Optimal universal programming of unitary gates

Yuxiang Yang,1,* Renato Renner,1,† and Giulio Chiribella2,3,4,5,†

1Institute for Theoretical Physics, ETH Zürich
2QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong
3Department of Computer Science, Parks Road, Oxford, OX1 3QD, UK
4Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
5The University of Hong Kong Shenzhen Institute of Research and Innovation, 5/F, Key Laboratory Platform Building, No.6, Yuxing 2nd Rd., Nanshan, Shenzhen 518057, China

A universal quantum processor is a device that takes as input a (quantum) program, containing an encoding of an arbitrary unitary gate, and a (quantum) data register, on which the encoded gate is applied. While no perfect universal quantum processor can exist, approximate processors have been proposed in the past two decades. A fundamental open question is how the size of the smallest quantum program scales with the approximation error. Here we answer the question, proving a fundamental bound on the size of the program and designing a concrete protocol that attains the bound in the asymptotic limit. Our result is based on a connection between optimal programming and the Heisenberg limit of quantum metrology, and establishes an asymptotic equivalence between the tasks of programming, learning, and estimating unitary gates.

Introduction. A universal quantum processor is the desideratum of quantum computing. Ideally, one would hope to realise quantum computing in the same way as its classical counterpart, i.e., by inserting data and programs, both in the form of quantum states, into a universal quantum computer. However, the no-programming theorem [1] asserts that any universal quantum processor must be approximate, or have a non-zero probability of failure [1–3].

It has been shown that approximate universal processors with a finite-size program register do exist [1, 4–9]. There one of the most important questions is to determine the cost-accuracy tradeoff or, more specifically, how the program cost, i.e., the number \( c_p \) of qubits required to store the optimal program, scales with the desired accuracy of implementation, which we quantify in terms of an approximation error \( \epsilon \).

Over the past two decades, many efforts have been dedicated to finding the optimal approximate universal processor [4, 5, 8, 9] (see also Table I). The state-of-the-art result, [9], asserts that the optimal program cost \( c_p \) for a \( d \)-dimensional quantum computer lies between \( c_{\text{low}} := [(1 - \epsilon)K]d - (2/3) \log d \) qubits and \( c_{\text{upp}} := d^2 \log (K/\epsilon) \) qubits, where \( K \) is a universal constant. Despite all efforts, the precise value for \( c_p \) remains largely unknown — especially in the small error regime, where the ratio \( c_{\text{upp}}/c_{\text{low}} \) diverges.

In this work, we close this gap by identifying the optimal scaling of the program cost with the accuracy and therefore solving a long-standing open problem of optimal quantum programming. Specifically, our program cost scales as \( [(d^2 - 1)/2] \log (1/\epsilon) \) in the small \( \epsilon \) regime, which reduces the cost of the best existing protocol (see \( c_{\text{upp}} \) above) by half. The optimal scaling is achieved with a gate learning protocol, where the program is prepared by sending a quantum state through \( n \) instances of the gate to learn it [10]. The gate information is later read out by measuring the program. Our protocol achieves an error scaling of \( 1/n^2 \) — well-known as the Heisenberg limit of quantum metrology [11–14]. We thus prove the asymptotic equivalence of quantum gate programming, metrology, and learning.

Preliminaries. We consider programming unitary gates of a system with a \( d \)-dimensional Hilbert space \( \mathcal{H} \). The gates, up to a global phase, form the special unitary group SU\((d)\). For a pure state \( |\psi\rangle \), we abbreviate its density matrix \( |\psi\rangle\langle\psi| \) by \( \psi \). Similarity, \( U(\cdot):=U(\cdot)U^{\dagger} \) denotes a unitary channel.

We will use the \( \bigOmega \) notation, the \( \bigO \) notation, and the \( \bigTheta \) notation for asymptotic behaviour of functions. For a function \( f(n) \), we write \( f(n) = \bigOmega(g(n)) \) if there exists a constant \( c_1 > 0 \) so that \( f(n) \geq c_1 g(n) \) for large enough \( n \), \( f(n) = O(g(n)) \) if there exists a constant \( c_2 > 0 \) such that \( f(n) \leq c_2 g(n) \) for
0 so that $|f(n)| \leq c_2 g(n)$ for large enough $n$, and $f(n) = \Theta(g(n))$ if $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$. We will also abbreviate log$_2$ by log.

**Approximate universal processors.** A universal quantum processor consists of two key elements: a set of programs $\{\psi_{P,U}\}_{U \in SU(d)}$, which are quantum states in $\mathcal{H}_P$, and the action of the processor $C$, which is a quantum channel (i.e. a completely positive trace-preserving linear map) acting on the composite Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_P$ of the system and the program. Notice that all information on $U$ should come from the program, and $C$ must be independent of $U$. The program cost $c_P$ is defined as log$_2 d_P$, with the program dimension $d_P$ being the dimension of $\text{Supp}(\psi_{P,U})_{U \in SU(d)}$.

As shown in Figure 1, to run any arbitrary unitary $U$ on the system, one selects the corresponding program $\psi_{P,U}$ and plugs it into the processor, resulting in the following channel on the system:

$$\mathcal{E}_U(\cdot) := \text{Tr}_P [C(\cdot \otimes \psi_{P,U})].$$

(1)

A tuple $(C, \{\psi_{P,U}\}_{U \in SU(d)})$ is called an $\epsilon$-universal processor, if

$$\frac{1}{2} \|U - \mathcal{E}_U\|_\diamond \leq \epsilon \quad \forall U \in SU(d).$$

(2)

Here $\| \cdot \|_\diamond$ denotes the diamond norm [19], which equals the maximum trace distance between the outputs of the two channels, maximized over all input states and over all possible reference systems.

The no-programming theorem [1] rules out perfect (i.e. $\epsilon = 0$) universal processors. This impossibility result raised the question: “Given a desired accuracy $1/\epsilon$, how big does the program need to be?” This question can of course be subdivided into two, namely to find upper and lower bounds on the program cost $c_P$. We summarise the best known results in Table I. Here we are providing both a new lower and a new upper bound, which match in terms of their asymptotic dependence on $1/\epsilon$.

**Lower bound on the program dimension.** Here we establish a fundamental lower bound on the program dimension. For this purpose, we exploit an alternative proof of the no-programming theorem, originally developed in the framework of general probabilistic theories [20]. The idea is that the exact implementation of a unitary gate requires the channel $C$ to leave the system and the program uncorrelated. Using this fact, the program can be recycled, thereby generating multiple copies of the desired unitary gate. The approximate version of this argument was first used by us to determine the energy requirement of quantum processors [21] and is further exploited here.

To approximate a unitary quantum gate $U$ with good precision, there should be almost no correlation between the system and the program after we apply $C$. This means that the complementary channel $\mathcal{E}_U$, defined as $\mathcal{E}_U(\cdot) := \text{Tr}_S [C(\rho_S \otimes \cdot)]$, is almost independent of $\rho_S$. It further suggests that, instead of discarding the program after one usage, we can recycle it: We can invert the action of $\mathcal{E}_U$ on the program state by a $(\rho_S)$-independent operation and get back the original program. The program can be further used, generating multiple uses of $U$ at the cost of an increased approximation error. Notice that the argument does not hold for noisy or classical processes. For instance, using a controlled unitary $|0\rangle \langle 0| \otimes I + |1\rangle \langle 1| \otimes \sigma_z$ and an ancillary qubit $(1/\sqrt{2})(|0\rangle + |1\rangle)$ one can (perfectly) implement the channel $\rho \mapsto (1/2)(\rho + \sigma_z \rho \sigma_z)$. However, the system and the ancillary qubit become strongly correlated after the implementation.

