Extreme Dimensionality Reduction with Quantum Modeling

Thomas J. Elliott,[1,2,3,*] Chengran Yang,[3,2,†] Felix C. Binder,[4] Andrew J. P. Garner,[4,5] Jayne Thompson,[5] and Mile Gu[3,2,5,‡]

[1]Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom
[2]Complexity Institute, Nanyang Technological University, Singapore 637335, Singapore
[3]Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore
[4]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, Vienna 1090, Austria
[5]Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore

(Received 16 May 2020; accepted 23 October 2020; published 22 December 2020)

Effective and efficient forecasting relies on identification of the relevant information contained in past observations—the predictive features—and isolating it from the rest. When the future of a process bears a strong dependence on its behavior far into the past, there are many such features to store, necessitating complex models with extensive memories. Here, we highlight a family of stochastic processes whose minimal classical models must devote unboundedly many bits to tracking the past. For this family, we identify quantum models of equal accuracy that can store all relevant information within a single two-dimensional quantum system (qubit). This represents the ultimate limit of quantum compression and highlights an immense practical advantage of quantum technologies for the forecasting and simulation of complex systems.

DOI: 10.1103/PhysRevLett.125.260501

Predicting the future based on past events is a cornerstone of life. From meteorologists forecasting the weather, through investors trading on stock markets, to a predator chasing its prey, the ability to identify causes and accurately anticipate effects is central to survival and success. To carry out these essential tasks, models must be formulated and information about past observations must be stored within memory.

In this context, processes with long historical dependence typically require models that store extensive information about past observations. This is because a model must ascribe each set of past causes that can give rise to distinct future effects to distinct configurations in its memory. When there are many such causes, the memory must support many configurations. Classically, the number of configurations is synonymous with the *dimension* of the memory—tracking a process with causes reaching far into the past typically requires a large memory with many dimensions.

In contrast, the number of configurations a quantum memory can take is separate from its dimension. This has led to quantum encodings with reduced memory dimension for several Markovian processes—where each output is conditional only on its immediate predecessor [1–4]. Here, we demonstrate that not only do these quantum advantages persist for non-Markovian processes, but that they become even more pronounced in this regime. We consider a family of such processes where the memory dimension required of a faithful classical model diverges with precision and identify corresponding quantum models that compress all configurations into two dimensions. This allows for all relevant history to be stored in a single two-state quantum system (qubit), evincing an extreme quantum advantage that scales without bound. Moreover, our protocol requires only a single probe qubit to extract the future statistics. This turns a problem from the converse scenario—that tracking a finite quantum system can require infinite classical resources [5–8]—into a useful tool.

This complements recent advances at the interface of complexity and quantum science, where it has been found that quantum models can drastically reduce the amount of past information—as measured by *information entropy*—that must be stored in memory to replicate the future behavior of a process [1,9–15]. Our Letter indicates that this advantage (along with its quantitative scaling divergences) also persists for the *memory dimension*. Crucially, this brings practical, verifiable, and significant quantum memory advantages within the reach of present technologies.

Framework and tools.—A stochastic process $\mathcal{X}$ can be characterized by an observation sequence $\overleftrightarrow{X}$, detailing what happens and when [16]. We can partition this sequence in two: a past $\overleftarrow{x}$ that describes everything that has happened up to the present, and a future $\vec{x}$ describing everything yet to come (we use upper case to denote random variables and lower case for their corresponding variates). The goal of causal modeling is to use the past (and only the past) to simulate the future [1,17–19]. Specifically, a causal model

M stores in its memory states $m \in \mathcal{M}$ determined from an encoding function of the past $f:\{\overleftarrow{x}\} \to \mathcal{M}$, such that it can produce futures $\vec{X}$ according to $P[\vec{X}|m = f(\overleftarrow{x})] = P(\vec{X}|\overleftarrow{x})$.

Two widely used metrics for a causal model's memory efficiency are [17] $C_{\mathsf{M}} \coloneqq -\sum_{m\in\mathcal{M}} P(m)\log_2[P(m)]$ and $D_{\mathsf{M}} \coloneqq \log_2[\dim(\mathcal{M})]$, where $P(m) = \sum_{\overleftarrow{x}\in m} P(\overleftarrow{x})$ is the probability of finding the memory in state m in the process' steady state. These measures, respectively, characterize the information stored by the memory and the dimension of the substrate into which it is encoded. Operationally, they represent the memory required to implement the model in an asymptotic ensemble (C_{M}) or single-shot (D_{M}) setting.

