Universal Superreplication of Unitary Gates

Article *in* Physical Review Letters · March 2015

DOI: 10.1103/PhysRevLett.114.120504 · Source: arXiv

3 authors:

G. Chiribella
University of Oxford
110 PUBLICATIONS **1,912** CITATIONS
SEE PROFILE

Yuxiang Yang
The University of Hong Kong
22 PUBLICATIONS **90** CITATIONS
SEE PROFILE

Cupjin Huang
University of Michigan
3 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Universal super-replication of unitary gates

G. Chiribella, Y. Yang and J. Huang
Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China

The manifold of all pure states of a quantum system satisfies an asymptotic no-cloning theorem, which forbids the replication of arbitrary identical and independently distributed (i.i.d.) sequences of states. In stark contrast, we show that arbitrary i.i.d. sequences of unitary gates can be replicated at a quadratic rate, with an error that can be made arbitrarily small on most inputs, except for an exponentially small fraction. By the same argument, we show that N parallel uses of a completely unknown unitary gate can be compressed into a single gate acting on $O(\log N)$ qubits, enabling an exponential saving of computational workspace.

A striking feature of quantum theory is the impossibility of constructing a universal copy machine, i. e. a machine that takes as input a quantum system in an arbitrary pure state and produces as output a number of exact replicas [1, 2]. Such an impossibility has important implications for quantum error correction [3, 4], key distribution [5, 6], and other cryptographic primitives, such as quantum money [7–10] and quantum secret sharing [11, 12].

The impossibility of universal copy machines is a hard fact: it equally affects deterministic [13–16] and probabilistic machines [17], these two types of machines having the same performances when it comes to copying completely unknown pure states [18, 19]. A similar no-go result holds when the universal machine is presented a large number $N \gg 1$ of identical copies and is required to produce a larger number $M > N$ of approximate replicas: if the replicas have non-vanishing overlap with the desired M-systems state, then the number of extra copies must be negligible compared to N [15, 20]. We refer to this fact as the *asymptotic no-cloning theorem* or *standard quantum limit for cloning* [19], expressing the fact that long i.i.d. sequences of pure states cannot be stretched by more than a linear factor.

The impossibility of universal cloning of pure states suggests the validity of similar results for arbitrary unitary gates. Along this line, a no-go theorem for exact universal gate cloning was proven in Ref. [21], showing that no quantum network can perfectly simulate two uses of an unknown unitary gate by querying it only once. On the other hand, very recently Dür and coauthors [22] considered a non-universal setup designed to approximately copy phase-gates, i. e. gates generated by time evolution with a known Hamiltonian. In this scenario, they came up with a quantum network that simulates with high fidelity up to N^2 uses of an unknown gate while using it only N times. This phenomenon of *gate super-replication* is an analog of the super-replication of phase-states discovered in Ref. [19], with the crucial difference that state replication has necessarily an exponentially small probability of success while the gate replication takes place deterministically. The main open question raised by Ref. [22] is whether deterministic super-replication occurs not only for the subset of phase gates, but also for arbitrary unitary gates. An affirmative answer would imply that the asymptotic no-cloning theorem only applies to states, whereas it is possible to stretch long i.i.d. sequences of reversible quantum gates by up to a quadratic factor.

In this letter we answer the question of Ref. [22] in the affirmative, establishing the possibility of universal super-replication of unitary gates, in stark contrast with the asymptotic no-cloning theorem for pure states. Given N uses of a completely unknown unitary gate U, we construct a quantum network that simulates up to $M = O(N^2)$ parallel uses of U, providing an output that is close to the ideal target for all possible input states except for an exponentially small fraction. The network is asymptotically optimal: every other network that produces replicas at a rate higher than quadratic will necessarily spoil their quality, delivering an output that has vanishing overlap with the desired output. In addition to cloning, we consider the task of gate compression, where one is given a black box and is asked to faithfully encode its action in a gate acting on a smaller quantum system. For gates acting on two-dimensional quantum systems (qubits), we show that N uses of a completely unknown gate can be encoded without any loss in a single gate acting only on $2\log N$ qubits. This exponential compression of workspace is the analogue of the exponential compression of i.i.d. sequences of quantum states [23], recently demonstrated experimentally with photons [24].