By the above argument, we can show (see Appendix A for details) that an $\epsilon$-universal processor for a single use of $U$ can be turned into a $(4m \sqrt{2}\epsilon)$-universal processor for $m$ uses of $U$ for any $m \geq 1$. This requires the original program to contain enough information for programming up to $1/\sqrt{\epsilon}$ uses of $U$. This fact, in turn, implies a bound on its minimum information content and therefore its size. This ultimately leads to the following theorem, which can be regarded as a quantitative version of the no-programming theorem [1]:

**Theorem 1 (Approximate no-programming theorem).** Consider any $\epsilon$-universal processor with program dimension $c_P$. For any $(\epsilon$-independent) parameter $\delta > 0$, the

| Previous works | Upper bounds | Lower bounds |
|----------------|--------------|--------------|
|                | $d^2 \log (K/\epsilon)$ [9] | $[(1 - \epsilon)K]d - (2/3) \log d$ [9] |
|                | $4d^2 \log d/e^2$ [8, 15, 16] | $2 \log (d/\epsilon)$ [17] |

TABLE I. Comparison of bounds on universal quantum gate programming. In the table we compare our results with the best previous results (summarised from Table I of Ref. [9]). In the vanishing error regime $\epsilon \to 0$, both our lower bound and our upper bound are tighter than all previous results, for the first time closing the gap between the lower and upper bound in this regime. The error is evaluated in terms of the diamond norm (2). $K$ denotes a universal constant.
program dimension is lower bounded as
\[ c_P \geq (1 - \delta - 4\sqrt{2}\epsilon)(d^2 - 1)\log \left( \frac{\delta}{4\sqrt{2}\epsilon(2d - 1)(d - 1)} \right) - 1. \] (3)

This immediately implies the expression for the lower bound stated in Table I. The key message from the above theorem is that, for any \( \alpha < (d^2 - 1)/2 \), the program dimension \( d_P = 2^\alpha \) satisfies
\[ d_P = \Omega(1/\epsilon^\alpha) \] (4)

Taking \( \epsilon \to 0 \) in Eq. (4), one gets \( d_P \to \infty \), recovering the original no-programming theorem [1].

**Optimal approximate universal processor.** Next we construct an approximate universal processor that achieves the bound in Theorem 1. Our processor works in a measure-and-operate (MO) fashion, as illustrated in Figure 2. It measures the input program \( \psi_{P,U} \) with a suitable POVM \( \{d\hat{U}M_U\}_{U \in SU(d)} \), where \( d\hat{U} \) is the Haar measure. The measurement yields an estimate \( \hat{U} \) of the gate \( U \), and the processor performs the corresponding gate on the system. Explicitly, our optimal processor obeys the following procedure:

**Protocol 1** A MO universal processor.
1: (Generating the program.)
   Apply \( U_{\psi_P} \) to a suitable quantum state \( |\psi_P\rangle \).
2: Measure \( |\psi_{P,U}\rangle := U_{\psi_P}^\dagger|\psi_P\rangle \) with \( \{M_U\}_{U \in SU(d)} \).
3: Apply \( \hat{U} \) to the state of the system, where \( \hat{U} \) is the measurement outcome.

The program in Protocol 1 is prepared by applying \( n \) parallel uses of \( U \) on a quantum state (called the probe state). The performance of this processor is then determined jointly by the choice of the probe state and the choice of the POVM \( \{d\hat{U}M_U\}_{U \in SU(d)} \). It is known from quantum metrology [12, 13, 22] that the performance of the measurement is optimised using non-product probe states and POVMs\(^3\). In Appendix B, we construct an estimation protocol which, when incorporated into Protocol 1, yields an optimal processor asymptotically achieving the \( (d^2 - 1)/2 \log(1/\epsilon) \) scaling bound of Theorem 1.

**Theorem 2.** Consider quantum metrology of a \( d \)-dimensional unitary given \( n \) uses. Suppose \( n \geq 2d(d-1) \) and \( d \geq 2 \). The error for the optimal unitary gate estimation is bounded as
\[ \epsilon \leq 2d \left( \frac{\pi(d-1)^2(3d-2)}{d \cdot n} \right)^2. \] (5)
The probe state has dimension bounded as
\[ d_P \leq \left( \frac{9n}{3d-2} \right)^{d^2-1}. \] (6)

Ref. [22] showed that the estimation of an arbitrary \( d \)-dimensional unitary given \( n \) uses can be done with an error scaling \( 1/n^2 \). The error was measured by the entanglement gate infidelity, which is upper bounded by \( (1 - \epsilon)^2 \). Here we refine this result by not only achieving the \( 1/n^2 \) scaling but also identifying an explicit expression of the constant of proportionality. In addition, our result holds for the more stringent error criterion \( \epsilon \), i.e., the diamond norm error, and we also determine how the probe state dimension scales with \( n \).

Combining Eq. (5) with Eq. (6), we get:

**Corollary 1.** The program cost \( c_P \) of Protocol 1 is upper bounded as
\[ c_P \leq \left( \frac{d^2 - 1}{2} \right) \log \left( \frac{162\pi^2(d-1)^4}{d \cdot \epsilon} \right). \] (7)

It is obvious from the above corollary that
\[ c_P \leq \left( \frac{d^2 - 1}{2} \right) \log \left( \frac{162\pi^2d^3}{\epsilon} \right), \] (8)
which matches Table I and achieves a quadratic reduction compared to known results.

**Asymptotic equivalence of programming, metrology, and learning.** From the previous discussion, we can see that an optimal way of programming a unitary is actually to let the processor learn and memorize it (see Figure 2). The task of learning a unitary \( U \) from \( n \) instances

\(^3\) Here we take a Bayesian perspective of quantum metrology, where the gate is assumed to be drawn randomly according to a uniform measure. This is to be distinguished from the minimax perspective (see, e.g., Refs. [23, 24]) which is also very common.
[3, 10, 25, 26] consists of a learning phase and an execution (or testing) phase. In the learning phase, the protocol makes \( n \) (not necessarily parallel) queries to \( U \). In the execution phase, the protocol emulates the learned unitary on an arbitrary input state. Notice that the execution phase happens after the learning phase, thus the protocol should be able to store the information of \( U \).

Learning and programming are without doubt two relevant tasks. A learning protocol induces a programmable processor in the sense that the learning phase can be used to generate a program. Nevertheless, one should keep in mind that they are not equivalent. Indeed, in the task of programming, the program does not have to be generated by learning, i.e., by applying multiple instances of \( U \) on a quantum state. As learning has this addition constraint, its resource requirement is at least as stringent as that of programming. Therefore, since Protocol 1 is an optimal processor, it is also an optimal learning protocol. The performance of optimal learning given \( n \) instances is thus given by Theorem 1, achieved by unitary gate metrology. In summary, for finite dimensional quantum gates, the performances of programming, metrology and learning are asymptotically equal:

\[
\text{(Asymptotically) programming} \approx \text{metrology} \approx \text{learning}.
\]

**Quantum versus classical advantage.** One may wonder if it is possible to simply use a classical program, e.g., to write down the description of the gate on a tape. Here we show, via a simple example, that our Protocol 1, which uses a quantum program, beats the best processor that uses classical programs in scaling.

To see this, let us consider the case of programming a phase gate \( U_\theta = |0\rangle \langle 0| + e^{-i\theta} |1\rangle \langle 1| \), where \( \theta \in [0, 2\pi] \) is the (unknown) phase. Fixing the program dimension \( d_p := 2^{\nu} \), the best classical strategy is nothing but dividing the range \([0, 2\pi] \) into \( d_p \) equal-width intervals. The tag of the interval that contains \( \theta \) is used as the program, and the processor runs \( U_\theta \) with \( \theta \) being the middle point of the interval. Since max \( |\hat{\theta} - \theta| = \pi/d_p \), the error of this approach is \( \epsilon_{\text{classical}} = \sqrt{(1 - \cos(\pi/d_p))/2} \approx \pi/(2d_p) \), which is inversely proportional to the program dimension. Notice that, in the more complex case of programming a \( d \)-dimensional unitary gate, this approach extends to the \( \epsilon \)-mesh construction of programs, which was employed by Ref. [9].

In contrast, we can employ our Protocol 1, where we use the sine state [11]

\[
|\psi\rangle = \sqrt{\sum_{m=0}^{d_p-1} \sin \frac{\pi(m + 1/2)}{d_p}} |m\rangle.
\]

as the probe state and the covariant POVM

\[
\left\{ \frac{d \hat{\theta}}{2\pi} |\hat{\eta}_{\theta}\rangle \langle \hat{\eta}_{\theta}| : |\hat{\eta}_{\theta}\rangle := \sum_{m=0}^{d_p-1} e^{-i\theta m} |m\rangle \right\}_{\hat{\theta}}
\]

as the measurement. The error can be evaluated as

\[
\epsilon_{\text{quantum}} \simeq \frac{\pi^2}{2d_p^2},
\]

which is inversely proportional to the square of the program dimension. In other words, the program dimension of a processor with classical programs is quadratically larger than that of our quantum processor. This proves the claimed quantum-over-classical advantage in programming.

