. Hayden, K. Milner and M. Wilde, arXiv:1308.5788.
- [39] C. Marriott and J. Watrous, *Computational Complexity* **14**, 122 (2005).
- [40] Actually the values of c and s can be chosen arbitrarily as long as $c - s \geq 1/\text{poly}(n)$, because this gap can be amplified to have exponentially small errors. See [35, 39] for the amplification of $\text{QMA}^{\text{LOCC}_1}(k)$ and QMA . The amplification of $\text{SymQMA}^{\text{LOCC}_1}(k)$ follows from Eq. (12) together with the amplification of QMA .
- [41] F. G. S. L. Brandão, PhD thesis, Imperial College, London (2008), arXiv:0810.0026.
- [42] R. Impagliazzo, R. Paturi and F. Zane, in *Proc. of the IEEE 39th Symp. on Found. of Comp. Sci. (FOCS 1998)*, pp. 653–662 (1998).
- [43] J. Chen and A. Drucker, arXiv:1011.0716.
- [44] J. B. Lasserre, *SIAM J. Opt.* **11**, 796 (2001).
- [45] P. A. Parrilo, PhD thesis, MIT, 2000.
- [46] B. Barak, J. Kelner and D. Steurer, in *Proc. of the 46th ACM Symposium on theory of computing (STOC 2014)*, pp. 31–40 (2014), arXiv:1312.6652.

Appendix

Inequivalence of $\|\cdot\|_{\text{LOCC}_1}$ and $\|\cdot\|_{\text{LOCC}_1^\bullet}$. Here we show the following: for all d there are constant C and $d \times d \times 2$ states ρ_{ABC} and σ_{ABC} such that $\|\rho_{ABC} - \sigma_{ABC}\|_{\text{LOCC}_1} = 2$ but $\|\rho_{ABC} - \sigma_{ABC}\|_{\text{LOCC}_1^\bullet} \leq C/\sqrt{d}$.

To see this, notice that for states of the form $\rho_{ABC} = \rho_{AB} \otimes |0\rangle\langle 0|$ and $\sigma_{ABC} = \sigma_{AB} \otimes |0\rangle\langle 0|$ we have

$$\begin{aligned}\|\rho_{ABC} - \sigma_{ABC}\|_{\text{LOCC}_1} &= \|\rho_{AB} - \sigma_{AB}\|_{\text{LOCC}_1}, \\ \|\rho_{ABC} - \sigma_{ABC}\|_{\text{LOCC}_1^\bullet} &= \|\rho_{AB} - \sigma_{AB}\|_{\text{LO}},\end{aligned}$$

where LO denotes the set of local measurements. We can then apply the existence of bipartite states with $\|\rho_{AB} - \sigma_{AB}\|_{\text{LOCC}_1} = 2$ and $\|\rho_{AB} - \sigma_{AB}\|_{\text{LO}} \leq C/\sqrt{d}$ as shown in Theorem 2.4 of [24].

Proof of Lemma 3. Let $\rho^{AZ_1 \dots Z_m} := \text{id}^A \otimes \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_m(\rho^{AB_1 \dots B_m})$, with $\mathcal{M}_\ell^{B_\ell \rightarrow Z_\ell}$ being measurement operations. Due to the chain rule of mutual information,

$$I(A; Z_1 \dots Z_m) = I(A; Z_1) + I(A; Z_2 | Z_1) + \dots + I(A; Z_m | Z_1 \dots Z_{m-1}). \quad (14)$$

Now we fix a special choice of \mathcal{M}_ℓ 's. Let \mathcal{M}_1 maximize $I(A; Z_1)$. Then under this choice of \mathcal{M}_1 , we choose \mathcal{M}_2 that maximizes $I(A; Z_2 | Z_1)$. Repeat this procedure and at last we pick \mathcal{M}_m that maximizes $I(A; Z_m | Z_1 \dots Z_{m-1})$ under the previously fixed measurements $\mathcal{M}_1, \dots, \mathcal{M}_{m-1}$. As a result, for each conditional mutual information we have

$$\begin{aligned}I(A; Z_\ell | Z_1 \dots Z_{\ell-1}) \\ \geq \min_{\mathcal{Q}'} \max_{\mathcal{P}'} I(A; Y_\ell | X_\ell)_{\text{id}^A \otimes \mathcal{P}' \otimes \mathcal{Q}'(\rho^{AB_1 \dots B_\ell})},\end{aligned} \quad (15)$$

