Optimal Classical Simulation of State-Independent Quantum Contextuality

Adán Cabello,[1,*] Mile Gu,[2,3,4] Otfried Gühne,[5] and Zhen-Peng Xu[6,1]

[1]*Departamento de Física Aplicada II, Universidad de Sevilla, E-41012 Sevilla, Spain*

[2]*School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore*

[3]*Complexity Institute, Nanyang Technological University, 18 Nanyang Drive, Singapore 637723, Singapore*

[4]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore*

[5]*Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Walter-Flex-Straße 3, D-57068 Siegen, Germany*

[6]*Theoretical Physics Division, Chern Institute of Mathematics, Nankai University, Tianjin 300071, People's Republic of China*

(Received 2 October 2017; revised manuscript received 6 February 2018; published 27 March 2018)

Simulating quantum contextuality with classical systems requires memory. A fundamental yet open question is what is the minimum memory needed and, therefore, the precise sense in which quantum systems outperform classical ones. Here, we make rigorous the notion of classically simulating quantum state-independent contextuality (QSIC) in the case of a single quantum system submitted to an infinite sequence of measurements randomly chosen from a finite QSIC set. We obtain the minimum memory needed to simulate arbitrary QSIC sets via classical systems under the assumption that the simulation should not contain any oracular information. In particular, we show that, while classically simulating two qubits tested with the Peres-Mermin set requires $\log_2 24 \approx 4.585$ bits, simulating a single qutrit tested with the Yu-Oh set requires, at least, 5.740 bits.

DOI: 10.1103/PhysRevLett.120.130401

Introduction.—Quantifying the resources needed to simulate quantum phenomena with classical systems is crucial to making precise the sense in which quantum systems provide an advantage over classical ones. While the extra resources needed for simulating entanglement and quantum nonlocality (i.e., the quantum violation of Bell inequalities [1]) have been studied extensively [2–8], the resources needed to simulate quantum contextuality [9,10], a natural generalization of quantum nonlocality to the case of non-spacelike separated systems and witnessed by the quantum violation of noncontextuality inequalities [11–15], have been less explored [16–18]. In a nutshell, while simulating quantum nonlocality with classical systems requires superluminal communication [2,5–8], simulating quantum contextuality requires memory [16–18] or, more precisely, the ability of storing and recovering a certain amount of classical information. It is known that, in some cases, the required memory is larger than the information-carrying capacity of the corresponding quantum system [16]. The problem is that only lower bounds to the minimum memory are known for some particular scenarios [16,18]. In addition, it is not known how the minimum memory scales with, e.g., the size of the set of possible measurements.

A particularly interesting case is that of quantum state-independent contextuality (QSIC) in experiments with sequential measurements [12–15] on a single recycled quantum system [16,19,20]. In this case, a single quantum system is submitted to an unlimited sequence of measurements, randomly chosen from a finite set of measurements, as illustrated in Fig. 1. After each measurement, the outcome is observed and recorded. The set of measurements has the peculiarity of being able to produce contextuality no matter what is the initial quantum state of the system. These sets are called QSIC sets [21,22] and, for each of them, there are optimal combinations of correlations for detecting contextuality [15]. The interest of this case comes from the fact that unbounded strings of data with contextual correlations can be produced using a *single* system initially prepared in an arbitrary state [20], a situation that strongly contrasts with the case of nonlocality generated through the violation of a Bell inequality, where thousands of spacelike separated pairs of quantum systems in an entangled quantum state are needed. The question we want to address in this Letter is what is the minimal amount of memory a classical system would need to simulate the predictions of quantum theory