**Conclusion and further discussions.** We identified the optimal scaling of the program dimension with accuracy in a universal quantum processor. The optimal scaling can be achieved with a measure-and-operate learning protocol. With this finding, we showed the asymptotic equivalence between programming, metrology, and learning.

The task we focused on is universal programming, which requires the processor to work well for every gate of a certain dimension. It is natural to expect that a smaller set of gates would lead a smaller program cost. Observe from Eq. (8) that the prefactor \((d^2 - 1)/2\) is exactly one half of the number of real parameters determining a qudit unitary gate (up to a global phase). We therefore conjecture a general formula

\[
c_p \sim \left( \frac{\nu}{2} \right) \log \left( \frac{C_{\nu,d}}{\epsilon} \right)
\]

where \( \nu \) is the number of variable real parameters required to identify the unitary and \( C_{\nu,d} \) is a parameter, possibly dependent on \( \nu \) and \( d \) but independent of \( \epsilon \).

Another key reason for making this conjecture is that the ultimate performances of quantum information processing tasks share similar forms in the asymptotic limit of “many copies”. In particular, one can consider the compression of identically prepared quantum systems, e.g., states of the form \( \rho^{\otimes n} \) with \( \rho \) unknown and \( n \) large. It turns out that the minimum cost of the memory, when requiring the error to be vanishing for large \( n \), is \((\nu/2) \log n \) (qu)bits in the leading order [27–32]. Here \( \nu \), the number of variable real parameters, appears again. Further pursuit in this direction could lead to the discovery of a universality rule, which governs the behaviour of optimal quantum devices in the limit of macroscopically many copies.

**ACKNOWLEDGMENTS**

This work is supported by the National Natural Science Foundation of China through grant 11675136, the Hong Kong Research Grant Council through grant 17300317, the Foundational Questions Institute through grant FQXi-RFP3-1325, the Croucher Foundation, the AFOSR via grant No. FA9550-19-1-0202, the Swiss National Science Foundation via the National Center for Competence in Research “QSIT” as well as via project No. 200020_165843, and the ETH Pauli Center for Theoretical Studies.
Appendix A: Proof of Theorem 1

Consider any \( \epsilon \)-universal processor \( \langle C, \{ \psi_{P,U} \} \rangle \). We prove Theorem 1, which is a lower bound on the dimension \( d_F \) of the program, i.e. the dimension of \( \text{Supp}\{\psi_{P,U}\} \).

We first show that programming one use of \( U \) with error \( \epsilon \) requires the same amount of information as programming \( m \) uses of \( U \) with error \( 4m\sqrt{2\epsilon} \) for any \( m \geq 1 \). Note that the proof here extends that of [21, Corollary 2]. First, we define the worst-case input (or minimum) fidelity between two arbitrary quantum channels \( A \) and \( B \), defined as [33]

\[
F_{\text{wc}}(A, B) := \inf_{\Psi} F\left( (A \otimes I_R)(\Psi), (B \otimes I_R)(\Psi) \right),
\]

where the infimum is taken over all pure states \( |\Psi\rangle \in \mathcal{H}_S \otimes \mathcal{H}_R \) with \( \mathcal{H}_R \simeq \mathcal{H}_S \) being a reference system, and

\([1] M. A. Nielsen and I. L. Chuang, Physical Review Letters 79, 321 (1997).\)
\([2] M. Hillery, V. Bužek, and M. Ziman, Physical Review A 65, 022301 (2002).\)
\([3] M. Sedlák, A. Bisio, and M. Ziman, Physical Review Letters 122, 170502 (2019).\)
\([4] J. Kim, Y. Cheong, J.-S. Lee, and S. Lee, Physical Review A 65, 012302 (2001).\)
\([5] M. Hillery, V. Bužek, and M. Ziman, Fortschritte der Physik: Progress of Physics 49, 987 (2001).\)
\([6] G. Vidal, L. Masanes, and J. I. Cirac, Physical Review A 52, 047905 (2000).\)
\([7] A. Brazier, V. Bužek, and P. L. Knight, Physical Review A 71, 032306 (2005).\)
\([8] S. Ishizaka and T. Hiroshima, Physical Review Letters 101, 240501 (2008).\)
\([9] A. M. Kubicki, C. Palazuelos, and D. Pérez-García, Physical Review Letters 122, 080505 (2019).\)
\([10] A. Bisio, G. Chiribella, G. M. D’Ariano, S. Facchini, and P. Perinotti, Physical Review A 81, 032324 (2010).\)
\([11] V. Bužek, R. Derka, and S. Massar, Physical Review Letters 82, 2207 (1999).\)
\([12] G. Chiribella, G. D’Ariano, P. Perinotti, and M. F. Sacchi, Physical Review Letters 93, 180503 (2004).\)
\([13] E. Bagan, M. Baig, and R. Munoz-Tapia, Physical Review A 70, 030301 (2004).\)
\([14] M. Hayashi, Physics Letters A 354, 183 (2006).\)
\([15] S. Beigi and R. König, New Journal of Physics 13, 090306 (2011).\)
\([16] M. Christandl, F. Leditzy, C. Majenz, G. Smith, F. SPEEDMAN, and M. Walter, arXiv preprint arXiv:1809.10751 (2018).\)
\([17] C. Majenz, Entropy in Quantum Information Theory—Communication and Cryptography, Ph.D. thesis, Faculty of Science, University of Copenhagen (2017).\)
\([18] D. Pérez-García, Physical Review A 73, 052315 (2006).\)
\([19] A. Y. Kitaev, Russian Mathematical Surveys 52, 1191 (1997).\)
\([20] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Physical Review A 81, 062348 (2010).\)
\([21] G. Chiribella, Y. Yang, and R. Renner, arXiv:1908.10884 (2019).\)
\([22] J. Kahn, Physical Review A 75, 022326 (2007).\)
\([23] V. Giovannetti, S. Lloyd, and L. Maccone, Physical Review Letters 96, 010401 (2006).\)
\([24] V. Giovannetti, S. Lloyd, and L. Maccone, Nature Photonics 5, 222 (2011).\)
\([25] S. Gammelmark and K. Mølmer, New Journal of Physics 11, 033017 (2009).\)
\([26] Y. Mo and G. Chiribella, New Journal of Physics 21, 113003 (2019).\)
\([27] M. Plesch and V. Bužek, Physical Review A 81, 032317 (2010).\)
\([28] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, Physical Review Letters 113, 160504 (2014).\)
\([29] G. Chiribella, Y. Yang, and C. Huang, Physical Review Letters 114, 120504 (2015).\)
\([30] Y. Yang, G. Chiribella, and D. Ebler, Physical Review Letters 116, 080501 (2016).\)
\([31] Y. Yang, G. Chiribella, and M. Hayashi, Physical Review Letters 117, 090502 (2016).\)
\([32] Y. Yang, G. Bai, G. Chiribella, and M. Hayashi, IEEE Transactions on Information Theory (2018).\)
\([33] M. A. Nielsen and I. Chuang, Quantum Information. Cambridge University Press, Cambridge (2000).\)
\([34] C. A. Fuchs and J. Van De Graaf, IEEE Transactions on Information Theory 45, 1216 (1999).\)
\([35] A. S. Holevo, Problemy Peredachi Informatsii 9, 3 (1973).\)
\([36] W. Fulton and J. Harris, Representation theory: a first course, Vol. 129 (Springer Science & Business Media, 2013).\)
\([37] R. Alicki and M. Fannes, Journal of Physics A: Mathematical and General 37, L55 (2004).\)
\([38] A. Winter, Communications in Mathematical Physics 347, 291 (2016).\)
\([39] M. Horodecki, P. Horodecki, and R. Horodecki, Physical Review A 60, 1888 (1999).\)
\([40] M. A. Nielsen, Physical Review Letters 303, 249 (2002).\)
\([41] M. Raginsky, Physics Letters A 290, 11 (2001).\)
\([42] G. Chiribella, G. D’Ariano, and M. Sacchi, Physical Review Letters 72, 042338 (2005).\)
\([43] S.-E. Ekström, C. Garoni, and S. Serra-Capizzano, Experimental Mathematics 27, 478 (2018).\)
\([44] J. Watrous, The theory of quantum information (Cambridge University Press, 2018).\)
\([45] C. Itzykson and M. Nauenberg, Reviews of Modern Physics 38, 95 (1966).\)
whose distance from an inequality $[34]$, we have

$$F_{wc}(E_U, U) \geq (1 - \epsilon)^2 \geq 1 - 2\epsilon \quad \forall U.$$  

(A2)

Denote by $V : \mathcal{H} \otimes \mathcal{H}_P \to \mathcal{H} \otimes \mathcal{H}_{P'}$ a Stinespring dilation of $C$, where $\mathcal{H}_{P'}$ is an ancillary space. There exists a state $|\phi_{P', U}\rangle \in \mathcal{H}_{P'}$ such that

$$F_{wc}(V \circ (I_S \otimes \psi_{P, U}), U \otimes \phi_{P', U}) \geq 1 - 2\epsilon.$$  

(A3)

Applying again the Fuchs - Van de Graaf inequality, we get

$$\|V \circ (I_S \otimes \psi_{P, U}) - U \otimes \phi_{P', U}\|_\diamond \leq 2\sqrt{2}\epsilon.$$  

(A4)

Notice that here $\psi_{P, U}$ is regarded as a channel that has trivial input and prepares the state $\psi_{P, U}$.