Universal super-replication of qubit gates. Let us start from the simplest case of qubit gates, represented by unitary matrices in $\mathsf{SU}(2)$. A generic gate can be represented as $U_{\varphi,\mathbf{n}} = \exp[-i\varphi\,\mathbf{n}\cdot\mathbf{j}]$, where $\varphi \in [0, 2\pi)$ is a rotation angle, $\mathbf{n} = (n_x, n_y, n_z)$ is a rotation axis, and $\mathbf{j} = (j_x, j_y, j_z)$ is the vector of angular momentum operators. We define $g := (\varphi, \mathbf{n})$ and label the unitary gate as U_g. Here both φ and $\mathbf{n}$ are completely unknown, differing from the setting of [22], where the rotation axis was fixed and only φ was varying. In order to replicate an unknown gate, we consider a network where N parallel uses of U_g are sandwiched between two quantum channels, $\mathcal{C}_1$ and $\mathcal{C}_2$, as in figure 1. The overall action of the network is de-

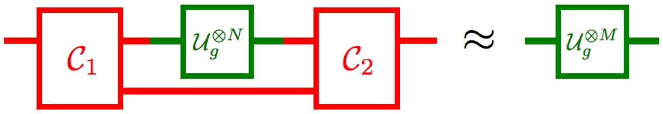

FIG. 1. **Quantum network for gate replication.** The network simulates M parallel uses of an unknown unitary gate $\mathcal{U}_g$, while querying it only N times. The simulation is obtained by transforming the input state of M systems into the joint state of N systems and an ancilla (quantum channel $\mathcal{C}_1$), applying the unknown gate on the N systems, and then recombining them with the ancilla via a quantum channel $\mathcal{C}_2$, which finally produces M output systems.

scribed by the channel $\mathcal{C}_g := \mathcal{C}_2\left(\mathcal{U}_g^{\otimes N} \otimes \mathcal{I}_A\right)\mathcal{C}_1$, where $\mathcal{U}_g$ is the channel $\mathcal{U}_g(\cdot) = U_g \cdot U_g^\dagger$ and $\mathcal{I}_A$ is the identity on a suitable ancilla.

In order to specify the channels $\mathcal{C}_1$ and $\mathcal{C}_2$, it is useful to decompose the Hilbert space of K qubits ($K = N, M$) into rotationally invariant subspaces. Choosing K to be even, we have

$$\mathscr{H}^{\otimes K} \simeq \bigoplus_{j=0}^{K/2} \left(\mathscr{R}_j^{(K)} \otimes \mathscr{M}_j^{(K)}\right) , \qquad (1)$$

where j is the quantum number of the total angular momentum, $\mathscr{R}_j^{(K)}$ is a representation space, of dimension $d_j = 2j+1$, and $\mathscr{M}_j^{(K)}$ is a multiplicity subspace, of dimension

$$m_j^{(K)} = \frac{2j+1}{K/2+j+1}\binom{K}{K/2+j} \qquad (2)$$

(see e.g. [25]). The isomorphism in Eq. (1) is called the *Schur transform* and can be implemented efficiently by a suitable quantum circuit [26]. Let us denote by $P_J^{(K)}$ the projector on the direct sum of all subspaces with angular momentum number smaller than J. Setting $K = M$, we define the encoding operation

$$\mathcal{E}_J^{(M)}(\rho) = P_J^{(M)} \rho P_J^{(M)} + \operatorname{Tr}\left[\left(I^{\otimes M} - P_J^{(M)}\right)\rho\right]\rho_0 \quad (3)$$

where ρ_0 is a fixed state, viz. $\rho_0 = P_0^{(M)}/\operatorname{Tr}[P_0^{(M)}]$. Denoting by $F_\Psi^{(M,J)}$ be the fidelity between a generic pure state $|\Psi\rangle \in \mathscr{H}^{\otimes M}$ and the state $\mathcal{E}_J^{(M)}(|\Psi\rangle\langle\Psi|)$, we have the following

Theorem 1. *If $|\Psi\rangle$ is chosen uniformly at random, then, for every fixed $\epsilon > 0$, the probability that $F_\Psi^{(M,J)}$ is smaller than $1-\epsilon$ satisfies the bound*

$$\mathsf{Prob}\left[F_\Psi^{(M,J)} < 1-\epsilon\right] < \frac{4\left(1-2^{-M}\right)e^{-\frac{2J^2}{M}}}{\epsilon} , \qquad (4)$$

up to a multiplicative correction of order $O(J/\sqrt{M})$.