When $\overleftrightarrow{X}$ is a bi-infinite, stationary sequence with discrete events, the ε-machine of computational mechanics [17–19] is the provably most efficient classical causal model according to both these metrics. The corresponding minimal measures are labeled as C_μ and D_μ and referred to as the statistical and topological complexity, respectively [17]. The key elements of these models are causal states $s \in \mathcal{S}$, a set of equivalence classes defined such that, if two pasts have identical future predictions, the (causal state) memory encoding function $f_\varepsilon:\{\overleftarrow{x}\} \to \mathcal{S}$ assigns them to the same state: $f_\varepsilon(\overleftarrow{x}) = f_\varepsilon(\overleftarrow{x}') \Leftrightarrow P(\vec{X}|\overleftarrow{x}) = P(\vec{X}|\overleftarrow{x}')$. Causal states are in essence a state of knowledge, minimally encapsulating all information relevant to future prediction that can be obtained from observations of the past; they closely mirror the belief states of reinforcement learning [20,21]. They represent the minimal (classical) sufficient statistic of the past with respect to the future [18]. The ε-machine describes a stochastic transition structure between causal states, with transitions accompanied by the output of a symbol; this can be represented by a hidden Markov model [18]. These complexity measures have been applied to study structure in systems from a variety of fields, including neuroscience [22,23], biology [24,25], economics [26], geophysics [27], meteorology [28], and condensed matter physics [29].

These optimality results do not hold within the quantum domain [9]. For quantum causal models [1–4,9–15,30–37], each past $\overleftarrow{x}$ is assigned a quantum state $|f(\overleftarrow{x})\rangle$ to be stored in the model memory. The efficiency metrics become $C_q \coloneqq -\mathrm{Tr}[\rho\log_2(\rho)]$ and $D_q \coloneqq \log_2[\mathrm{rank}(\rho)]$, where $\rho = \sum_{\overleftarrow{x}} P(\overleftarrow{x})|f(\overleftarrow{x})\rangle\langle f(\overleftarrow{x})|$. We refer to these as the quantum statistical memory and quantum topological memory of a model, respectively; they inherit the same operational significance in the quantum regime as the corresponding classical quantities [9]. As with classical causal models, these quantum memory states encode information from the past of the process and must not contain any information that can only be obtained from its future; the full description of a quantum model then includes the means by which its memory is probed to produce a sample of the future statistics given the observed past, which must similarly be drawn from $P[\vec{X}|m = |f(\overleftarrow{x})\rangle] = P(\vec{X}|\overleftarrow{x})$. For definiteness, we remark that, while the model and its associated memory are quantum, the data (i.e., the modeled stochastic process) remain classical.

Current state-of-the-art constructions for quantum causal models [3] assign memory states directly from causal states $s \to |s\rangle$, though the optimal quantum encoding strategy is presently unknown for general processes [1,34]—we therefore do not designate these quantum metrics as complexity measures. Nevertheless, it has been shown that, in general, there exists a quantum model with $C_q \leq C_\mu$ [9]. This quantum advantage exploits the possibility to store quantum information in nonorthogonal states [38], enabling efficient isolation of predictive features. It has recently been shown that quantum models can also exhibit $D_q < D_\mu$ [1–4].

Dual Poisson processes.—Consider a system that undergoes a series of Poissonian decay events through one of two channels with rates γ_1 and γ_2. After each event, the decay channel for the next emission is chosen randomly, with probability p or $\bar{p} = 1 - p$, respectively. The choice of channel is hidden internally in the system, such that an external observer can only see when the decay events occur. Specifically, we consider an observer operating on discrete time steps Δt, recording a 1 when an event occurs and 0 otherwise. We call the resultant stochastic process a dual Poisson process, and it manifests as a series of 1s separated by strings of 0s. Note that the probabilistic choice of channel occurs only after events (1s) and remains unchanged across nonevents (0s). The probability that a contiguous string of 0s (bookended by 1s) is of at least length n is given by the so-called survival probability $\Phi(n)$,

$$\Phi(n) = p\Gamma_1^n + \bar{p}\Gamma_2^n, \tag{1}$$

where $\Gamma_j = \exp(-\gamma_j\Delta t)$. We shall now look at the scaling of the memory metrics of causal models for such processes as the temporal precision Δt is refined—making the process increasingly non-Markovian. With arbitrary $\Phi(n)$, this framework describes general renewal processes [39].