Next, define the pseudoinverse of $V$, $E_{inv, V} : \mathcal{H}_S \otimes \mathcal{H}_{P'} \to \mathcal{H}_S \otimes \mathcal{H}_P$, as the following quantum channel:

$$E_{inv, V}(\cdot) := V^\dagger \circ \rho_{\mathcal{H}_V}(\cdot) + \text{Tr} \left[ P_{\mathcal{H}_V^\perp}(\cdot) \right] \pi_{\mathcal{H}_S \otimes \mathcal{H}_{P'}}.$$  

(A5)

where for any Hilbert space $K$ we denote by $\pi_K$ the maximally mixed state, $\mathcal{H}_V$ is the image of $V$, and $P_{\mathcal{H}_V}$ or $P_{\mathcal{H}_V^\perp}$ is the projection operation into $\mathcal{H}_V$ or $(\mathcal{H}_S \otimes \mathcal{H}_{P'}) \setminus \mathcal{H}_V$. Then we have

$$V \circ E_{inv, V}(\cdot) = \rho_{\mathcal{H}_V}(\cdot) + \text{Tr} \left[ P_{\mathcal{H}_V^\perp}(\cdot) \right] \pi_{\mathcal{H}_V} \geq P_{\mathcal{H}_V}(\cdot).$$  

(A6)

It follows that

$$F_{wc}(V \circ (U^\dagger \otimes \psi_{P, U}), \rho_{\mathcal{H}_V}(I_S \otimes \phi_{P', U})) \leq F_{wc}(V \circ (U^\dagger \otimes \psi_{P, U}), V \circ E_{inv, V}(I_S \otimes \phi_{P', U})).$$  

(A7)

Since $(V \otimes I_R)((U^\dagger \otimes I_R)\rho \otimes |\psi_{P, U}\rangle) \in \mathcal{H}_V \otimes \mathcal{H}_R$ for any input state $|\rho\rangle \in \mathcal{H} \otimes \mathcal{H}_R$, we have

$$F_{wc}(V \circ (I_S \otimes \psi_{P, U}), U \otimes \phi_{P', U}) = F_{wc}(V \circ (U^\dagger \otimes \psi_{P, U}), \rho_{\mathcal{H}_V}(I_S \otimes \phi_{P', U})) \leq F_{wc}(V \circ (U^\dagger \otimes \psi_{P, U}), V \circ E_{inv, V}(I_S \otimes \phi_{P', U}))$$  

(A8)

$$= F_{wc}(U^\dagger \otimes \psi_{P, U}, E_{inv, V}(I_S \otimes \phi_{P', U})),$$  

(A9)

having used the property that $V$ (as an isometry) preserves fidelity in the last step. Applying again the Fuchs - Van de Graaf inequality, we get

$$\|E_{inv, V} \circ (I_S \otimes \phi_{P', U}) - U^\dagger \otimes \psi_{P, U}\|_\diamond \leq 2\sqrt{2}\epsilon.$$

(A11)

Now, we apply $V$ and $E_{inv, V}$ separately on two replicas of the system. Using Eqs. (A4) and (A11) as well as basic properties (the triangle inequality and the data processing inequality) of the diamond norm, we have

$$\left\| E_{inv, V} \circ \mathcal{P}_{S_1} \circ \mathcal{S}_{P, U} \circ \mathcal{P}_{S_2} \circ \mathcal{S}_{P, U} \otimes \mathcal{U} \right\|_\diamond \leq 4\sqrt{2}\epsilon,$$

(A12)

where the superscript in $\mathcal{S}_{P, U}$ indicates the registers that $V$ acts upon. Repeating this procedure for $m$ times and discarding the program in the end, we get a cascade of channels which acts on $2m$ replicas $S_1, S_2, \ldots, S_{2m}$ of the system:

$$\tilde{N}_U(\cdot) := T_{P, U} \circ E_{inv, V} \circ \mathcal{P}_{S_1} \circ \mathcal{S}_{P, U} \circ \mathcal{P}_{S_2} \circ \mathcal{S}_{P, U} \circ \cdots \circ E_{inv, V} \circ \mathcal{P}_{S_1} \circ \mathcal{S}_{P, U} (\cdot) \otimes \psi_{P, U}$$

(A13)

whose distance from $m$ uses of the unitary channel $U \otimes U^\dagger$ is bounded as

$$\left\| \tilde{N}_U - (U \otimes U^\dagger)^\otimes m \right\|_\diamond \leq 4m\sqrt{2}\epsilon.$$  

(A14)

For simplicity of calculation, we now discard half of the systems $\{S_{2j}\}_{j=1}^m$ in the above formula, obtaining

$$\left\| N_U - U^\otimes m \right\|_\diamond \leq 4m\sqrt{2}\epsilon.$$  

(A15)
We argue that the program state has to contain almost the same amount of information as χ [35], defined for an ensemble of quantum states.

This concludes the first part of the proof. Observe that, on one hand, all information in $\mathcal{M}_U$ on $U$ comes from the program state; on the other hand, $\mathcal{M}_U$ is $(4m\sqrt{e})$-close to $m$ uses of $U$. By comparing the amount of information, we argue that the program state has to contain almost the same amount of information as $m$ uses of $U$, for any $m \ll 1/\sqrt{e}$.

Next, we make the above argument quantitative. As a measure of information, we consider the Holevo information $\chi$, defined for an ensemble of quantum states $\{\rho_x, dx\}_{x \in X}$ as

$$I_H(\{\rho_x, dx\}) := H\left(\int_{x \in X} dx \rho_x\right) - \int_{x \in X} dx H(\rho_x)$$

(A17)

where $H$ denotes the von Neumann entropy.

Now let us derive an upper bound of the Holevo information of the program. Consider inputting an arbitrary state $\Phi_m$ to $\mathcal{M}_U$. Notice that $\chi$ is non-increasing under data processing on the system side. We get

$$I_H(\{\psi_{P,U}, dU\}) = I_H(\{\psi_{P,U} \otimes \Phi_m, dU\}) \geq I_H(\{\mathcal{M}_U(\Phi_m), dU\}),$$

(A18)

where $dU$ is the Haar measure of SU(d).

We choose $\Phi_m$ to maximise $I_H(\{U^\otimes m(\Phi_m), dU\})$. By the Schur-Weyl duality (see, e.g., Ref. [36]), the $m$-qudit Hilbert space can be decomposed as

$$\mathcal{H}^\otimes m \simeq \bigoplus_{\lambda \in S_m} \mathcal{H}_\lambda \otimes M_{\lambda,m},$$

(A19)

where $S_m := \{\lambda \in \mathbb{N}^d \mid |\lambda| := \sum_i \lambda_i = m, \lambda_i \geq \lambda_j \forall i < j\}$, each $\lambda$ is called a Young diagram, each $\mathcal{H}_\lambda$ is an irreducible subspace of SU(d) characterized by the Young diagram $\lambda$, and $M_{\lambda,m}$ is the corresponding multiplicity subspace. With this decomposition, $m$ parallel uses of $U \in SU(d)$ can also be decomposed as

$$U^\otimes m \simeq \bigoplus_{\lambda \in S_m} U_\lambda \otimes I_{\lambda,m},$$

(A20)

where $U_\lambda$ is the irreducible representation of SU(d) characterized by the Young diagram $\lambda$ and $I_{\lambda,m}$ is the identity of the corresponding multiplicity subspace.