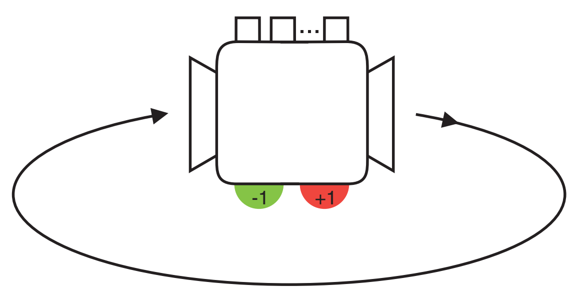

FIG. 1. Contextuality experiment on a recycled system. Buttons represent possible measurements. Light bulbs represent possible outcomes. We consider experiments in which there are as many buttons as elements of the QSIC set, and all of them have two outcomes.

for QSIC experiments with unlimited sequential measurements. Contrary to the previous approaches [16,18], we aim at simulating all statistics arising in quantum theory and not only the perfect correlations leading to a violation of a contextuality inequality. We consider the most general simulation under the restriction that the classical model used for simulation should not contain oracular information, as explained below.

Scenario.—We consider ideal experiments in which successive measurements are performed on a single quantum system at times $t_1 < t_2 < \cdots$. At each t_i, a measurement belonging to a QSIC set is randomly chosen and performed. We assume that the quantum state after the measurement at t_i is the quantum state before the measurement at t_{i+1}. The process is repeated infinitely many times. Our aim is to extract conclusions valid for any QSIC set. However, for the sake of clarity, we will present our results using two famous QSIC sets.

The Peres-Mermin set.—The QSIC set with the smallest number of observables known has nine 2-qubit observables and it is shown in Fig. 2. It was introduced by Peres [23] and Mermin [24] and first implemented in experiments with sequential measurements by Kirchmair *et al.* [25] on trapped ions and by Amselem *et al.* [26] on single photons. In addition, it has been recently implemented on entangled photons by Liu *et al.* [27].

When one uses this set for unlimited sequential measurements on a single system, from the moment two different observables that are in the same row or column in Fig. 2 are measured consecutively, the system remains in one of the 24 quantum states defined in Fig. 2. After that, any other subsequent measurement leaves the system in one of these 24 quantum states and each of them occurs with the same probability.

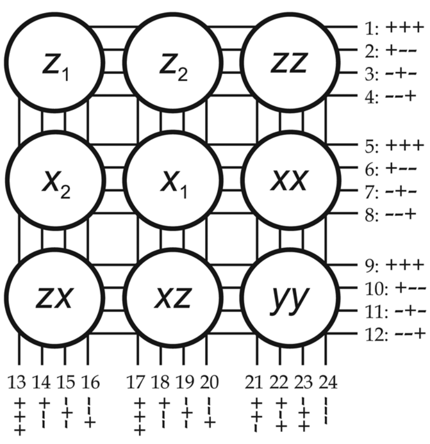

FIG. 2. Observables in the Peres-Mermin set. z_1 denotes the quantum observable represented by the operator $\sigma_z^{(1)} \otimes \mathbb{1}^{(2)}$. Similarly, zx denotes $\sigma_z^{(1)} \otimes \sigma_x^{(2)}$. Observables in each row or column are mutually compatible and their corresponding operators have four common eigenstates. In the figure, these eigenstates are represented by straight lines numbered from 1 to 24. For example, quantum state $|2\rangle$ is the one satisfying $\sigma_z^{(1)} \otimes \mathbb{1}^{(2)}|2\rangle = |2\rangle$, $\mathbb{1}^{(1)} \otimes \sigma_z^{(2)}|2\rangle = -|2\rangle$, and $\sigma_z^{(1)} \otimes \sigma_z^{(2)}|2\rangle = -|2\rangle$.

The Yu-Oh set.—As proven in Ref. [22], the QSIC set with the smallest number of observables represented by rank-one projectors has 13 single-qutrit observables. It was introduced by Yu and Oh [14] and is a subset of a QSIC set previously considered by Peres [28]. Its associated optimal noncontextuality inequality was found by Kleinmann *et al.* [15]. It inspired a photonic experiment by Zu *et al.* [29] (see also Amselem *et al.* [30]) and was implemented as an experiment with sequential measurements on a single ion by Zhang *et al.* [31], and, recently, it was used to implement the scheme in Fig. 1 by Leupold *et al.* [20].

