

Contextuality and Wigner-function negativity in qubit quantum computation

Robert Raussendorf,¹ Dan E. Browne,² Nicolas Delfosse,^{3,4,5} Cihan Okay,⁶ and Juan Bermejo-Vega^{7,8}

¹*Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*

²*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*

³*Département de Physique, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada*

⁴*IQIM, California Institute of Technology, Pasadena, California 91125, USA*

⁵*Department of Physics and Astronomy, University of California, Riverside, California 92521, USA*

⁶*Department of Mathematics, University of Western Ontario, London, Ontario N6A 5B7, Canada*

⁷*Max-Planck Institute for Quantum Optics, Theory Division, Garching 85748, Germany*

⁸*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin 14195, Germany*

(Received 16 November 2016; published 17 May 2017)

We describe schemes of quantum computation with magic states on qubits for which contextuality and negativity of the Wigner function are necessary resources possessed by the magic states. These schemes satisfy a constraint. Namely, the non-negativity of Wigner functions must be preserved under all available measurement operations. Furthermore, we identify stringent consistency conditions on such computational schemes, revealing the general structure by which negativity of Wigner functions, hardness of classical simulation of the computation, and contextuality are connected.

DOI: [10.1103/PhysRevA.95.052334](https://doi.org/10.1103/PhysRevA.95.052334)

I. INTRODUCTION

Contextuality [1–9] and negativity of Wigner functions [10–14] have recently been established as a necessary resources for quantum computation by injection of magic states (QCSI) [15–19]. This was first achieved for systems of qudits [15,17], where the local degrees of freedom have an odd number of states. These findings beg the question: Under which conditions are contextuality and Wigner function negativity computational resources for QCSI on n -qubit systems?

This is the question we address in the present paper. As a first step into this direction, it was shown in Ref. [16] that Wigner function negativity and contextuality of magic states are necessary resources for QCSI on rebits, which are systems of local dimension 2 whose state vector is real valued. Here we generalize that approach.

Before we state our results, we give a brief description of the technical terms that have appeared above. *Quantum computation with magic states* (QCSI) [20] is closely related to the standard circuit model. It deviates in that the allowed state preparations, unitary transformations, and measurements are restricted to be nonuniversal and, in fact, efficiently classically simulable. These operations are called “free.” Computational universality is restored by the capability to inject so-called magic states, which are states that cannot be prepared by the free operations. The source of computational power thus shifts from the gates to the magic states, and it is natural to ask which quantum features must be present in these states to allow for quantum speedup.

Wigner functions [10–13] describe quantum states in phase space. They are quasiprobability distributions, and as such the closest quantum analog to joint probability distributions of position and momentum in classical statistical mechanics. The difference is that Wigner functions can take negative values, and this negativity is a signature of quantumness [21,22].

Contextuality means that measurements cannot be viewed as deterministically revealing pre-existing properties of a system. This happens in Hilbert spaces of dimension $d \geq 3$ [2]. There, it is impossible to consistently assign pre-existing

values λ to all observables, in such a way that $\lambda(A)$ is a property of the observable A alone and not also of compatible observables that are measured simultaneously.

State-dependent contextuality is a weaker form that can be attributed to quantum states rather than observables [4], and it is the notion we are interested in here. The set of measurable observables may be restricted in such a way that consistent noncontextual value assignments exist. Then, a probabilistic mixture of such value assignments is a noncontextual hidden variable model (ncHVM). For certain quantum states, it may happen that no probability distribution over the value assignments correctly reproduces all measurement statistics. Quantum states of this kind are called contextual.

The phenomenon of state-independent contextuality with regard to Pauli observables (which are the free observables in QCSI) represents an obstacle towards establishing contextuality of the magic states as a computational resource. Namely, because of it, for $n \geq 2$ qubits, there does not even exist a single consistent noncontextual value assignment for the Pauli observables. Hence, no quantum state of $n \geq 2$ qubits, not even the completely mixed state, can be described by a ncHVM. Every quantum state is thus contextual. And if contextuality is ubiquitous, it cannot be a resource.

In this paper, we overcome the problem posed by state-independent contextuality by restricting to QCSI schemes on qubits which satisfy an additional condition. Under this condition, contextuality and Wigner function negativity of the magic states are necessary computational resources for QCSI on qubits. The condition is as follows:

(P1) Non-negative Wigner functions remain non-negative under free measurements.