To this end, we can define $\mathcal{T} : \mathcal{H}^\otimes m \to \bigoplus_{\lambda \in S_m} U_\lambda$ to be the quantum channel that first incorporates the isometry $\mathcal{H}^\otimes m \to \bigoplus_{\lambda \in S_m} U_\lambda \otimes M_{\lambda,m}$ and then discards the multiplicity parts $\{M_{\lambda,m}\}$. Since $U^\otimes m$ is invariant on the multiplicity subspace, we have

$$I_H(\{\mathcal{T} \circ U^\otimes m(\Phi_m), dU\}) = I_H(\{U^\otimes m(\Phi_m), dU\})$$

(A21)

for any $\Phi_m$. The point of applying $\mathcal{T}$ is that the dimension is reduced from $d^2m$ to

$$d_m := \sum_{\lambda \in S_m} d_{\lambda}^2$$

(A22)

with $d_\lambda$ being the dimension of $\mathcal{H}_\lambda$. $d_m$ grows only polynomially instead of exponentially in $m$. Explicitly, we have the following lemma, whose proof is provided in Appendix C:

**Lemma 1.** For $d \geq 2$, we have

$$d_m \geq \left(\frac{m}{(2d - 1)(d - 1)}\right)^{d-1}.$$  

(A23)

We then take $\Phi_m$ to be

$$|\Phi_m\rangle := \bigoplus_{\lambda \in S_m} \frac{d_\lambda}{\sqrt{d_m}} \left(|\Phi_\lambda^+\rangle \otimes |\psi_0\rangle\right)$$

(A24)
where $|\Phi^+_\lambda\rangle \in \mathcal{H}_\lambda \otimes \mathcal{H}_\lambda$ is the maximally entangled state and $|\psi_0\rangle$ is an arbitrary state in the multiplicity spaces. This choice of $\Phi_m$ achieves the maximum Holevo information

$$I_H \left( \{ T \circ U^\otimes m(\Phi_m), d U \} \right) = \log d_m. \tag{A25}$$

Therefore, we have

$$I_H \left( \{ \psi_{P,U}, d U \} \right) \geq I_H \left( \{ U^\otimes m(\Phi_m), d U \} \right)$$

$$\geq I_H \left( \{ U^\otimes m(\Phi_m), d U \} \right) - (4m\sqrt{2}\epsilon) \log d_m - 1 \tag{A26}$$

$$= I_H \left( \{ T \circ U^\otimes m(\Phi_m), d U \} \right) - (4m\sqrt{2}\epsilon) \log d_m - 1, \tag{A27}$$

having used Eq. (A15) and the Fannes-Alicki-Winter inequality [37, 38] to get the second inequality. Taking into account the bound $\log d_P \geq I_H \left( \{ \psi_{P,U}, d U \} \right)$, the inequality (A28) becomes

$$\log d_P \geq (1 - 4m\sqrt{2}\epsilon) \log d_m - 1. \tag{A29}$$

For an arbitrarily $\epsilon$-independent parameter $\delta > 0$, we choose

$$m = \left\lceil \frac{\delta}{4\sqrt{2}\epsilon} \right\rceil, \tag{A30}$$

where $\lceil \cdot \rceil$ denotes the ceiling function. Substituting this choice of $m$ into the bound, we get

$$d_P \geq \frac{1}{2} \left( \frac{\delta}{4\sqrt{2}\epsilon(2d - 1)(d - 1)} \right)^{(1 - \delta - 4\sqrt{2}\epsilon)(d^2 - 1)}. \tag{A31}$$

With this, we conclude that, for any $\alpha < (d^2 - 1)/2$, we have

$$d_P = O \left( \frac{1}{\epsilon^\alpha} \right). \tag{A32}$$

The bound in Eq. (A31) decreases with $d$, which seems to contradict the intuition that it is more demanding to program a larger unitary gate. In fact, we can show an alternative bound that increases with $d$ by modifying a few steps in the above proof.

First, observe that Lemma 1 does not work very well in the case of fixed $m$ and large $d$. Instead of using Lemma 1 in Eq. (A29), we employ directly the definition (A22):

$$\log d_P \geq (1 - 4m\sqrt{2}\epsilon) \log \left( \sum_{\lambda \in S_m} d^2_\lambda \right) - 1 \tag{A33}$$

$$\geq (1 - 4m\sqrt{2}\epsilon) \log \left( d_{(m,0,...,0)} \right)^2 - 1 \tag{A34}$$

$$= 2(1 - 4m\sqrt{2}\epsilon) \log \left( \frac{d + m - 1}{m - 1} \right) - 1. \tag{A35}$$

Suppose $\epsilon$ is small enough, e.g., $\epsilon = 0.001$. We take $m$ to be the closest integer to $1/(8\sqrt{2}\epsilon)$ in the above inequality, which becomes

$$\log d_P \geq \log \left( \frac{d + 1/(8\sqrt{2}\epsilon) - 1}{1/(8\sqrt{2}\epsilon) - 1} \right) - 1. \tag{A36}$$

Therefore, we get an alternative bound

$$d_P \geq \frac{1}{2} \left( \frac{d + 1/(8\sqrt{2}\epsilon) - 1}{1/(8\sqrt{2}\epsilon) - 1} \right) \sim d_{1/8\sqrt{2}\epsilon}^{-1}. \tag{A37}$$
Appendix B: Proof of Theorem 2

In this section we prove Theorem 2 on the performance of qudit gate estimation. The estimation task consists of two steps: The first step is to prepare a suitable probe state $|\psi\rangle$ and then to apply $n$ parallel uses of $U$ on it. The second step is to measure the resultant state, denoted by $|\psi_{U,n}\rangle$, with a suitable POVM $\{M_\hat{U}\}_{\hat{U}\in SU(d)}$, which outputs an estimate $\hat{U}$ of $U$.

Here we measure the performance of unitary gate metrology by the diamond norm error (2):

$$\epsilon := \sup_{U \in SU(d)} \frac{1}{2} \| U - E_{mo,U} \|_\diamond. \quad (B1)$$

Here $E_{mo,U}$ is the measure-and-operate (MO) channel

$$E_{mo,U}(\cdot) := \int d\hat{U} p(\hat{U}|U) \hat{U}(\cdot), \quad (B2)$$

where $p(\hat{U}|U)$ is the probability of getting the estimate $\hat{U}$ (when the actual gate is $U$) defined as

$$p(\hat{U}|U) := \text{Tr} [\psi_{U,n} M_\hat{U}] . \quad (B3)$$

We remark that the performance of unitary gate metrology can also be characterised by other figures of merit, e.g., the (average) gate fidelity [39, 40]. Here we are using a more demanding error measure.

The proof can be sketched as the following:

1. We first measure the performance of estimation protocols using the entanglement fidelity [41]:

$$F_{ent}(A, B) := F \left( (A \otimes I_R) (|\Phi^+\rangle\langle\Phi^+|) , (B \otimes I_R) (|\Phi^+\rangle\langle\Phi^+|) \right), \quad (B4)$$

where $|\Phi^+\rangle$ is the maximally entangled state of the system $S$ and a reference $R \simeq S$. In general, $F_{ent}$ serves as an upper bound on $F_{wc}$ and is easier to evaluate.

2. We derive a formula of $F_{ent}$ for a class of estimation protocols, which include the optimal protocol that achieves the maximum of $F_{ent}$ over all protocols. The optimal protocol and its $F_{ent}$ can be evaluated numerically from the formula.

3. Next, we show that, for the above class of protocols, $\epsilon \leq d \cdot (1 - F_{ent})$.

4. We fix an estimation protocol and prove that it achieves the performance $F_{ent} \geq 1 - c_d/n^2$. Combining with the point above, we obtain an upper bound on $\epsilon$ in terms of $n$.

5. We also determine, for the same protocol, the relation between the dimension of the probe and $n$.

1. A formula for $F_{ent}$

In this subsection, we focus first on the entanglement fidelity $F_{ent}$. Before starting, we recall a few concepts from the Schur-Weyl decomposition [cf. Eq. (A19)]. We will make frequent uses of the Young diagrams $\lambda = (\lambda_1, \lambda_2, \ldots)$ and the irreducible representation $U_\lambda$ characterised by the Young diagram $\lambda$ [see Eq. (A20)]. In particular, we define $e_i$ to be the vector whose $i$-th entry is one and other entries are zero. By definition, $e_1$ corresponds to a legitimate Young diagram whose associated representation is the $d$-dimensional self-representation, and we use the abbreviation $U := U_{e_1}$. We will use the double-ket notation $|A\rangle := \sum_{n,m} (n|A|m) |n\rangle\langle m|$ ({$|n\rangle$} being an orthonormal basis) for a matrix $A$ and denote by $|\Phi^+_{U,\lambda}\rangle$ the maximally entangled state $|U_\lambda\rangle/\sqrt{d_\lambda}$.