Optimal classical causal model.—Since the observer is unaware of the choice of decay channel, the information they must track reflects their confidence in the chosen rate based on the time since last emission. Let $\{\overleftarrow{n}\}$ denote clusterings of all pasts with the same number n of 0s since the last 1, and $\{\vec{n}\}$ denote clusters of futures with the same number n of 0s until the next 1. Then, a causal model of a dual Poisson process must track the number of 0s ($\overleftarrow{n}$) since the last 1 in order to predict how many more 0s ($\vec{n}$) until the next 1 appears; the direction of the arrows signifies that this is information about observations either in the past or in the future. The relevant conditional future distribution is given by

$$P(\vec{N} = \vec{n}|\overleftarrow{n}) = \frac{\Phi(\overleftarrow{n}+\vec{n}) - \Phi(\overleftarrow{n}+\vec{n}+1)}{\Phi(\overleftarrow{n})}. \tag{2}$$

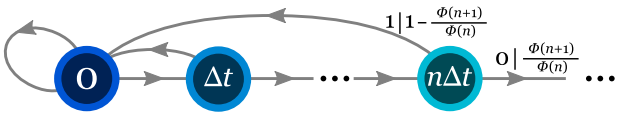

FIG. 1. Tracking dual Poisson processes. Causal models of dual Poisson processes track the confidence in chosen emission rate based on the time since last emission; the number of possible states diverges with refinement of time steps. Since all states have different future distributions they each correspond to different causal states—the model depicted is the ϵ-machine. The notation $x|T$ indicates that with probability T the marked transition occurs while symbol x is output.

When $\gamma_1 \neq \gamma_2$ and $p \neq 0, 1$ this conditional distribution is different for every $\bar{n}$. We can thus treat $\bar{n}$ as being synonymous with the causal states; the causal states are in effect counting the number of 0s since the last event. The ϵ-machine of the process is shown in Fig. 1.

From previous studies on the computational mechanics of renewal processes [40,41], we can immediately identify that C_μ and D_μ are infinite in the continuum limit ($\Delta t \to 0$), as storing $\bar{n}$ involves tracking an infinity of states with non-negligible occupation probabilities; we can understand this as arising from the increasing lengths of strings of 0s as time steps are refined. Moreover, since $\Phi(\bar{n})$ remains nonzero for all $\bar{n}$, D_μ is also infinite at any level of discretization. Note that a continuous-variable classical memory must analogously support a distinguishable mode for each $\bar{n}$ and so will exhibit similar divergences. However, the differences between conditional probabilities become increasingly small for states at large $\bar{n}\Delta t$, and the probability of reaching such states, is very small. We hence introduce a truncated form of the model, where after sufficiently large $\bar{n}\Delta t$ the causal states are all merged together (see Supplemental Material [42]), and study the associated complexities $\tilde{C}_\mu$ and $\tilde{D}_\mu$ of this model. Their scaling with increasing precision (i.e., decreasing Δt) for $\gamma_1 = 12, \gamma_2 = 1$, and $p = 0.9$ is shown in Fig. 2. Note that the qualitative features of this plot are typical for any nonextremal choice of parameters (i.e., $P \neq 0, 1$ and $\gamma_1 \neq \gamma_2$).

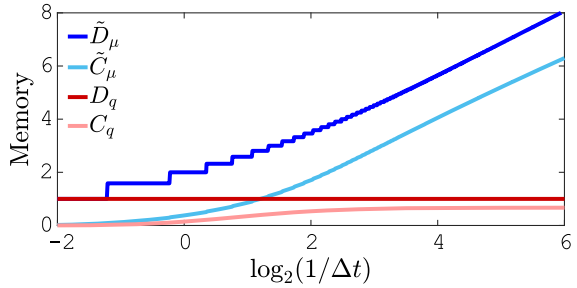

FIG. 2. Scaling of memory metrics with precision. Both classical memory metrics diverge with increasing precision, wherein the interval Δt is refined. In contrast, the quantum metrics remain finite, evincing an unbounded advantage: C_q tends to a bounded value, while D_q remains constant. Plot shown for $\gamma_1 = 12$, $\gamma_2 = 1$, and $p = 0.9$ (the qualitative features are typical for any nonextremal parameter choice).