where $\mathcal{P}' : B_\ell \rightarrow Y_\ell$ and $\mathcal{Q}' : B_1 \dots B_{\ell-1} \rightarrow X_\ell$ are measurement operations. Further relax the minimization to allow the measurement \mathcal{Q}' to be performed on all the B systems except for B_ℓ . Then we can set ℓ to be 1 without changing the value due to the symmetry of the state. Thus Eq. (15) is further lower bounded by

$$\min_{\mathcal{Q}} \max_{\mathcal{P}} I(A; Y | X)_{\text{id}^A \otimes \mathcal{P} \otimes \mathcal{Q}(\rho^{AB_1 \dots B_m})}$$

with measurement operations $\mathcal{P} : B_1 \rightarrow Y$ and $\mathcal{Q} : B_2 \dots B_m \rightarrow X$. This, combined with Eq. (14) lets us conclude that

$$\begin{aligned}\log |A| &\geq I(A; Z_1 \dots Z_m) \\ &\geq m \min_{\mathcal{Q}} \max_{\mathcal{P}} I(A; Y | X)_{\text{id}^A \otimes \mathcal{P} \otimes \mathcal{Q}(\rho^{AB_1 \dots B_m})},\end{aligned}$$

and we are done. \square

Proof of Corollary 4. We first prove this result with the error measured by the fully one-way LOCC norm. Then we explain that a slight adaptation works for the case of relative entropy.

Let $\ell \geq k$ be an integer. Introduce quantum systems $\bar{A}_i := A_1^i A_2^i \dots A_k^i$ with $A_j^i \cong A_j$, for all $0 < i \leq \ell$ and $0 < j \leq k$. We search for a state $\tilde{\rho}_{\bar{A}_1 \dots \bar{A}_\ell}$ such that $\tilde{\rho}_{\bar{A}_1 \dots \bar{A}_\ell}$ is permutation-invariant and $\tilde{\rho}_{A_1^1 A_2^2 \dots A_k^k} = \rho_{A_1 A_2 \dots A_k}$. If such a state exists, we feed back “separable”. Otherwise we conclude that it is entangled. For our purpose we set $\ell = 2(k-1)^2 \epsilon^{-2} \sum_i \ln |A_i| + k$.

This search can be done using semidefinite programming in time polynomial of the total dimension of $\tilde{\rho}_{\bar{A}_1 \dots \bar{A}_\ell}$, which coincides with the claimed result.

To analyze the correctness, first assume that such an extension exists. Then we apply Theorem 1 to see that there is certain probability measure μ such that

$$\left\| \tilde{\rho}_{\bar{A}_1 \dots \bar{A}_\ell} - \int \sigma^{\otimes k} d\mu(\sigma) \right\|_{\text{LOCC}_1} \leq \epsilon.$$

By definition, if we restrict the measurement to be performed only on systems $A_1^1, A_2^2, \dots, A_k^k$ then the above inequality implies that there exists a separable state $\sigma_{A_1 \dots A_k}$ such that $\|\rho_{A_1 \dots A_k} - \sigma_{A_1 \dots A_k}\|_{\text{LOCC}_1} \leq \epsilon$. So if $\rho_{A_1 \dots A_k}$ is ϵ -away from any separable state, the required extension can not exist. But if $\rho_{A_1 \dots A_k}$ is separable, it is obvious that such an extension does exist. As a result in both cases the above procedure works correctly.

The above argument works as well if the error is measured by the relative entropy. The small modification needed is just to replace $\|\cdot\|_{\text{LOCC}_1}$ by D_{LOCC_1} and here we set $\ell = (k-1)^2 \epsilon^{-1} \sum_i \log |A_i| + k$. \square

Closeness of extendible states to being separable.

We also show how an extendible multipartite state is close to the set of separable states, under fully one-way LOCC distinguishability. A state $\rho_{A_1 \dots A_k}$ is ℓ -extendible if there is an extension $\tilde{\rho}_{\bar{A}_1 \dots \bar{A}_\ell}$ with $\bar{A}_i := A_1^i A_2^i \dots A_k^i$, such that for all $0 < j \leq k$ the state $\tilde{\rho}_{\bar{A}_1 \dots \bar{A}_\ell}$ is invariant under any permutations over subsystems $A_j^1, A_j^2, \dots, A_j^\ell$ and for any $0 < i_1, \dots, i_k \leq \ell$ we have $\rho_{A_1 \dots A_k} = \tilde{\rho}_{A_1^{i_1} \dots A_k^{i_k}}$. Obviously $\tilde{\rho}_{\bar{A}_1 \dots \bar{A}_\ell}$ is permutation-invariant and $\tilde{\rho}_{A_1^1 A_2^1 \dots A_k^1} = \rho_{A_1 A_2 \dots A_k}$. So similar to the argument in the proof of Corollary 4, a use of Theorem 1 lets us obtain:

$$E_r^{\text{LOCC}_1}(\rho_{A_1 \dots A_k}) \leq \frac{(k-1)^2 \sum_i \log |A_i|}{\ell - k},$$

$$\min_{\sigma \in \text{SEP}} \|\rho_{A_1 \dots A_k} - \sigma_{A_1 \dots A_k}\|_{\text{LOCC}_1} \leq \sqrt{\frac{2(k-1)^2 \sum_i \ln |A_i|}{\ell - k}}$$

holds for any ℓ -extendible state $\rho_{A_1 \dots A_k}$. Here SEP denotes the set of all separable states.

Proof of Corollary 6. Restricting the verification to be performed on the first proof in the multi-prover protocols, we see that

$$\text{QMA}_{m,c,s} \subseteq \text{SymQMA}^{\text{LOCC}_1}(k)_{m,c,s}, \quad (16)$$

$$\text{QMA}_{m,c,s} \subseteq \text{QMA}^{\text{LOCC}_1}(k)_{m,c,s}. \quad (17)$$

By definition, Eq. (16) and Eq. (12) imply $\text{SymQMA}^{\text{LOCC}_1}(\text{poly}) = \text{QMA}$. Similarly, Eq. (17) and Eq. (13) imply $\text{QMA}^{\text{LOCC}_1}(\text{poly}) = \text{QMA}$. Note that we can use the amplification of $\text{QMA}_{m,c,s}$ (see [39]) to keep $c = 2/3$ and $s = 1/3$.

To prove Eq. (12), we show a way of simulating a $\text{SymQMA}^{\text{LOCC}_1}(k)_{m,c,s}$ protocol in a single-proof QMA system. The prover provides the verifier with a proof of size $0.6m^2k^2\epsilon^{-2}$, which consists of $\ell = 0.6mk^2\epsilon^{-2}$ subsystems each of size m qubits. Then the verifier makes a uniformly random permutation over the subsystems and then performs the $\text{SymQMA}^{\text{LOCC}_1}(k)_{m,c,s}$ verification on the first k (denoted as A_1, A_2, \dots, A_k) of them. No matter what the initial state ρ_{initial} of the proof is, Theorem 1 implies that the state on $A_1 \dots A_k$, $\rho_{A_1 \dots A_k}$, can be approximated as $\|\rho_{A_1 \dots A_k} - \int \sigma^{\otimes k} d\mu(\sigma)\|_{\text{LOCC}_1} \leq 2\epsilon$

with certain probability measure μ . Let $\{(M_x, \mathbb{1} - M_x)\}_x$ be the one-way LOCC measurements in the $\text{SymQMA}^{\text{LOCC}_1}(k)_{m,c,s}$ protocol for a language L . Then the soundness constant s' in this simulation can be

$$\begin{aligned} s' &= \max_{x \notin L} \max_{\rho_{\text{initial}}} \text{Tr } M_x \rho_{A_1 \dots A_k} \\ &\leq \max_{x \notin L} \max_{\mu} \int d\mu(\sigma) \text{Tr } M_x \sigma^{\otimes k} + \epsilon \\ &\leq s + \epsilon. \end{aligned}$$

One the other hand, suppose the proof for an accepted instance in $\text{SymQMA}^{\text{LOCC}_1}(k)_{m,c,s}$ is $\omega^{\otimes k}$. Then in the simulation the state $\omega^{\otimes \ell}$ gives the same probability of acceptance. So completeness does not change.

For Eq. (13), we will prove

$$\text{QMA}^{\text{LOCC}_1}(k)_{m,c,s} \subseteq \text{SymQMA}^{\text{LOCC}_1}(k)_{km,c,s}.$$

This, together with Eq. (12), leads to Eq. (13). The argument is similar to that in [35] (Lemma 38), where the same relation with “LOCC₁” replaced by “ALL” was proved. The strategy is to divide each proof in the $\text{SymQMA}^{\text{LOCC}_1}(k)_{km,c,s}$ system into k subsystems of m qubits, then simulate the $\text{QMA}^{\text{LOCC}_1}(k)_{m,c,s}$ protocol on the i th subsystem from the i th proof for all $i = 1, \dots, k$. \square

Polynomial optimization over hyperspheres. An immediate consequence of Theorem 1 is that we can enlarge in [10] the class of polynomials, for which the optimization over multiple hyperspheres admits efficient SOS approximation. We also provide another class of polynomials whose optimization over the single hypersphere has a similar feature, supplementing a result of [46] on polynomials with nonnegative coefficients.