When one uses the Yu-Oh set for unlimited sequential measurements on a single system, if at any point the system is in one of the 13 pure states of the Yu-Oh set and one measures one randomly chosen projector onto one of these 13 states, then the number of possible postmeasurement states does not remain constant but grows with the number of sequential measurements. In fact, some states are more probable than others (see Fig. 3). This contrasts with the case of the Peres-Mermin set, where the number of possible postmeasurement states is constant and all states are equally probable.

The notion of simulation and relation to previous works.—When talking about a classical simulation of a temporal process, it is important to specify what precisely shall be simulated and which conditions a simulation apparatus should meet. A general strategy for simulating temporal correlations is to use hidden Markov models (HMMs) [34] or, when deterministic effects are considered, Mealy machines [35]. There, the simulation apparatus is always in a definite internal state k, and for each internal state k, there is an output mechanism (e.g., a table R_k containing all the results of the potential measurements) and an update mechanism (e.g., a table U_k that describes the change of the internal state depending on the measurement). In such a model, however, it can easily happen that the simulation apparatus contains information about the future that cannot be derived from the past. By this we mean the following: consider two persons, where the first one only knows the current internal state of the machine and the second one only knows the past observation of measurements and results. Clearly, if the simulation apparatus simulates all the correlations properly, the first person can predict the future as well as the second person. For many processes, however, it can happen that the first person can predict the future better, e.g., if the given internal state k predicts a deterministic outcome for the next measurement, which cannot be deduced from the past. In this way, a simulation apparatus can contain oracular information (i.e., information that cannot be obtained from the past) [36].

For our simulation, we restrict our attention to a simulation without oracular information. This leads to

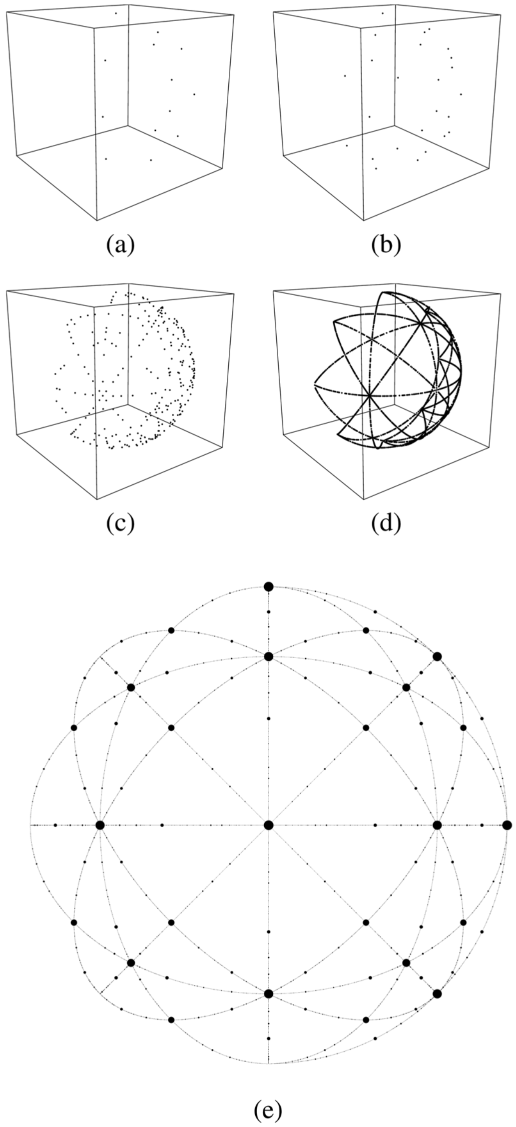

FIG. 3. Assuming that the experiment starts with a qutrit in one of the 13 quantum states of the Yu-Oh set, represented by the dots over a semisphere in (a), the successive figures show the possible postmeasurement quantum states after one (b), three (c), five (d), and seven measurements (e). The number of possible postmeasurement states is 25, 265, 3649, and 50 293, respectively. All the states lie in one of the 13 semicircles corresponding to the states with real components orthogonal to the 13 states of the Yu-Oh set. However, not all of them occur with the same probability. To illustrate this, the volume of each point in (e) is proportional to the probability with which the corresponding state appears. Collectively, these figures exemplify the typical behavior of QSIC sets (e.g., [32,33]).