To maximise the entanglement fidelity of metrology, it is enough to consider probe states of the form [42, Theorem 1]

$$|\psi\rangle = \bigoplus_{\lambda \in S_{\text{Young}}} \sqrt{q_\lambda} |\Phi^+_{\lambda}\rangle \otimes |\Phi^+_m\rangle, \quad (B5)$$
Here $\mathcal{S}_{\text{Young}} \subset \mathcal{S}_n$ is a suitable set containing Young diagrams of $n$ boxes, $|\Phi^+\rangle$ denotes the maximally entangled state (of the corresponding Hilbert spaces), and $\{q_\lambda\}$ is a suitable probability distribution. We assume that any Young diagram $\lambda \in \mathcal{S}_{\text{Young}}$ has strictly decreasing row numbers. After the application of $U \otimes^n$, the probe state is in the form

$$|\psi_{U,n}\rangle = \bigoplus_{\lambda \in \mathcal{S}_{\text{Young}}} \sqrt{q_\lambda} |\Phi^+_{U,\lambda}\rangle \otimes |\Phi^+_{m,\lambda}\rangle.$$  \hspace{1cm} (B6)

The optimal measurement [42] is the covariant POVM $\{d \hat{U}, M_U\}$ with $d \hat{U}$ being the Haar measure and

$$M_U := |\eta_U \rangle \langle \eta_U| \quad |\eta_U\rangle := \bigoplus_{\lambda \in \mathcal{S}_{\text{Young}}} d_\lambda |\Phi^+_{U,\lambda}\rangle \otimes |\Phi^+_{m,\lambda}\rangle.$$  \hspace{1cm} (B7)

Denoting by $\chi_{U,\lambda} := \text{Tr} [U_\lambda]$ the characters of $SU(d)$, the probability of getting the outcome $\hat{U}$ when the actual gate is $U$ can be expressed as

$$p(\hat{U}|U) = \left| \sum_{\lambda \in \mathcal{S}_{\text{Young}}} \sqrt{q_\lambda} \chi_{U \hat{U}^{-1}, \lambda} \right|^2.$$  \hspace{1cm} (B8)

We can then express the entanglement fidelity as

$$F_{\text{ent}} (\mathcal{E}_{\text{mo},U}, U) = \inf_{U' \in SU(d)} \frac{1}{d^2} \int d \hat{U} \left| \sum_{\lambda \in \mathcal{S}_{\text{Young}}} \sqrt{q_\lambda} \chi_{U \hat{U}^{-1}, \lambda} \right|^2.$$  \hspace{1cm} (B9)

where $\chi_{U \hat{U}^{-1}} := \chi_{U \hat{U}^{-1}, e_\lambda^1}$ is the character of the self-representation $e_1$. To proceed, we decompose the characters as

$$\chi_{U \hat{U}^{-1}, \lambda} = \sum_{\lambda' \in O_1(\lambda)} \chi_{U \hat{U}^{-1}, \lambda'},$$  \hspace{1cm} (B10)

where $O_1(\lambda) := \{ \lambda + e_i \mid i : \lambda_i < \lambda_{i-1} \}$. Using the group invariance property of the Haar measure and the orthogonality of characters, we have

$$F_{\text{ent}} (\mathcal{E}_{\text{mo},U}, U) = \inf_U \frac{1}{d^2} \int d \hat{U} \left| \sum_{\lambda, \lambda' \in \mathcal{S}_{\text{Young}}} \sqrt{q_\lambda q_{\lambda'}} \chi_{U \hat{U}^{-1}, \lambda'} \right|^2.$$  \hspace{1cm} (B11)

$$= \inf_U \frac{1}{d^2} \left( \sum_{\lambda, \lambda' \in \mathcal{S}_{\text{Young}}} \sqrt{q_\lambda q_{\lambda'}} \sum_{\lambda' \in O_1(\lambda), \lambda' \in O_1(\lambda)} \int d \hat{U} \chi_{U \hat{U}^{-1}, \lambda'} \chi_{U \hat{U}^{-1}, \lambda'} \right)^2.$$  \hspace{1cm} (B12)

$$= \frac{1}{d^2} \left( \sum_{\lambda, \lambda' \in \mathcal{S}_{\text{Young}}} \sqrt{q_\lambda q_{\lambda'}} \sum_{\lambda' \in O_1(\lambda), \lambda' \in O_1(\lambda)} \delta_{\lambda, \lambda'} \right)^2.$$  \hspace{1cm} (B13)

Rearranging terms, we have

$$F_{\text{ent}} (\mathcal{E}_{\text{mo},U}, U) = \frac{1}{d^2} \sum_{\lambda' \in S_{\text{ent}}} \left( \sum_{\lambda' \in O_{\lambda'}} \sqrt{q_\lambda} \right)^2,$$  \hspace{1cm} (B14)

where $O_{\lambda'} := \{ \lambda \in \mathcal{S}_{\text{Young}} \mid \exists i : \lambda_i = \lambda_i + e_i \}$. Equivalently, the entanglement fidelity can be expressed as

$$F_{\text{ent}} (\mathcal{E}_U, U) = \frac{1}{d^2} (\tilde{q}^T S \tilde{q})_1,$$  \hspace{1cm} (B15)

where $\tilde{q}$ is a unit vector (i.e. $\tilde{q} \cdot \tilde{q} = 1$) supported by $\mathcal{S}_{\text{Young}}$ and $S$ is the score matrix defined by

$$S_{\lambda, \lambda'} := \begin{cases} d \quad &d_{\text{Young}}(\lambda, \lambda') = 0 \\ 1 \quad &d_{\text{Young}}(\lambda, \lambda') = 2 \\ 0 \quad &\text{else} \end{cases}.$$  \hspace{1cm} (B16)
Here \( d_{\text{Young}}(\lambda, \lambda') := \sum_i |\lambda_i - \lambda'_i| \) is a distance measure between Young diagrams. Summarizing the above derivation, we have shown that:

**Lemma 2.** Assume that any Young diagram \( \lambda \in S_{\text{Young}} \) has strictly decreasing row numbers. The entanglement fidelity of the optimal estimation is given by the optimization in Eq. (B15).

The same result, in a slightly different form, was first obtained by Kahn [22]. We remark that, though the optimal estimation performance is just the maximum eigenvalue of \( S \) (B16), it is not easy to show the \( 1/n^2 \) error scaling. The matrix \( S \) is a banded multilevel Toeplitz matrix, whose eigensystem problem remains open to the best of our knowledge (see, e.g., Ref. [43]).

### 2. Switching between the diamond norm error \( \epsilon \) and \( F_{\text{ent}} \) for covariant protocols

Here we show that for any covariant estimation protocol, defined as follows, it is enough to evaluate the entanglement fidelity:

**Definition 1 (Covariant estimation protocols).** An estimation protocol \( (\psi, \{M_U\}) \) is covariant if the probability distribution (B3) of the estimate satisfies

\[
p(W\hat{U}V^\dagger|WUV^\dagger) = p(\hat{U}|U) \quad \forall W, V \in SU(d).
\]

(B17)

One can directly check that protocols mentioned in the previous subsection, whose \( p(\hat{U}|U) \) has the form (B8), are covariant. For covariant protocols, the channel \( \mathcal{E}_{\text{mo},U} \) is covariant when \( U = I \), and we have the following lemma:

**Lemma 3.** For any covariant estimation protocol, the following bound holds

\[
\epsilon \leq d \cdot (1 - F_{\text{ent}}(\mathcal{E}_{\text{mo},I}, I)).
\]

(B18)

Therefore, it is enough to consider the quantity \( F_{\text{ent}}(\mathcal{E}_{\text{mo},I}, I) \).