Unbounded quantum compression advantage.—We now show that this scaling divergence is a purely classical phenomenon and need not persist in the quantum regime. By constructing quantum causal models of such processes for which the memory metrics are finite at any level of precision, we show unbounded quantum advantages in compression, forming our main result.

Main result. A quantum causal model with $C_q \leq 1$ and $D_q \leq 1$ exists for any dual Poisson process at any level of precision Δt.

Our models work by encoding the memory into one qubit and using another to probe it (Fig. 3). At each time step, a constant unitary interaction U acts on both the memory and probe qubits, after which measurement (in the computational basis $\{|0\rangle, |1\rangle\}$) of the probe qubit generates the corresponding output for the time step [3,14,31]. (The same U is applied at every time step, which depends only on the parameters defining the particular dual Poisson process and the desired precision, and is not conditioned on any external counter.) We define a set of quantum memory states $\{|\varsigma(n)\rangle\}$ corresponding to having observed n 0s since the last 1. We require

$$U|\varsigma(n)\rangle|0\rangle = \sqrt{\frac{\Phi(n+1)}{\Phi(n)}}|\varsigma(n+1)\rangle|0\rangle + \sqrt{1-\frac{\Phi(n+1)}{\Phi(n)}}|\varsigma(0)\rangle|1\rangle, \tag{3}$$

where the first subspace corresponds to the memory and the second the probe (reset to $|0\rangle$ at each time step). To understand this criterion, consider the required action of U—for any quantum memory state $|\varsigma(n)\rangle$, it must take the current memory state and blank probe state (left-hand side) to a state such that (i) the measurement statistics of the probe in the computational basis are correct according to Eq. (2) (setting $\bar{n} = n$, with $\bar{n} = 0$ and the cumulative

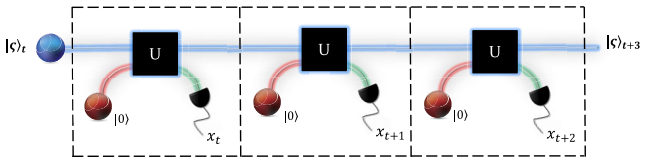

FIG. 3. Two-qubit quantum model. Our quantum models need only two qubits—one for the memory, and one to probe it. At each time step, a blank probe (red) is interacted with the memory qubit (blue) according to U and subsequently measured (green). The measurement outcome forms the output of the process, and the memory automatically updates conditional on this outcome (the conditional dependence is not explicitly depicted here). The dashed boxes delineate the repeated fundamental building block of the model, representing each time step.

$\vec{n} > 0$ to obtain the probability for 1 and 0, respectively); and (ii) the quantum memory state is updated correctly according to this outcome ($|\varsigma(n+1)\rangle$ for nonevents 0, and reset to $|\varsigma(0)\rangle$ for events 1). It can be seen that this is satisfied by the right-hand side of the condition, with the weightings corresponding to the probability (amplitudes) of the desired measurement statistics. Note that, in principle, we have the freedom to add a phase factor to the second term on the right-hand side; we do not include this here as it is not necessary for our construction.

In the Supplemental Material [42], we show that for any dual Poisson process the condition Eq. (3) is satisfied by the set of quantum memory states given by

$$|\varsigma(n)\rangle = \frac{\sqrt{p\Gamma_1^n} + ig\sqrt{\bar{p}\Gamma_2^n}}{\sqrt{\Phi(n)}}|0\rangle + \frac{i\sqrt{(1-g^2)\bar{p}\Gamma_2^n}}{\sqrt{\Phi(n)}}|1\rangle, \qquad (4)$$

where g is defined in the Supplemental Material, along with an explicit expression for U. Crucially, Eq. (4) evinces that the memory states can be encoded into a single qubit, guaranteeing $C_q \leq 1$ and $D_q \leq 1$. Moreover, since the process is generically not memoryless, and a binary system is the smallest possible memory, we can conclude that our model is (single-shot) minimal and that $D_q = 1$ is the quantum topological complexity. In Fig. 2 we compare the scaling of the quantum memory metrics with those of the minimal classical model. We thus observe the unbounded scaling of the quantum compression advantage in both ensemble and single-shot settings.