the notion of causal models and, more specifically, to ε transducers, as explained below. These are also so-called unifilar processes, meaning that they are special HMMs, where the output derived from the internal state k determines the update of the internal state. We note that with more general HMMs the memory required for the simulation can sometimes be reduced [36,37] and that such models have been used to simulate the Peres-Mermin set [16,18]. Our restriction to causal models, however, is physically motivated by the demand that only the past observations should be used for simulating the future.

Tools.—To calculate the minimum memory that a classical system must have, a key observation is that our ideal experiments are examples of stochastic input-output processes that can be analyzed in information-theoretic terms. A stochastic process $\overleftrightarrow{\mathcal{Y}}$ is a one-dimensional chain $\ldots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \ldots$ of discrete random variables $\{Y_t\}_{t\in\mathbb{Z}}$ that take values $\{y_t\}_{t\in\mathbb{Z}}$ over a finite or countably infinite alphabet $\mathcal{Y}$. An input-output process $\overleftrightarrow{Y}|\overleftrightarrow{X}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is a collection of stochastic processes $\overleftrightarrow{Y}|\overleftrightarrow{X} \equiv \{\overleftrightarrow{Y}|\overleftrightarrow{x}\}_{\overleftrightarrow{x}\in\overleftrightarrow{\mathcal{X}}}$, where each such process $\overleftrightarrow{Y}|\overleftrightarrow{x}$ corresponds to all possible output sequences $\overleftrightarrow{Y}$ given a particular bi-infinite input sequence $\overleftrightarrow{x}$. It can be represented as a finite-state automaton or, equivalently, as a hidden Markov process. In our experiment, x_t is the observable measured at time t and y_t is the corresponding outcome. By $\overleftarrow{X}$ we denote the chain of previous measurements, $\ldots, X_{t-2}, X_{t-1}$, by $\vec{X}$ we denote $X_t, X_{t+1}, \ldots$, and by $\overleftrightarrow{X}$ we denote the chain $\ldots, X_{t-1}, X_t, X_{t+1}, \ldots$. Similarly, $\overleftarrow{Y}$, $\vec{Y}$, and $\overleftrightarrow{Y}$ denote the past, future, and all outcomes, respectively, while $\overleftarrow{Z}$, $\vec{Z}$, and $\overleftrightarrow{Z}$ denote the past, future, and all pairs of measurements and outcomes. For deriving physical consequences, we have to consider the minimal and optimal representation of this process.

As proven in Ref. [38], the fact that each of our experiments is an input-output process implies that for each of them there exists a unique minimal and optimal predictor of the process, i.e., a unique finite-state machine with minimal entropy over the state probability distribution and maximal mutual information with the process's future output given the process's input-output past and the process's future input. This machine is called the process's ε transducer [38] and is the extension of the so-called ε machines [39,40]. An ε transducer of an input-output process is a tuple $(\mathcal{X}, \mathcal{Y}, \mathcal{S}, \mathcal{T})$ consisting of the process's input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, the set of causal states $\mathcal{S}$, and the set of corresponding conditional transition probabilities $\mathcal{T}$. The causal states $s_{t-1} \in \mathcal{S}$ are the equivalence classes in which the set of input-output pasts $\overleftarrow{\mathcal{Z}}$ can be partitioned in such a way that two input-output pasts $\overleftarrow{z}$ and $\overleftarrow{z}'$ are equivalent if and only if the probabilities $P(\vec{Y}|\vec{X}, \overleftarrow{Z} = \overleftarrow{z})$ and $P(\vec{Y}|\vec{X}, \overleftarrow{Z} = \overleftarrow{z}')$ are equal. The causal states are a so-called sufficient statistic of the process. They store all the information about the past needed to predict the output and as little as possible of the remaining information overhead contained in the past. The Shannon entropy over the stationary distribution of the causal states $H(\mathcal{S})$ is the so-called statistical complexity and represents the minimum internal entropy needed to be stored to optimally compute future measurement outcomes (this quantity generally depends on how our measurements $\overleftrightarrow{X}$ are selected; here, we assume each X_t is selected from a uniform probability distribution). The set of conditional transition probabilities $\mathcal{T} \equiv \{P(S_{t+1} = s_j, Y_t = y | S_t = s_i, X_t = x)\}$ governs the evolution.