Proof. By Markov's inequality, one has $\mathsf{Prob}\left[F_\Psi^{(M,J)} < 1-\epsilon\right] < (1 - \mathbb{E}\left[F^{(M,J)}\right])/\epsilon$, where $\mathbb{E}\left[F^{(M,J)}\right]$ is the average of the fidelity over all pure states. In turn, the average fidelity can be expressed as $\mathbb{E}\left[F^{(M,J)}\right] = \left[2^M F_E^{(M,J)} + 1\right]/(2^M+1)$, where $F_E^{(M,J)}$ is the entanglement fidelity [27, 28]. Finally, the entanglement fidelity satisfies the bound

$$\begin{aligned} F_E^{(M,J)} &\geq \left|\langle\Phi|P_J^{(M)}|\Phi\rangle\right|^2 = \left[\sum_{j=0}^{J} \frac{d_j m_j^{(M)}}{2^M}\right]^2 \\ &\equiv \left[\sum_{j=0}^{J} p_j^{(M)}\right]^2 , \end{aligned} \qquad (5)$$

where the probability distribution $p_j^{(M)} := d_j m_j^{(M)}/2^M$ is known as the *Schur-Weyl measure* (see e.g. [29, 30]). Now, for large M the Schur-Weyl measure is concentrated on an interval of size $\sqrt{M}$ around $j = 0$. Explicit calculation (cf. Appendix A.1) then gives the bound $F_E^{(M,J)} \geq 1 - 4\,e^{-\frac{2J^2}{M}}\left[1 + O\left(J/\sqrt{M}\right)\right]$. The desired result follows by inserting this bound into the expression of the average fidelity and using Markov's inequality. □

Theorem 1 guarantees that, except for an exponentially small fraction of the Hilbert space, almost all pure M-qubit states are approximately in the subspace $\mathscr{H}_J^{(M)} = P_J^{(M)}\mathscr{H}^{\otimes M}$, provided that J is large compared to $\sqrt{M}$. To achieve super-replication of qubit gates, we combine this fact with another observation, namely that for $J \leq N/2$, the states in $\mathscr{H}_J^{(M)}$ can be faithfully encoded into $\mathscr{H}^{\otimes N} \otimes \mathscr{H}_A$, where $\mathscr{H}_A$ the Hilbert space of a suitable ancilla. The encoding is achieved by an isometry $V_J : \mathscr{H}_J^{(M)} \to \mathscr{H}^{\otimes N} \otimes \mathscr{H}_A$ that commutes with all rotations, namely

$$V_J\, U_g^{\otimes M} = \left(U_g^{\otimes N} \otimes I_A\right) V_J \qquad \forall U_g \in \mathsf{SU}(2) . \qquad (6)$$

In order for this to be possible, the dimension of the ancilla must satisfy the condition $m_l^{(N)} d_A \geq m_l^{(M)}$ for all $l \leq J$, and therefore has the minimum value

$$\begin{aligned} d_{\min} &= \frac{(N/2+J+1)\binom{M}{M/2+J}}{(M/2+J+1)\binom{N}{N/2+J}} \\ &\approx 2^{M-N}\sqrt{\frac{N}{M}}\,\frac{N/2+J+1}{M/2+J+1}\,e^{-\frac{2(M-N)J^2}{MN}} , \quad M, N \gg 1 \end{aligned} \qquad (7)$$

We are now ready to specify the channels $\mathcal{C}_1$ and $\mathcal{C}_2$ in the gate replication network. For channel $\mathcal{C}_1$, we choose $\mathcal{C}_1 = \mathcal{V}_J \mathcal{E}_J^{(M)}$, where $\mathcal{V}_J$ is the isometric channel $\mathcal{V}_J(\cdot) = V_J \cdot V_J^\dagger$ and J is given by $J = \min\left\{\lfloor\sqrt{M^{1+\delta}}\rfloor, N/2\right\}$ for an arbitrary small $\delta > 0$. For channel $\mathcal{C}_2$, we choose

$$\mathcal{C}_2(\rho) = V_J^\dagger \rho V_J + \operatorname{Tr}\left[\left(I^{\otimes N} - V_J V_J^\dagger\right)\rho\right]\rho_0 . \qquad (8)$$