Relationship to other works.—We have shown that quantum dimensional advantages in causal modeling of classical stochastic processes can grow without bound. The highly cross-disciplinary nature of this work necessarily invites comparison with a range of prior and current research directions, and remarks on these relationships are in order. Foremost, a number of previous studies have shown unbounded quantum memory advantages in ensemble settings [12–15], where the advantage is contingent on an asymptotically large set of simulators acting in parallel with a shared memory. A scaling advantage in terms of dimension has previously been found for a Markovian process [1], albeit at the cost of an unboundedly large alphabet (and hence output register). Thus, while theoretically demonstrating the scaling of quantum memory advantages, these advantages are not presently experimentally feasible due to the need to either implement many simulators at once or assign an ancilla of unbounded dimension (e.g., a continuous-variable mode) for the output register. In contrast, our proposal requires only two qubits to demonstrate its advantage (and the associated scaling) and so is eminently more practical to implement; moreover, our proposal is the smallest possible that could ever demonstrate such an advantage, in the sense that if either the memory or output register of a model has fewer than two states then the process it simulates is trivial and/or memoryless.

The modeling of quantum dynamics with classical simulators is well studied; several works approaching this problem from a variety of angles show that it typically requires unbounded classical resources to track the dynamics of a finite quantum system, due to the continuous nature of the Hilbert space it occupies [5–8]. Here, by reversing the scenario, we show that this problem can turn into an asset—the very properties of (even simple) quantum systems that make them appear complex to classical systems can turn complex classical problems into simple quantum ones.

Despite a degeneracy in nomenclature, our framework is distinct from quantum causal models [45] in the sense of causal inference [46]. In such works, the goal is to identify causal relationships between variables, e.g., to determine if one variable *causes* the value of another or if both stem from a common cause; on the other hand, we start from the proposition that the past causes the future and seek to identify *what* the information in the past observations is that gives rise to (i.e., causes) the future statistics. Finally, we also note a resemblance between our discretized models using an ancillary system to interrogate a memory qubit and recent work on models of quantum clocks [47].

Concluding remarks.—The single-shot setting of our advantage is ideal for current and near-future quantum technologies. Crucially, such dimensional advantages can be more readily verified than corresponding entropic advantages; one need only count the dimension of the memory system, rather than perform full tomography [4]. The small-scale quantum systems required for our proposal are highly amenable to present experimental capabilities; they could, for instance, be implemented with current 2-qubit ion trap experiments, where sequential interaction-measurement-feedback cycles have been realized [48]. Moreover, photonic setups have already been used to experimentally realize the compression of a process with three causal states into a two-dimensional quantum memory [4], as well as quantum stochastic simulation over multiple time steps [49]; together, these form the two main aspects required for a proof-of-principle demonstration of our proposal. Consideration of resources in the experimentally more straightforward single-shot regime has garnered notable interest in other contexts [50–60].

While we have shown an unbounded dimensional scaling advantage for the dual Poisson process specifically, there are many other examples that can be found. For example, the behavior of a broader range of renewal processes can be captured by generalizing Eq. (4) to have different (potentially complex) amplitudes and include additional states. With the target of compressed simulation of given stochastic processes in mind, our findings motivate future work on developing the mapping of processes with large numbers of causal states into exact and near-exact quantum models with

low-dimensional memory—for renewal processes and beyond. A further direction would be to extend these techniques for scaling advantages to other continuous parameters such as spatial coordinates [12] or more abstract settings such as continuous belief spaces [20].

Interestingly, while both our quantum model and the optimal classical model provide an approximation of the fully continuous process for finite time steps Δt, in the quantum case a decrease in time step size is not accompanied by an increase in memory size. The quantum model memory size D_q is entirely independent of the time step and does not exhibit the classical scaling of memory with precision. This indicates that the limiting factor in the accuracy of quantum models of such processes is not the available memory, but the accuracy with which it can be addressed. Our results already in some sense indicate a robustness of the quantum advantage: errors in the implementation of the quantum model can be accounted for by limiting the precision Δt to not exceed that achievable by the experiment—and the problem of noise exceeding the difference in future statistics for large $\bar{n}$ is mitigated by the truncated process. Ultimately, while it would not be possible to witness the scaling difference all the way up to the continuum limit, it can still be shown up to the best achievable precision. We note that, while errors present in current quantum technologies would not prevent us from demonstrating that our quantum models can achieve better precision than any classical model at a fixed number of (qu) bits, the possibility to address larger numbers of classical bits with smaller errors than qubits on quantum computers would presently allow classical computers to achieve a higher level of precision. Nevertheless, our results suggest compression tasks as a potential future route for demonstrating absolute superiority of quantum technologies over classical devices and as a critical application of these incipient devices.