Minimum memory needed to simulate QSIC.—The ε transducers associated with the QSIC experiments have a particular property, namely, that there is a one-to-one correspondence between causal states s_t and quantum states $|\psi_t\rangle \in \Phi$, where Φ is the set of possible states occurring after a measurement (for completeness, a proof is presented in the Supplemental Material [41]). Therefore, the minimum number of bits a finite-state classical machine must have to simulate the predictions of quantum theory for a QSIC experiment with unlimited sequential measurements chosen uniformly at random is given by the Shannon entropy

$$H = -\sum_i p_i \log_2 p_i. \tag{1}$$

In (1), p_i is the probability of each quantum state achievable during the experiment's occurrence and, in general, depends on the distribution in which different measurements are chosen.

For the Peres-Mermin set, there are 24 causal states, each occurring with equal probability (see Fig. 2). Hence, a simulation with an ε transducer requires $\log_2(24) = 4.585$ bits to imitate a quantum system of 2 qubits. This classical memory is significantly higher than the classical information-carrying capacity of the quantum system that produces these correlations.

For the Yu-Oh set, the calculations are more involved. The reason is that the longer the measurement sequence is, the more possible quantum states can occur as postmeasurement states. In addition, the quantum states do not occur with the same probability; see Fig. 3. For small sequences up to length ten, however, all the states and probabilities can be analytically computed. The results imply that if only the last ten measurements and results are included, at least 5.740 bits are required for the simulation (see Fig. 4).

A proper comparison with the amount of memory required to simulate noncontextual sets is obtained by noticing that the memory required to reproduce the predictions of quantum mechanics when we restrict the measurements to subsets (of the QSIC sets) that cannot produce contextuality, is 2 bits for the Peres-Mermin set and $\log_2 3 \approx 1.585$ bits for the Yu-Oh set. These values are obtained as follows. Contextuality is an impossibility of a joint probability distribution over a single probability space. For sequences of projective measurements, incompatibility implies the nonexistence of a joint probability

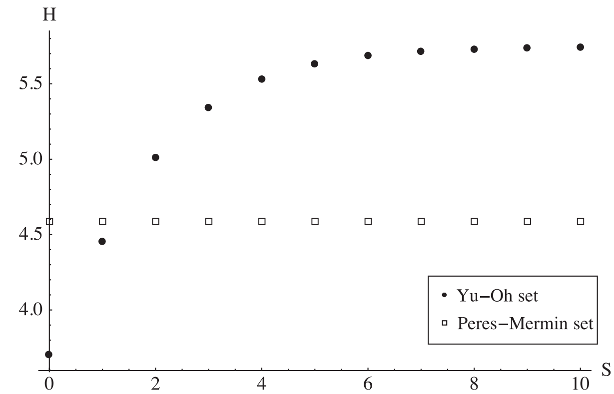

FIG. 4. Classical memory in bits needed to simulate sequential Yu-Oh and Peres-Mermin experiments, as given by Eq. (1), as a function of the number of steps, as defined in the caption of Fig. 3 for the case of Yu-Oh (and, similarly, for the case of Peres-Mermin). Values are obtained from considering all possible measurement sequences of a given number of steps and then analytically calculating the corresponding results and postmeasurement states.