The action of the network on a generic M-qubit state U_g is then given by

$$\begin{aligned}\mathcal{C}_g(|\Psi\rangle\langle\Psi|) &= \mathcal{C}_2(\mathcal{U}_g^{\otimes N}\otimes\mathcal{I}_A)\mathcal{C}_1(|\Psi\rangle\langle\Psi|)\\ &= \mathcal{U}_g^{\otimes M}\mathcal{C}_2\mathcal{C}_1(|\Psi\rangle\langle\Psi|)\\ &= \mathcal{U}_g^{\otimes M}\mathcal{E}_J^{(M)}(|\Psi\rangle\langle\Psi|)\,,\end{aligned}$$

having used Eq. (6) in the first equality. Clearly, this implies that the fidelity between the output state and the ideal target $U_g^{\otimes M}|\Psi\rangle$ is equal to the fidelity between $\mathcal{E}_J^{(M)}(|\Psi\rangle\langle\Psi|)$ and $|\Psi\rangle\langle\Psi|$, which is equal to $F_\Psi^{(M,J)}$ and is arbitrarily close to one with high probability whenever M is small compared to N^2, by virtue of theorem 1.

In summary, given N copies of a completely unknown gate, our network simulates the action of $M \ll N^2$ copies one a large fraction of the input states, provided that M is small compared to N^2. Note that the simulation works with probability exponentially close to 1, but can fail on some specific inputs: notably, it fails for all the inputs of the i.i.d. form $|\varphi\rangle^{\otimes M}$, for which the fidelity with the desired output state is *zero*. This fact is in complete analogy with other phenomena based on measure concentration, such as equilibration in large systems [31] and entanglement typicality [32]. Among other things, the fact that our network does not work on i.i.d. states implies that it cannot be used to circumvent the asymptotic no-cloning theorem for pure states: even if one had at disposal N uses of a gate U_ψ that prepares the state $|\psi\rangle$ from a fixed state $|0\rangle$, one would still be unable to produce $M \gg N$ copies of $|\psi\rangle$ by gate super-replication. In this, universal super-replication differs remarkably from the super-replication of phase-gates [22], in that the latter allows one to deterministically generate up to N^2 copies of a phase-state from N uses of the corresponding gate.

Entanglement replication. Mathematically there is a one-to-one correspondence between qubit gates U_g and two-qubit maximally entangled states $|\Psi_g\rangle$, set up by the Choi isomorphism [33, 34]. Based on it, the universal super-replication of quantum gates may suggest that one could transform N copies of an unknown maximally entangled state into up to N^2 almost perfect copies. This intuition, however, is faulty. Since the set of all maximally entangled states contains the one-parameter family $|\Phi_t\rangle = (|0\rangle|0\rangle + e^{i\varphi}|1\rangle|1\rangle)/\sqrt{2}$, $\varphi\in[0,2\pi)$ a deterministic super-replication of maximally entangled states would lead to a violation of the standard quantum limit for phase estimation [19, 35]. As a matter of fact, the fidelity of the optimal deterministic cloning machine transforming N perfect copies into M approximate replicas scales like $(N/M)^{3/2}$ for asymptotically large N and M [20], clearly forbidding deterministic super-replication.

The difference between unitary gates and maximally entangled states is due to the probabilistic nature of the Choi isomorphism, which allows one to retrieve a qubit gate from the corresponding maximally entangled state only with probability 1/4 [36, 37]. On the other hand, the probabilistic reversibility of the Choi isomorphism implies that super-replication of maximally entangled states *is* possible, albeit with an exponentially small probability $1/4^N$. The protocol works as follows: *i)* prepare M pairs of qubits, with each pair in the maximally entangled state $|\Phi^+\rangle = (|0\rangle|0\rangle + |1\rangle|1\rangle)/\sqrt{2}$, *ii)* on the first qubit of each pair, apply the operation $\mathcal{C}_1$ of the gate super-replication protocol, thus producing N qubits, *iii)* teleport each of the N qubits, using one of the N copies of the unknown entangled state $|\Phi_g\rangle$ as resource, and finally *iv)* apply the operation $\mathcal{C}_2$ of the gate super-replication protocol, thus producing M qubits. In the lucky case where all the Bell measurements used to teleport the N qubits give the favourable outcome, this protocol outputs M pairs of qubits in a state that has fidelity $F_E \geq 1 - 4e^{-\frac{N^2}{2M}}\left[1 + O(J/\sqrt{M})\right]$ with M perfect copies of $|\Phi_g\rangle$ (cf. the proof of theorem 1). This argument is interesting not only because it provides an explicit protocol for the super-replication of maximally entangled states, but also because it allows one to prove the optimality of the gate super-replication protocol: if there existed a protocol producing $M \gg N^2$ almost perfect copies out of N copies of an unknown gate, such a protocol could be converted into a (probabilistic) protocol producing $M \gg N^2$ almost perfect copies out of N copies of an unknown maximally entangled state. But this is known to be impossible, because it would lead to a violation of the Heisenberg limit [19]. Even more strongly, since Ref. [19] showed that the fidelity of state replication must vanish for $M \gg N^2$, the bound of Eq. (5) leads to the conclusion that a gate replication protocol with $M \gg N^2$ will have vanishing fidelity on most input states. This conclusion equally applies to deterministic and probabilistic protocols.