We use a d -dimensional complex vector to encode $2d$ real variables.

Corollary 7 For $1 \leq i \leq k$, let A_i and A'_i be identical quantum systems of dimension d . Let $|\alpha_i\rangle \in \mathcal{H}_{A_i}$ and $|\beta\rangle^{\otimes k} \in \mathcal{H}_{A_1 A'_1} \otimes \dots \otimes \mathcal{H}_{A_k A'_k}$ be complex vectors. Let $0 \leq M \leq \mathbb{1}$ be a matrix on $\mathcal{H}_{A_1 \dots A_k}$ such that $\{M, \mathbb{1} - M\} \in \text{LOCC}_1$. The two optimizations

$$(\mathcal{O}1:) \max \langle \alpha_1 | \otimes \dots \otimes \langle \alpha_k | M | \alpha_1 \rangle \otimes \dots \otimes | \alpha_k \rangle$$

subject to $\langle \alpha_i | \alpha_i \rangle = 1, i = 1, \dots, k$

$$(\mathcal{O}2:) \max \langle \beta |^{\otimes k} (\mathbb{1} \otimes M) | \beta \rangle^{\otimes k} \text{ subject to } \langle \beta | \beta \rangle = 1$$

can be solved to within additive error ϵ efficiently, via a hierarchy of SDP relaxations (SOS), respectively in time $\exp(O(\epsilon^{-2}k^4 \log^2(d)))$ and $\exp(O(\epsilon^{-2}k^2 \log^2(d)))$.

The advantage of Corollary 7 is that, for constant ϵ and k , the runtime of these two optimizations is only quasi-polynomial of the number of variables, instead of exponential time of exhaustive search.

Proof. The analysis of $\mathcal{O}1$ is the same as that of the polynomial-optimization problem considered in [10]. We only need to employ our Theorem 1 when the de Finetti theorem is used.

It is easy to see that the maximum in $\mathcal{O}2$ equals

$$\max_{\sigma} \text{Tr } M \sigma^{\otimes k},$$

with σ a normalized quantum state. This, in turn, is bounded as

$$\begin{aligned} & \max_{\rho_{A_1 \dots A_\ell}} \text{Tr}(M \otimes \mathbb{1}) \rho_{A_1 \dots A_\ell} - \sqrt{\frac{(k-1)^2 \ln d}{2(\ell-k)}} \\ & \leq \max_{\sigma} \text{Tr } M \sigma^{\otimes k} \\ & \leq \max_{\rho_{A_1 \dots A_\ell}} \text{Tr}(M \otimes \mathbb{1}) \rho_{A_1 \dots A_\ell}, \end{aligned}$$

where $\ell \geq k$ and the maximization in the first and last

lines are over *permutation-invariant* state $\rho_{A_1 \dots A_\ell}$. Note that here the first inequality follows from a direct application of Theorem 1, and the second inequality is by restricting the maximization in the last line to over ℓ -fold states of the form $\sigma^{\otimes \ell}$. So the problem $\mathcal{O}2$ can be approximated to within additive error $\sqrt{\frac{(k-1)^2 \ln d}{2(\ell-k)}}$, by the lever- ℓ SDP hierarchy (SOS hierarchy)

$$\begin{aligned} & \text{maximize} \quad \text{Tr}(M \otimes \mathbb{1}) \rho_{A_1 \dots A_\ell} \\ & \text{subject to} \quad \rho_{A_1 \dots A_\ell} \geq 0, \text{Tr } \rho_{A_1 \dots A_\ell} = 1, \\ & \quad \rho_{A_1 \dots A_\ell} \text{ being permutation-invariant.} \end{aligned}$$

This can be done in time $\exp(O(\ell \log d))$, namely, polynomial of the dimension of $\rho_{A_1 \dots A_\ell}$. At last, to obtain the claimed result, we choose $\ell = \frac{1}{2}\epsilon^{-2}(k-1)^2 \ln d + k$. \square