distribution. Therefore, the memory needed to simulate noncontextual sets is the one required to reproduce the predictions of quantum mechanics for subsets of mutually compatible measurements of the QSIC set, which is $\log_2 d$ bits for any QSIC set of dimension d. Notice that contextuality requires incompatibility, but also that measurements can be grouped into mutually compatible subsets so that each measurement belongs to at least two of them. Therefore, simulating a set of incompatible measurements not restricted by these rules may require more memory.

One might conjecture that the minimal memory necessary to classically simulate QSIC must be related to the degree of contextuality. However, the relation is difficult to trace. For example, while the minimal memory necessary to simulate classically the Peres-Mermin and Yu-Oh sets is larger for Yu-Oh, the degree of contextuality that can be measured by, e.g., the ratio between the violation and the noncontextual bound for the optimal noncontextuality inequalities [15] is 1.5 for Peres-Mermin and 1.107 for Yu-Oh, showing that contextuality is higher for Peres-Mermin. The same conclusion can be reached by adopting other measures of contextuality [42,43]. Therefore, understanding the connection between memory and the degree of contextuality is an interesting open problem that should be addressed in the future. Here, also the effects of noise and imperfections should be considered.

Conclusions.—The question of which classical resources are needed for simulating quantum effects is central for the connection of the foundations of quantum theory with quantum information. By applying the tools of complexity science, we have shown how to calculate the amount of memory a classical system would need to simulate quantum state-independent contextuality in the case of a single quantum system submitted to an infinite sequence of measurements randomly chosen from any finite set. Our result precisely quantifies the quantum vs classical advantage of a phenomenon, quantum state-independent contextuality, discovered 50 years ago and shows how profitable may be combining previously unrelated disciplines, such as complexity and quantum information.

Our result opens a way to test systems for their quantumness. Suppose we have a system whose internal functioning is unknown and that is submitted to sequential measurements for which a classical simulation requires more memory than the one allowed by the Bekenstein bound. Here, the Bekenstein bound refers to the limit on the entropy that can be contained in a physical system with given size and energy [44]. We may assume that no system can store and process information beyond the Bekenstein bound and can test whether the system is not emitting heat due to Landauer's principle (which states that the erasure of classical information implies some heat emission [45]). If this heat is not found, then our result allows us to certify that the system is in fact quantum and not a classical simulation. Therefore, we can use its quantum features for information processing.

On the other hand, our result could also inspire new techniques in complexity science, where there is a growing interest in the value of quantum theory for simulating otherwise difficult to simulate classical processes (e.g., [46,47]). In this respect, our result could pinpoint the properties of classical processes that make them particularly amenable to improved modeling using quantum systems and thus also further catalyze the use of quantum methods in complexity science.

We thank Matthias Kleinmann, Jan-Åke Larsson, and Karoline Wiesner for discussions and Jayne Thompson for help in the Supplemental Material [41]. This work was supported by Project No. FIS2014-60843-P, "Advanced Quantum Information" (MINECO, Spain) with FEDER funds, the project "Photonic Quantum Information" (Knut and Alice Wallenberg Foundation, Sweden), and the Singapore National Research Foundation Fellowship No. NRF-NRFF2016-02. A. C. is also supported by the FQXi Large Grant "The Observer Observed: A Bayesian Route to the Reconstruction of Quantum Theory." M. G. is also supported by the John Templeton Foundation Grant No. 53914 "Occam's Quantum Mechanical Razor: Can Quantum Theory Admit the Simplest Understanding of Reality?" and the FQXi Large Grant "Observer-Dependent Complexity: The Quantum-Classical Divergence over 'What is Complex?'" O. G. is also supported by the DFG and the ERC (Consolidator Grant No. 683107/TempoQ). Z.-P. X. is also supported by the Natural Science Foundation of China (Grant No. 11475089) and the China Scholarship Council.