Gate compression. The argument used to prove gate super-replication can also be applied to the task of quantum gate compression, which consists in encoding the action of an unknown gate U_x in a gate U'_x acting on a smaller physical system. Gate compression protocols are useful in a distributed scenario where a service provider (Alice) has to apply the gate U_x to the input state provided by a client (Bob). In this task, it is natural to minimize the total amount of communication between client and provider, by compressing the client's input and the provider's gate into the smallest physical support. Ideally, the compression should be faithful, in the sense that the action of U_x on the original input state can be retrieved to a good degree of approximation. Now, consider the specific scenario where U_x is an N-qubit gate of the i.i.d. form $U_g^{\otimes N}$. The case where g is a rotation around a fixed axis has been considered in [22], but again here we want to address the universal case, where the gate is completely unknown. Using theorem 1 one can prove that for large N any such gate can be compressed to a gate U'_g act-

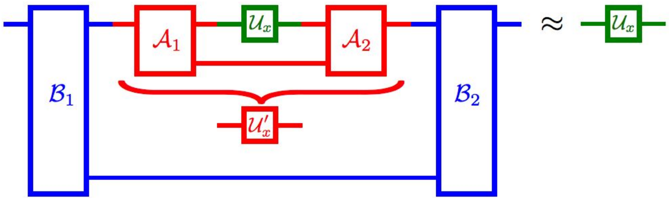

FIG. 2. **Gate compression.** Alice sandwiches the given gate $\mathcal{U}_x$ (in green) between two quantum channels $\mathcal{A}_1$ and $\mathcal{A}_2$ (in red), thus reducing it to a gate $\mathcal{U}'_x$ acting on a smaller quantum system. The action of $\mathcal{U}_x$ is retrieved through Bob's decoding operations (in blue). The protocol reduces the amount of quantum communication needed to simulate the application of Alice's gate to Bob's input.

ing only on $O(\log N)$ qubits, with an error that vanishes on almost all input states. The protocol works as follows: first, Bob applies the encoding operation $\mathcal{E}_J^{(N)}$ to the input state, a generic state of N qubits. Thanks to theorem 1, for $J \gg \sqrt{N} \gg 1$ this operation will cause little disturbance, except on an exponentially small fraction of the inputs. Specifically, one can set J to scale like $\sqrt{N^{1+\delta}}$ for every desired $\delta > 0$. Now, the output of $\mathcal{E}_J^{(N)}$ can be encoded faithfully into the state of a composite system AB, where system A has Hilbert space $\mathscr{H}_A = \bigoplus_{j=0}^{J} \mathscr{R}_j$ and system B has Hilbert space $\mathscr{H}_B = \mathscr{M}_J$. Let W_J be the isometry that achieves such encoding. After the application of W_J, Bob sends system A to Alice and keeps system B in his laboratory. On her hand, Alice has only to implement the gate $U'_g = \bigoplus_{j=0}^{J} U_g^{(j)}$ on system A: if she succeeds in doing that, then she can send system A back to Bob, who can decode by applying the inverse of W_J. Overall, this sequence of operations is equivalent to the application of the gate $U_g^{\otimes N}$ to the truncated state, which is close to the original input state with probability exponentially close to 1. Implementing the gate U'_g with N uses of U_g is the actual gate compression. Alice can achieve it by following these instructions: *i)* take a copy of system B, call it B', and initialize it in a fixed state $|\mu\rangle$, *ii)* apply the inverse of W_J, thus producing N qubits, *iii)* apply the gate $U_g^{\otimes N}$, apply *iv)* apply W_J and *v)* discard system B'.