This work was funded by the Imperial College Borland Fellowship in Mathematics, the Lee Kuan Yew Endowment Fund (Postdoctoral Fellowship), Grants No. FQXi-RFP-1809, No. FQXi-RFP-1815, and No. FQXi-RFP-IPW-1903 from the Foundational Questions Institute and Fetzer Franklin Fund (a donor advised fund of Silicon Valley Community Foundation), Singapore Ministry of Education Tier 1 Grant No. RG190/17, National Research Foundation Fellowship NRF-NRFF2016-02, National Research Foundation and Agence Nationale de la Recherche Joint Project No. NRF2017-NRFANR004 VanQuTe and the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 801110 and the Austrian Federal Ministry of Education, Science and Research (BMBWF). T. J. E., C. Y., and F. C. B. thank the Centre for Quantum Technologies for their hospitality.

T. J. E. and C. Y. contributed equally to this work.

*physics@tjelliott.net
†yangchengran92@gmail.com
‡mgu@quantumcomplexity.org

[1] J. Thompson, A. J. P. Garner, J. R. Mahoney, J. P. Crutchfield, V. Vedral, and M. Gu, Phys. Rev. X **8**, 031013 (2018).
[2] S. P. Loomis and J. P. Crutchfield, J. Stat. Phys. **176**, 1317 (2019).
[3] Q. Liu, T. J. Elliott, F. C. Binder, C. D. Franco, and M. Gu, Phys. Rev. A **99**, 062110 (2019).
[4] F. Ghafari, N. Tischler, J. Thompson, M. Gu, L. K. Shalm, V. B. Verma, S. W. Nam, R. B. Patel, H. M. Wiseman, and G. J. Pryde, Phys. Rev. X **9**, 041013 (2019).
[5] A. Monras and A. Winter, J. Math. Phys. (N.Y.) **57**, 015219 (2016).
[6] A. Cabello, M. Gu, O. Gühne, J.-Å. Larsson, and K. Wiesner, Phys. Rev. A **94**, 052127 (2016).
[7] A. Cabello, M. Gu, O. Gühne, and Z.-P. Xu, Phys. Rev. Lett. **120**, 130401 (2018).
[8] P. Warszawski and H. M. Wiseman, Quantum **3**, 192 (2019).
[9] M. Gu, K. Wiesner, E. Rieper, and V. Vedral, Nat. Commun. **3**, 762 (2012).
[10] J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Sci. Rep. **6**, 20495 (2016).
[11] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde, Sci. Adv. **3**, e1601302 (2017).
[12] A. J. P. Garner, Q. Liu, J. Thompson, V. Vedral *et al.*, New J. Phys. **19**, 103009 (2017).
[13] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield, Sci. Rep. **7**, 6735 (2017).
[14] T. J. Elliott and M. Gu, npj Quantum Inf. **4**, 18 (2018).
[15] T. J. Elliott, A. J. P. Garner, and M. Gu, New J. Phys. **21**, 013021 (2019).
[16] A. Khintchine, Math. Ann. **109**, 604 (1934).
[17] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).
[18] C. R. Shalizi and J. P. Crutchfield, J. Stat. Phys. **104**, 817 (2001).
[19] J. P. Crutchfield, Nat. Phys. **8**, 17 (2012).
[20] A. R. Cassandra, L. P. Kaelbling, and M. L. Littman, in *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence* (AAAI Press, Seattle, Washington, 1994), pp. 1023–1028.
[21] L. P. Kaelbling, M. L. Littman, and A. W. Moore, J. Artif. Intell. Res. **4**, 237 (1996).
[22] R. Haslinger, K. L. Klinkner, and C. R. Shalizi, Neural Comput. **22**, 121 (2010).
[23] S. E. Marzen, M. R. DeWeese, and J. P. Crutchfield, Front. Comput. Neurosci. **9**, 105 (2015).
[24] C.-B. Li, H. Yang, and T. Komatsuzaki, Proc. Natl. Acad. Sci. U.S.A. **105**, 536 (2008).
[25] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner, PLoS One **7**, e29703 (2012).
[26] J. B. Park, J. W. Lee, J.-S. Yang, H.-H. Jo, and H.-T. Moon, Physica (Amsterdam) **379A**, 179 (2007).
[27] R. W. Clarke, M. P. Freeman, and N. W. Watkins, Phys. Rev. E **67**, 016203 (2003).
[28] A. J. Palmer, C. W. Fairall, and W. A. Brewer, IEEE Trans. Geosci. Remote Sensing **38**, 2056 (2000).
[29] D. P. Varn, G. S. Canright, and J. P. Crutchfield, Phys. Rev. B **66**, 174110 (2002).