*adan@us.es

[1] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics **1**, 195 (1964).

[2] G. Brassard, R. Cleve, and A. Tapp, Cost of Exactly Simulating Quantum Entanglement with Classical Communication, Phys. Rev. Lett. **83**, 1874 (1999).

[3] M. Steiner, Towards quantifying non-local information transfer: Finite-bit non-locality, Phys. Lett. A **270**, 239 (2000).

[4] N. J. Cerf, N. Gisin, and S. Massar, Classical Teleportation of a Quantum Bit, Phys. Rev. Lett. **84**, 2521 (2000).

[5] D. Bacon and B. F. Toner, Bell Inequalities with Auxiliary Communication, Phys. Rev. Lett. **90**, 157904 (2003).

[6] B. F. Toner and D. Bacon, Communication Cost of Simulating Bell Correlations, Phys. Rev. Lett. **91**, 187904 (2003).

[7] S. Pironio, Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations, Phys. Rev. A **68**, 062102 (2003).

[8] N. J. Cerf, N. Gisin, S. Massar, and S. Popescu, Simulating Maximal Quantum Entanglement without Communication, Phys. Rev. Lett. **94**, 220403 (2005).

[9] J. B. Bell, On the problem of hidden variables in quantum mechanics, Rev. Mod. Phys. **38**, 447 (1966).

[10] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, J. Math. Mech. **17**, 59 (1967).

[11] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Simple Test for Hidden Variables in Spin-1 Systems, Phys. Rev. Lett. **101**, 020403 (2008).

[12] A. Cabello, Experimentally Testable State-Independent Quantum Contextuality, Phys. Rev. Lett. **101**, 210401 (2008).

[13] P. Badziąg, I. Bengtsson, A. Cabello, and I. Pitowsky, Universality of State-Independent Violation of Correlation Inequalities for Noncontextual Theories, Phys. Rev. Lett. **103**, 050401 (2009).

[14] S. Yu and C. H. Oh, State-Independent Proof of Kochen-Specker Theorem with 13 Rays, Phys. Rev. Lett. **108**, 030402 (2012).

[15] M. Kleinmann, C. Budroni, J.-Å. Larsson, O. Gühne, and A. Cabello, Optimal Inequalities for State-Independent Contextuality, Phys. Rev. Lett. **109**, 250402 (2012).

[16] M. Kleinmann, O. Gühne, J. R. Portillo, J.-Å. Larsson, and A. Cabello, Memory cost of quantum contextuality, New J. Phys. **13**, 113011 (2011).

[17] A. Cabello, The role of bounded memory in the foundations of quantum mechanics, Found. Phys. **42**, 68 (2012).

[18] G. Fagundes and M. Kleinmann, Memory cost for simulating all quantum correlations of the Peres-Mermin scenario, J. Phys. A **50**, 325302 (2017).

[19] M. Wajs, S.-Y. Lee, P. Kurzyński, and D. Kaszlikowski, State-recycling method for testing quantum contextuality, Phys. Rev. A **93**, 052104 (2016).

[20] F. M. Leupold, M. Malinowski, C. Zhang, V. Negnevitsky, J. Alonso, A. Cabello, and J. P. Home, Sustained state-independent quantum contextual correlations from a single ion, arXiv:1706.07370.

[21] A. Cabello, M. Kleinmann, and C. Budroni, Necessary and Sufficient Condition for Quantum State-Independent Contextuality, Phys. Rev. Lett. **114**, 250402 (2015).

[22] A. Cabello, M. Kleinmann, and J. R. Portillo, Quantum state-independent contextuality requires 13 rays, J. Phys. A **49**, 38LT01 (2016).

[23] A. Peres, Incompatible results of quantum measurements, Phys. Lett. A **151**, 107 (1990).

[24] N. D. Mermin, Simple Unified Form for the Major No-Hidden-Variables Theorems, Phys. Rev. Lett. **65**, 3373 (1990).