**Proof of Lemma 3.** Applying Eq. (B17) we have

\[
\mathcal{E}_{\text{mo},U} \circ U' = \int d\hat{U} p(\hat{U}|U) \hat{U} \circ U'
\]

(B19)

\[
= \int d\hat{U} p(\hat{U}|I) \hat{U}_0 = \mathcal{E}_{\text{mo},I}.
\]

(B20)

Therefore, by unitary invariance of the diamond norm, we have \( \epsilon = \frac{1}{2} \| \mathcal{E}_{\text{mo},I} - I \|_{\diamond} \). What remains is to relate the diamond norm \( \| \mathcal{E}_{\text{mo},I} - I \|_{\diamond} \) to the entanglement fidelity \( F_{\text{ent}}(\mathcal{E}_{\text{mo},I}, I) \). Indeed, we have

\[
\mathcal{E}_{\text{mo},U} \circ U' = \int d\hat{U} p(\hat{U}|U) \hat{U} \circ U'
\]

(B22)

\[
= \int d(U'\hat{U}_1U'^\dagger)p(U'\hat{U}_1U'^\dagger|I) U' \circ \hat{U}_1 = U'^\dagger \hat{U} U'
\]

(B23)

\[
= U' \circ \mathcal{E}_{\text{mo},U'U'^\dagger}(U'^\dagger)
\]

(B24)

for any \( U' \). Taking \( U \) to be the identity, it is immediate that \( \mathcal{E}_{\text{mo},I} \) is covariant with respect to \( SU(d) \), i.e. \( \mathcal{E}_{\text{mo},I} \circ U' = U' \circ \mathcal{E}_{\text{mo},I} \). For covariant channels, we have the following general result:

For any quantum channel \( \mathcal{A} \) acting on a \( d \)-dimensional system, define its Choi state as

\[
\mathcal{A} := \left( \mathcal{A} \otimes I \right)(\Phi^+)
\]

(B25)

with \( \Phi^+ \) being the maximally entangled state in \( \mathcal{H} \otimes \mathcal{H} \). When \( \mathcal{A} \) is covariant, we have

\[
[A, U \otimes U^*] = 0, \quad \forall U \in SU(d).
\]

(B26)
By Schur’s lemma, the Choi state of a covariant channel $A$ can be decomposed as
\[ A = (1-a) \cdot \Phi^+ + a \cdot \rho^\perp \quad \rho^\perp := \frac{1}{d^2-1} \left( I \otimes I - \Phi^+ \right) \] (B27)
for some $a \in [0,1]$. It follows immediately from the above expression that
\[ 1 - F_{\text{ent}}(A, \mathcal{I}) = \frac{1}{2} \| A - \Phi^+ \|_1, \] (B28)
which is the trace distance error between the Choi state of $A$ and the maximally entangled state. Combining with the inequality $\| A - \mathcal{I} \|_0 \leq d \cdot \| A - \Phi^+ \|_1$ (see, e.g., [44, Exercise 3.6]), we get
\[ \frac{1}{2} \| A - \mathcal{I} \|_0 \leq d \cdot (1 - F_{\text{ent}}(A, \mathcal{I})) \] (B29)
as desired.

\[ \square \]

3. Proof of Eq. (5)

Now, we show that there exists a covariant protocol with worst-case fidelity given by Eq. (5). Due to Lemma 3, it is enough to show the bound for the entanglement fidelity.

The covariant estimation protocol we are going to discuss is of the structure described previously: Its input state is of the form (B5), its POVM is given by Eq. (B7), and its entanglement fidelity is given by Eq. (B15). What remains to be done is to specify the distribution \{q_s\}.

For this purpose, we first define a parameter $N$ that depends on $n$ as
\[ N = \left\lfloor \frac{1}{(3d - 2)} \left( \frac{2n}{d - 1} + d - 2 \right) \right\rfloor \] (B30)
and $n_0 := n - ((3d - 2)N - d + 2)(d - 1)/2$. By definition, $N$ is bounded as
\[ N \in \left[ c_{\text{min},d} \cdot n, c_{\text{max},d} \cdot n \right], \quad c_{\text{min},d} := \frac{2 \left( 1 - \frac{d(d-1)}{n} \right)}{(3d - 2)(d - 1)} \quad c_{\text{max},d} := \frac{2 \left( 1 + \frac{(d-2)(d-1)}{2n} \right)}{(3d - 2)(d - 1)} \] (B31)
with $c_{\text{min},d}$ and $c_{\text{max},d}$ depending only on $d$ when $n \to \infty$. Define $\mu_0 \in S_{n_0}$ as the most flat Young diagram with $n_0$ boxes:
\[ \mu_0 := (\mu_{0,1}, \mu_{0,2}, \ldots, \mu_{0,d}) \quad \text{s.t.} \quad \sum_i |\mu_{0,i}| = n_0 \quad \text{and} \quad \mu_{0,j} + 1 \geq \mu_{0,i} \geq \mu_{0,j} \quad \forall j > i. \] (B32)

Now we define the following viable subset of Young diagrams with $d$ rows and $n$ boxes, on which our probe state has support:
\[ S_{\text{Young}} := \left\{ \lambda \in S_n \mid \lambda_i = \mu_{i,0} + N(2d - 3) + 1 - (N + 1)(i - 1) + \check{\lambda}_i, \forall i \leq d - 1 \quad \exists \check{\lambda} \in [N - 1]^{(d-1)} \right\}. \] (B33)

Obviously, the above definition satisfies the assumption that any Young diagram $\lambda \in S_{\text{Young}}$ has strictly decreasing row numbers. This choice is to minimise the boundary set, which contains those elements of $S_{\text{Young}}$ with some of their adjacent (i.e. $d_{\text{Young}} = 2$) Young diagrams not in the set. One can see from Eq. (B16) that this makes the score higher. Moreover, as shown later, dimensions of elements in $S_{\text{Young}}$ are easy to bound.

Each Young diagram in $S_{\text{Young}}$ is now uniquely characterized by $\check{\lambda} \in [N - 1]^{(d-1)}$, so from now on we use $\check{\lambda}$ as the notion for Young diagrams. Note that the relevant elements of $S$ for the Young diagrams we consider are
\[ S(\check{\lambda}, \check{\lambda}') = \begin{cases} 1 & \check{\lambda} - \check{\lambda}' = \pm f_{ij} \quad \exists i > j \\ 1 & \check{\lambda} - \check{\lambda}' = \pm e_i \quad \exists i \\ d & \check{\lambda} = \check{\lambda}' \end{cases}. \] (B34)
Here \( \{e_i\}_{i=1}^{d-1} \) is the natural basis of \([N-1]^{(d-1)}\), and \( f_{ij} := e_i - e_j \). We denote by \( g_k \) the following distribution over \([N-1]^{(d-1)}\)

\[
g_k := \frac{2}{N} \sin^2 \left( \frac{\pi (2k+1)}{2N} \right) \tag{B35}
\]

and by \( \epsilon_g \) the quantity

\[
\epsilon_g := 1 - \sum_{k=0}^{N-2} \sqrt{g_k g_{k+1}} \leq \frac{\pi^2}{N^2}. \tag{B36}
\]

The inequality can be shown by straightforward calculation. Consider the product form distribution

\[
q^{*}_{\lambda} := \prod_{i=1}^{d-1} g_{\lambda_i}, \tag{B37}
\]

where \( g \) is the distribution defined in Eq. (B35).

Now, we show that the covariant protocol, specified by Eq. (B37), has entanglement fidelity as follows.

**Lemma 4.** The entanglement fidelity of the protocol specified by Eq. (B37) is lower bounded as

\[
F_{\text{ent}} (\mathcal{E}^{*}_{\text{mo}, j}, \mathcal{I}) \geq 1 - 2 \left( \frac{\pi (d-1)}{d \cdot c_{\text{min}, d} \cdot n} \right)^2, \tag{B38}
\]

where \( c_{\text{min}, d} \) is given in Eq. (B31).