Since the dimension of system A is $d_A = \sum_{j=0}^{J} (2j + 1) = \Theta\left(J^2\right) = \Theta(N^{1+\delta})$, we have that a completely unknown i.i.d. sequence $U_g^{\otimes N}$ can be compressed to a gate acting on $k = (1+\delta) \log N$ qubits for arbitrarily small $\delta > 0$. Quite surprisingly, this is the same compression rate that can be achieved if one tries to compress an arbitrary i.i.d. state $|\varphi\rangle^{\otimes N}$ of N qubits [23, 24]. This equality of compression rates comes rather unexpected if one consider the differences between state and gate replication. Note however that if one wants a compression protocol that works well on *every* input state, then one has to set $J = N/2$, which implies that the dimension of the Hilbert space $\mathscr{H}_A$ scales like N^2 and that the number of qubits required for compressing the gate is $2 \log N$. This is still an exponential saving of workspace with respect to the initial N qubits.

Gate super-replication in higher dimensions. We introduced the universal gate-replication protocol for qubits. However, the same idea can be easily extended to higher dimensions. An immediate extension is to the super-replication of rotation gates for spin-k particles with $k \geq 1/2$, that is, gates of the form $U_g^{(k)} = e^{-i\varphi \mathbf{n}\cdot\mathbf{j}}$ with $j_x^2 + j_y^2 + j_z^2 = k(k+1)\, I$. The super-replication protocol has the same structure of the protocol for qubit gates, with the only difference is that the probability measure in theorem 1 is now given by $p_j^{(M,k)} = d_j m_j^{(M,k)}/(2j+1)^M$, where $m_j^{(M,k)}$ is the multiplicity of the irreps with quantum number j in the decomposition of the tensor product representation $U_g^{(k)\,\otimes M}$. For large M, the multiplicity approaches the value [38]

$$m_j^{(M,k)} = \sqrt{\frac{27(2k+1)^{2M}(2j+1)^2}{8\pi M^3 k^3 (k+1)^3}}\, e^{-\frac{3j^2}{2Mk(k+1)}} \qquad (9)$$

and the entanglement fidelity in Eq. (5) can be bounded by

$$F_E^{(k)} \geq 1 - \sqrt{\frac{8(k+1)M}{3\pi k N^2}}\, e^{-\frac{3kN^2}{2(k+1)M}} \qquad (10)$$

(cf. Appendix A.2). Hence, deterministic super-replication can be achieved for all rotation gates on spin-k particles. The minimal ancilla has dimension $d_{\min}^{(k)} = \sqrt{M^3/N^3}(2k+1)^{M-N}$, i. e. it can be realized by adding $\Theta(M)$ spin-k systems. With to a more technical treatment, it is also possible to show that gate super replication can be achieved also for arbitrary gates in arbitrary dimensions, leveraging on the concentration of the Schur-Weyl measure on $\mathsf{SU}(d)$ [39]. In the same way, it is possible to show that N uses of a completely unknown gate can be compressed into a single gate acting on $(d-1)(d/2+1) \log N$ qubits (cf. Appendix A.3).

In conclusion, we showed that N uses of a completely unknown unitary gate are sufficient to simulate $M \ll N^2$ uses of the same gate, with high accuracy on all input states except for an exponentially small fraction. The protocol is optimal, in the sense that any attempt to simulate N^2 or more uses is doomed to have vanishing fidelity on most input states. The arguments developed in this paper do not apply only to gate replication, but also to the task of gate compression. For this task, we demonstrated that an unknown i.i.d. sequence of N unitary gates can be compressed down to a single gate acting only on $O(\log N)$ qubits, thus achieving an exponential saving of computational workspace.

Acknowledgments. This work is supported by the National Basic Research Program of China (973) 2011CBA00300 (2011CBA00301), by the National Natural Science Foundation of China through Grants

11350110207, 61033001, and 61061130540, by the Foundational Questions Institute through the Large Grant "The fundamental principles of information dynamics", and by the 1000 Youth Fellowship Program of China. GC is grateful to P Hayden for pointing out the concentration result for the Schur-Weyl measure in Ref. [39] and to W Dür for valuable comments on the manuscript.

[1] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).

[2] D. Dieks, Physics Letters A **92**, 271 (1982).