[25] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, State-independent experimental test of quantum contextuality, Nature (London) **460**, 494 (2009).

[26] E. Amselem, M. Rådmark, M. Bourennane, and A. Cabello, State-Independent Quantum Contextuality with Single Photons, Phys. Rev. Lett. **103**, 160405 (2009).

[27] B.-H. Liu, X.-M. Hu, J.-S. Chen, Y.-F. Huang, Y.-J. Han, C.-F. Li, G.-C. Guo, and A. Cabello, Nonlocality from Local Contextuality, Phys. Rev. Lett. **117**, 220402 (2016).

[28] A. Peres, Two simple proofs of the Kochen-Specker theorem, J. Phys. A **24**, L175 (1991).

[29] C. Zu, Y.-X. Wang, D.-L. Deng, X.-Y. Chang, K. Liu, P.-Y. Hou, H.-X. Yang, and L.-M. Duan, State-Independent Experimental Test of Quantum Contextuality in an Indivisible System, Phys. Rev. Lett. **109**, 150401 (2012).

[30] E. Amselem, M. Bourennane, C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, J.-Å. Larsson, and M. Wieśniak, Comment on "State-Independent Experimental Test of Quantum Contextuality in an Indivisible System", Phys. Rev. Lett. **110**, 078901 (2013).

[31] X. Zhang, M. Um, J. Zhang, S. An, Y. Wang, D.-L. Deng, C. Shen, L.-M. Duan, and K. Kim, State-Independent Experimental Test of Quantum Contextuality with a Single Trapped Ion, Phys. Rev. Lett. **110**, 070401 (2013).

[32] A. Cabello, J. M. Estebaranz, and G. García-Alcaine, Bell-Kochen-Specker theorem: A proof with 18 vectors, Phys. Lett. A **212**, 183 (1996).

[33] P. Lisoněk, P. Badziąg, J. R. Portillo, and A. Cabello, Kochen-Specker set with seven contexts, Phys. Rev. A **89**, 042101 (2014).

[34] L. R. Rabiner and B. H. Juang, An introduction to hidden Markov models, IEEE ASSP Mag. **3**, 4 (1986).

[35] G. H. Mealy, A method for synthesizing sequential circuits, Bell Syst. Tech. J. **34**, 1045 (1955).

[36] J. P. Crutchfield, C. J. Ellison, R. G. James, and J. R. Mahoney, Synchronization and control in intrinsic and designed computation: An information-theoretic analysis of competing models of stochastic computation, Chaos **20**, 037105 (2010).

[37] W. Löhr and N. Ay, in *Complex Sciences. Complex 2009, Part I*, edited by J. Zhou, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Vol. 4 (Springer, Berlin, 2009), p. 265.

[38] N. Barnett and J. P. Crutchfield, Computational mechanics of input-output processes: Structured transformations and the ε-transducer, J. Stat. Phys. **161**, 404 (2015).

[39] J. P. Crutchfield and K. Young, Inferring Statistical Complexity, Phys. Rev. Lett. **63**, 105 (1989).

[40] C. R. Shalizi and J. P. Crutchfield, Computational mechanics: Pattern and prediction, structure and simplicity, J. Stat. Phys. **104**, 817 (2001).

[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.120.130401 for an explanation of the one-to-one correspondence between causal states and quantum states.

[42] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, New J. Phys. **13**, 113036 (2011).

[43] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, Quantifying Contextuality, Phys. Rev. Lett. **112**, 120401 (2014).

[44] J. D. Bekenstein, Black holes, and the second law, Lett. Nuovo Cimento **4**, 737 (1972).

[45] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[46] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, Using quantum theory to simplify input-output processes, npj Quantum Inf. **3**, 6 (2017).

[47] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde, Experimentally modeling stochastic processes with less memory by the use of a quantum processor, Sci. Adv. **3**, e1601302 (2017).