**Proof of Lemma 4.** By definition (B37), we have

\[
q^{*} T S q^{*} = \sum_{\lambda, \lambda' \in [N-1]^{(d-1)}} \sqrt{q^{*}_{\lambda} S(\lambda, \lambda')} \sqrt{q^{*}_{\lambda'}} \tag{B39}
\]

\[
= \sum_{\lambda \in [N-1]^{(d-1)}} \sum_{i \neq j} \sqrt{q^{*}_{\lambda + f_{ij}}} + \sum_{\lambda \in [N-1]^{(d-1)}} \sum \sqrt{q^{*}_{\lambda + e_i}} + \sum_{\lambda \in [N-1]^{(d-1)}} \sum \sqrt{q^{*}_{\lambda - e_i}} + d \tag{B40}
\]

\[
= \sum_{i \neq j} \sum_{\lambda \in [N-1]^{(d-1)}} \sqrt{q^{*}_{\lambda + f_{ij}}} + \sum_{i} \left( \sum_{\lambda \in [N-1]^{(d-1)}} \sqrt{q^{*}_{\lambda + e_i}} + \sum_{\lambda \in [N-1]^{(d-1)}} \sqrt{q^{*}_{\lambda - e_i}} \right) + d. \tag{B41}
\]

For an arbitrary pair of \((i, j)\) such that \(i \neq j\), using Eqs. (B35), (B36) and (B37) we can explicitly evaluate the term in the first summation as

\[
\sum_{\lambda \in [N-1]^{(d-1)}} \sqrt{q^{*}_{\lambda + f_{ij}}} = \left( \sum_{\lambda_i = 0}^{N-2} \sqrt{g_{\lambda_i} g_{\lambda_i+1}} \right) \left( \sum_{\lambda_j = 1}^{N-1} \sqrt{g_{\lambda_j} g_{\lambda_j-1}} \right) = (1 - \epsilon_g)^2. \tag{B42}
\]

Similarly, for arbitrary \(i\), the term in the second and summation can be expressed as

\[
\sum_{\lambda \in [N-1]^{(d-1)}} \sqrt{q^{*}_{\lambda + e_i}} + \sum_{\lambda \in [N-1]^{(d-1)}} \sqrt{q^{*}_{\lambda - e_i}} = 2(1 - \epsilon_g). \tag{B43}
\]

Substituting the above back into Eq. (B41), we have

\[
q^{*} T S q^{*} = d + (d-1)(d-2)(1 - \epsilon_g)^2 + 2(d-1)(1 - \epsilon_g) \geq d^2 - 2(d-1)^2 \epsilon_g. \tag{B44}
\]

Combining the above inequality with Lemma 2 and Eqs. (B31) and (B36), we get the bound

\[
F_{\text{ent}} (\mathcal{E}^{*}_{\text{mo}, j}, \mathcal{I}) \geq 1 - \frac{2\pi^2 (d-1)^2}{d^2} \left( \frac{1}{N} \right)^2 \tag{B46}
\]

\[
= 1 - 2 \left( \frac{\pi (d-1)}{d \cdot c_{\text{min}, d} \cdot n} \right)^2 \tag{B47}
\]

as desired. □
4. Proof of Eq. (6)

We conclude our proof of Theorem 2 by showing Eq. (6), which is a bound on the dimension of the probe state (B37).

**Lemma 5.** The probe state specified by Eq. (B37) has dimension bounded as

\[
d_P \leq (2(d-1)c_{\text{max},d} \cdot n + 3)^{d^2-1},
\]

where \(c_{\text{max},d}\) is given in Eq. (B31).

**Proof of Theorem 2.** Finally, putting together all ingredients (Lemmas 3, 4, and 5) yields Theorem 2. We also used the bounds \(c_{\text{min},d} \geq \frac{1}{3(d-2)(d-1)}\), \(c_{\text{max},d} \leq \frac{3}{3(d-2)(d-1)}\) and \(3 \leq 3n/(3d-2)\), which come from the assumptions on \(n\) and \(d\), to simplify the expressions. \(\square\)

**Proof of Lemma 5.** The irreducible representation \(\lambda\) of SU(d) has dimension [45, Eq. (III.10)]

\[
d_\lambda = \prod_{1 \leq i < j \leq d} (\lambda_i - \lambda_j - i + j) / \prod_{1 \leq i \leq k < d} k!
\]

The viable set \(S_{\text{Young}}\) [cf. Eq. (B33)] is so defined that, for any \(\lambda \in S_{\text{Young}}\) and any \(i < j\),

\[
\lambda_i - \lambda_j \leq \begin{cases} 
N(j - i + 1) + (2j - 2i - 1) & j < d \\
N(2j - i - 1) + (j - 2i + 1) & j = d
\end{cases}
\]

Therefore, using Eq. (B31), for any \(\hat{\lambda} \in S_{\text{Young}}\), its dimension is upper bounded by

\[
d_{\hat{\lambda}} \leq \left( \prod_{1 \leq i < j < d} (\lambda_i - \lambda_j - i + j) \right) \left( \prod_{1 \leq i < d} (\lambda_i - \lambda_d - l + d) \right)
\]

\[
\leq C_{\text{max},d} (c_{\text{max},d} \cdot n)^{\frac{d(d-1)}{2}}.
\]

Here

\[
C_{\text{max},d} := \left( \prod_{1 \leq i < j < d} \left( j - i + 1 + \frac{3(d-1)}{c_{\text{max},d} \cdot n} \right) \right) \left( \prod_{1 \leq l < d} \left( 2d - l - 1 + \frac{2(d-1)}{c_{\text{max},d} \cdot n} \right) \right)
\]

\[
\leq \left( 2(d-1) + \frac{3}{c_{\text{max},d} \cdot n} \right)^{\frac{d(d-1)}{2}}.
\]

Since \(|S_{\text{Young}}| = N^{d-1}\), we have

\[
d_P = \sum_{\lambda \in S_{\text{Young}}} d_\lambda^2
\]

\[
\leq \left( 2(d-1) + \frac{3}{c_{\text{max},d} \cdot n} \right)^{\frac{d(d-1)}{2}} (c_{\text{max},d} \cdot n)^{d^2-1}
\]

\[
\leq (2(d-1)c_{\text{max},d} \cdot n + 3)^{d^2-1}.
\]

\(\square\)

**Appendix C: Proof of Lemma 1**

To prove Lemma 4 we explicitly constructed a probe state on a viable set of Young diagrams. Similarly, we are going to construct another viable set \(S'_{\text{Young}} \subset S_m\). Then, by definition, the dimension

\[
d_{\text{Young}} := \sum_{\lambda \in S'_{\text{Young}}} d_\lambda^2
\]

(C1)
lower bounds $d_m$. By deriving a lower bound on $d_{\text{Young}}$, we get Lemma 1.

We first define a parameter $M$ that depends on $m$ as

$$M = \left\lfloor \frac{m}{(2d-1)(d-1)} \right\rfloor$$

and $m_0 := m - (2d - 1)(d - 1)M$. By definition, we have the lower bound

$$M + 1 \geq c'_d \cdot m, \quad c'_d := \frac{1}{(2d-1)(d-1)}.$$  \hspace{1cm} (C3)

Define $\mu'_0 \in S_{m'_0}$ as the most flat Young diagram with $m'_0$ boxes [cf. Eq. (B32)]. Now we define the following viable subset of Young diagrams with $d$ rows and $m$ boxes, on which our probe state has support:

$$S'_{\text{Young}} := \left\{ \lambda \in S_m \mid \lambda_i = \mu'_{i,0} + M(3d - 2 - 2i) + \tilde{\lambda}'_i, \forall i \leq d - 1 \quad \exists \tilde{\lambda}' \in [M]^{(d-1)} \right\}.$$  \hspace{1cm} (C4)

By definition, elements of $S'_{\text{Young}}$ is uniquely determined by $\tilde{\lambda}' \in [M]^{(d-1)}$. The viable set $S'_{\text{Young}}$ [cf. Eq. (C4)] is so defined that, for any $\lambda \in S'_{\text{Young}}$ and any $i < j$,

$$\lambda_i - \lambda_j \geq M(j - i).$$  \hspace{1cm} (C5)

Therefore, using Eqs. (B49) and (C3), for any $\lambda \in S'_{\text{Young}}$, its dimension is upper bounded by

$$d_\lambda = \frac{\prod_{1 \leq i < j \leq d}(\lambda_i - \lambda_j - i + j)}{\prod_{k=1}^{d-1} k!} \geq \frac{\prod_{1 \leq i < j \leq d}(j - i)}{\prod_{k=1}^{d-1} k!} (M + 1)^{\frac{d(d-1)}{2}} = (c'_d m)^{\frac{d(d-1)}{2}}.$$  \hspace{1cm} (C6)

Since $|S'_{\text{Young}}| = (M + 1)^{d-1}$, we have

$$d_{\text{Young}} = \sum_{\lambda \in S'_{\text{Young}}} d_\lambda^2 \geq (c'_d m)^{d(d-1)} (c'_d m)^{d-1} \geq (c'_d m)^{d^2-1}.$$  \hspace{1cm} (C7)

Since $d_m \geq d_{\text{Young}}$, the same bound holds for $d_m$. \hspace{1cm} \Box