[3] D. Gottesman, in *Quantum Information Science and Its Contributions to Mathematics, Proceedings of Symposia in Applied Mathematics*, Vol. 68 (2009) pp. 13–58.

[4] D. A. Lidar and T. A. Brun, *Quantum error correction* (Cambridge University Press, 2013).

[5] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, Rev. Mod. Phys. **77**, 1225 (2005).

[6] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, Rev. Mod. Phys. **81**, 1301 (2009).

[7] S. Wiesner, SIGACT News **15**, 78 (1983).

[8] S. Aaronson, 24th Annual IEEE Conference on Computational Complexity , 229 (2009).

[9] E. Farhi, D. Gosset, A. Hassidim, A. Lutomirski, and P. Shor, Proceedings of the 3rd Innovations in Theoretical Computer Science Conference , 276 (2012).

[10] A. Molina, T. Vidick, and J. Watrous, Proceedings of the Conference on Theory of Quantum Computation, Communication, and Cryptography , 45 (2013).

[11] M. Hillery, V. Bužek, and A. Berthiaume, Phys. Rev. A **59**, 1829 (1999).

[12] R. Cleve, D. Gottesman, and H.-K. Lo, Phys. Rev. Lett. **83**, 648 (1999).

[13] V. Bužek and M. Hillery, Phys. Rev. A **54**, 1844 (1996).

[14] N. Gisin and S. Massar, Phys. Rev. Lett. **79**, 2153 (1997).

[15] R. F. Werner, Phys. Rev. A **58**, 1827 (1998).

[16] M. Keyl and R. F. Werner, Journal of Mathematical Physics **40**, 3283 (1999).

[17] L.-M. Duan and G.-C. Guo, Phys. Rev. Lett. **80**, 4999 (1998).

[18] J. Fiurášek, Phys. Rev. A **70**, 032308 (2004).

[19] G. Chiribella, Y. Yang, and A. C.-C. Yao, Nat. Comm. **4** (2013).

[20] G. Chiribella and Y. Yang, New Journal of Physics **16**, 063005 (2014).

[21] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 180504 (2008).

[22] W. Dür, P. Sekatski, and M. Skotiniotis, arXiv preprint (2014), arXiv 1410.6008.

[23] M. Plesch and V. Bužek, Phys. Rev. A **81**, 032317 (2010).

[24] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, Phys. Rev. Lett. **113**, 160504 (2014).

[25] J. I. Cirac, A. K. Ekert, and C. Macchiavello, Phys. Rev. Lett. **82**, 4344 (1999).

[26] D. Bacon, I. L. Chuang, and A. W. Harrow, Phys. Rev. Lett. **97**, 170502 (2006).

[27] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

[28] M. A. Nielsen, Physics Letters A **303**, 249 (2002).

[29] P. Méliot, arXiv preprint (2010), arXiv 1009.4034.

[30] P. Biane, International Mathematics Research Notices **2001**, 179 (2001).

[31] S. Popescu, A. J. Short, and A. Winter, Nat. Phys. **2**, 754 (2006).

[32] P. Hayden, D. W. Leung, and A. Winter, Communications in Mathematical Physics **265**, 95 (2006).

[33] M.-D. Choi, Linear Algebra and its Applications **10**, 285 (1975).

[34] D. W. Leung, *Towards Robust Quantum Computation*, Ph.D. thesis, Stanford University (2000).

[35] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[36] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[37] D. Genkina, G. Chiribella, and L. Hardy, Phys. Rev. A **85**, 022330 (2012).

[38] G. Chiribella, R. Chao, and Y. Yang, arXiv preprint (2014), arXiv 1411.3439.

[39] A. W. Harrow, arXiv preprint quant-ph/0512255 (2005).