[30] F. Ghafari Jouneghani, M. Gu, J. Ho, J. Thompson, W. Y. Suen, H. M. Wiseman, and G. J. Pryde, arXiv:1711.03661.
[31] F. C. Binder, J. Thompson, and M. Gu, Phys. Rev. Lett. **120**, 240502 (2018).
[32] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield, Phys. Lett. A **381**, 1223 (2016).
[33] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Phys. Rev. A **93**, 052317 (2016).
[34] W. Y. Suen, J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, Quantum **1**, 25 (2017).
[35] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, npj Quantum Inf. **3**, 6 (2017).
[36] C. Aghamohammadi, S. P. Loomis, J. R. Mahoney, and J. P. Crutchfield, Phys. Rev. X **8**, 011025 (2018).
[37] C. Yang, F. C. Binder, V. Narasimhachar, and M. Gu, Phys. Rev. Lett. **121**, 260602 (2018).
[38] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[39] W. L. Smith, J. R. Stat. Soc. Ser. B **20**, 243 (1958).
[40] S. E. Marzen and J. P. Crutchfield, Entropy **17**, 4891 (2015).
[41] S. Marzen and J. P. Crutchfield, J. Stat. Phys. **168**, 109 (2017).
[42] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.125.260501 for further technical details, which includes additional Refs. [43,44].
[43] F. Vatan and C. Williams, Phys. Rev. A **69**, 032315 (2004).
[44] B. Kraus and J. I. Cirac, Phys. Rev. A **63**, 062309 (2001).
[45] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Phys. Rev. X **7**, 031021 (2017).
[46] J. Pearl, *Causality: Models, Reasoning and Inference* (Springer, New York, 2000), Vol. 29.
[47] M. P. Woods, R. Silva, G. Pütz, S. Stupar, and R. Renner, arXiv:1806.00491.
[48] V. Negnevitsky, M. Marinelli, K. Mehta, H.-Y. Lo, C. Flühmann, and J. P. Home, Nature (London) **563**, 527 (2018).
[49] F. Ghafari, N. Tischler, C. Di Franco, J. Thompson, M. Gu, and G. J. Pryde, Nat. Commun. **10**, 1630 (2019).
[50] R. Renner, Int. J. Quantum. Inform. **06**, 1 (2008).
[51] L. Del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, Nature (London) **474**, 61 (2011).
[52] F. G. S. L. Brandao and N. Datta, IEEE Trans. Inf. Theory **57**, 1754 (2011).
[53] L. Wang and R. Renner, Phys. Rev. Lett. **108**, 200501 (2012).
[54] M. Tomamichel and M. Hayashi, IEEE Trans. Inf. Theory **59**, 7693 (2013).
[55] J. Åberg, Nat. Commun. **4**, 1925 (2013).
[56] M. Horodecki and J. Oppenheim, Nat. Commun. **4**, 2059 (2013).
[57] N. Y. Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, New J. Phys. **17**, 095003 (2015).
[58] M. Tomamichel, *Quantum Information Processing with Finite Resources: Mathematical Foundations* (Springer, New York, 2015), Vol. 5.
[59] K. Korzekwa, M. Lostaglio, J. Oppenheim, and D. Jennings, New J. Phys. **18**, 023045 (2016).
[60] B. Regula, K. Fang, X. Wang, and G. Adesso, Phys. Rev. Lett. **121**, 010401 (2018).