A.1 The entanglement fidelity in Theorem 1

By the definition of Schur-Weyl measure and Eq. (2) we have

$$\begin{aligned} p_j^{(M)} &= \frac{2(2j+1)^2}{2^M(M+2j+2)}\binom{M}{M/2+j} \\ &= \frac{2j+1}{2^M}\left[\binom{M}{M/2+j} - \binom{M}{M/2+j+1}\right] . \end{aligned} \tag{11}$$

Inserting Eq. (11) into the expression of the entanglement fidelity, one obtains

$$\begin{aligned} F_E^{(M,J)} &= \left\{\sum_{j=0}^{J}\frac{2j+1}{2^M}\left[\binom{M}{M/2+j} - \binom{M}{M/2+j+1}\right]\right\}^2 \\ &= \left\{\frac{1}{2^M}\left[\sum_{j=M/2-J}^{M/2+J}\binom{M}{j} - (2J+1)\binom{M}{M/2+J+1}\right]\right\}^2 . \end{aligned}$$

When $M \gg 1$, we can bound the first term within the parentheses of the last equality with Hoeffding's inequality as

$$\frac{1}{2^M}\sum_{j=M/2-J}^{M/2+J}\binom{M}{j} \geq 1 - 2e^{-\frac{2J^2}{M}}.$$

The second term can be estimated using de Moivre-Laplace formula, which gives

$$\frac{2J+1}{2^M}\binom{M}{M/2+J+1} = O\left(\sqrt{\frac{J^2}{M}}e^{-\frac{2(J+1)^2}{M}}\right).$$

Clearly, for $J \ll \sqrt{M}$ this value is negligible compared to the value of the first term, which scales like $e^{-\frac{2J^2}{M}}$. Combining the above observations, the entanglement fidelity is lower bounded as

$$\begin{aligned} F_E^{(M,J)} &\geq 1 - 4e^{-\frac{2J^2}{M}} - O\left(\sqrt{\frac{J^2}{M}}\,e^{-\frac{2(J+1)^2}{M}}\right) \\ &\geq 1 - 4\,e^{-\frac{2J^2}{M}}\left[1 + O(J/\sqrt{M})\right]. \end{aligned}$$

A.2 Gate replication fidelity in higher dimensions

The entanglement fidelity for the super-replication of rotation gates for spin-k particles can be derived when $N, M \gg 1$, by inserting the multiplicity expression (9) into Eq. (5) and substituting the sum by an integral, as

$$\begin{aligned} F_E^{(k)} &= \left[\sum_{j=0}^{kN} p_j^{(M,k)}\right]^2 \\ &= \left[\int_0^{kN} \underset{\cdot}{\mathrm{j}}\ \frac{3\sqrt{3}(2j+1)^2}{2\sqrt{2\pi M^3 k^3 (k+1)^3}} e^{-\frac{3j^2}{2Mk(k+1)}}\right]^2 \\ &= \mathrm{erf}^2\left[\sqrt{\frac{3kN^2}{2(k+1)M}}\right] + O(M^{-\frac{1}{2}}) \\ &\geq 1 - \sqrt{\frac{8(k+1)M}{3\pi k N^2}}\, e^{-\frac{3kN^2}{2(k+1)M}}. \end{aligned} \qquad (12)$$

Here $\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\underset{\cdot}{\mathrm{t}}$ is the error-function for the normal distribution and the lower-bound comes from the first-term Taylor expansion of $1 - \mathrm{erf}(x)$ when x is large, i. e. for $N^2/M \gg 1$.

A.3 Gate compression in higher dimensions

In the case of d-dimensional quantum systems (qudits), the Hilbert space of N identical systems can be decomposed as

$$\mathscr{H}^{\otimes N} \simeq \bigoplus_{j \in \mathcal{P}_{N,d}} \left(\mathscr{R}_j^{(N)} \otimes \mathscr{M}_j^{(K)}\right), \qquad (13)$$

where the sum runs over the set $\mathcal{P}_{N,d}$ of all partions of N consisting of d non-negative numbers, $\mathscr{R}_j^{(N)}$ is a representation space and $\mathscr{M}_j^{(N)}$ is a multiplicity space. Now, it is possible to show that the dimension of each representation space is upper bounded by $d_j = (N+d)^{d(d-1)/2}$ (see e.g. [39]). Since the total number of partitions of N into d non-negative integers is equal to $\binom{N+d-1}{d-1} \leq (N+1)^{d-1}$, one has $\sum_{y \in \mathcal{P}_{N,d}} d_j \leq (N+1)^{d-1}(N+d)^{d(d-1)/2}$, which approaches $N^{(d-1)(d/2+1)}$ for large N. In this regime, the compression protocol that encodes the gate $U^{\otimes N}$ into a gate acting on the Hilbert space $\mathscr{H}_A = \bigoplus_{j \in \mathcal{P}_{N,d}}$ can be implemented using $(d-1)(d/2+1)\log N$ qubits.