Condition P1 is the very basis for the usefulness of Wigner functions in the description of QCSI, namely to reveal the near classicality of QCSI without the magic states. This property also holds for the previously discussed cases of qudits and rebits. P1 restricts the Pauli observables that are available for measurement. To motivate this constraint, we observe that many constructions for fault-tolerant quantum computation

are QCSI schemes in which the free gates are a strict subset of the Clifford gates. One example is the first concrete QCSI scheme ever proposed, with the free operations provided by braiding and fusion of Ising anyons, and magic states enabling CNOT and $\pi/8$ gates [23].

However, Condition P1 is trickier than might at first appear. For a start, we do not require a counterpart of P1 for the unitary operations available in QCSI, and imposing it would indeed be too restrictive. Those unitaries may introduce large amounts of negativity into the Wigner function without compromising efficient classical simulability.

Condition P1 can be used to define the free measurements in a QCSI scheme, given a Wigner function. Typically, but not always, the resulting set of free measurements is small. Thus the question arises whether the corresponding QCSI schemes really operate on the intended number of qubits. To filter out the true n -qubit QCSI schemes, we therefore impose a second condition:

(P2) The available measurements are tomographically complete.

Condition P2 means that with the free operations of an n -qubit QCSI scheme the density matrix ρ of any n -qubit quantum state can be fully measured. Rebit QCSI [16], for example, does not satisfy it.

Conditions P1 and P2 are hard to satisfy jointly. Informally speaking, for typical Wigner functions, P1 generates small sets of free measurements, whereas P2 demands large sets. Yet, sets of freely measurable observables that satisfy both P1 and P2 do exist for every number n of qubits.

We show that for all n -qubit QCSI schemes satisfying P1, contextuality and Wigner function negativity are necessary for quantum computational universality and quantum speedup. The result on contextuality is strictly stronger than our corresponding result on Wigner functions. Unlike the qudit case [15,24], magic states can have negative Wigner functions but still be noncontextual. Consequently, the fundamental notion of classicality for qubit QCSI schemes subject to P1 is a noncontextual hidden variable model, not a positive Wigner function.

Our result on contextuality is an extension of the corresponding results for qudits [15], where the Hilbert space dimension of the local systems is an odd prime or a power of an odd prime, and for rebits [16], where the Hilbert space dimension of the local systems is 2, but the density matrix is constrained to be real.

Our result on Wigner negativity as a resource extends [17] from qudits to qubits, and is complementary to the earlier work [19] on qubits. In Ref. [19], many Wigner functions are simultaneously used to characterize the classical region for the magic states, whereas our work only involves one. On the other hand, the present result only applies for QCSI schemes that satisfy the Condition P1, whereas Ref. [19] applies in general. For QCSI schemes where both methods can be applied, the present result labels a larger region of the state space as guaranteed classical.

II. MOTIVATION, RESULTS, AND OUTLINE

In this section, we give a broader motivation for our work, explain why the qubit case requires a separate discussion,

summarize our results, and provide an outline of the remainder of the paper.

A. Motivation

Which quantum property makes the magic states valuable for QCSI? This is the question we address in the present paper, and to approach it, we first need to clarify in which sense the free operations of QCSI by themselves are not fully quantum and not computationally valuable. These operations are certainly not entirely classical. For example, highly entangled states can be created by them.

In the present paper, the term *qudit* always refers to systems of odd Hilbert space dimension. In this benign case of qudits, the near classicality of the free operations is revealed by the following characteristics: The free operations (which are Clifford operations) preserve the properties of quantum states (i) to be stabilizer states and hence to be efficiently classically simulable by the stabilizer formalism [25], (ii) to have a non-negative Wigner function, and (iii) to be describable by a noncontextual hidden variable model.

Thus there are different angles from which to view classicality in QCSI—the quantum optics angle, focusing on Wigner functions; the quantum foundations angle, focusing on ncHVMS; and the stabilizer angle, focusing on efficient classical simulability. Remarkably, in the qudit case, those different angles amount to essentially the same. First, the notions of contextuality and Wigner function negativity agree for qudits [15,24] and for the Wigner function proposed in Refs. [11,12]. Next, if the magic states have a non-negative Wigner function, the resulting QCSIs can be efficiently classically simulated [17]. The domain of applicability of this simulation method strictly contains the domain of applicability of the stabilizer method.

When transitioning from qudits to qubits, this well-rounded picture of classicality splinters. The stabilizer simulation method survives unharmed. Wigner functions can still be defined, although their definition cannot be straightforwardly adapted from the infinite-dimensional or finite odd-dimensional case [13,26,27]. Their non-negativity is no longer preserved under all Clifford gates.

Worst fares the contextuality of the magic states. As already noted in the introduction, due to the phenomenon of state-independent contextuality with Pauli observables [3], ncHVMS can no longer be consistently defined. Hence, every quantum states on $n \geq 2$ becomes contextual [15], and contextuality of the magic states is no longer a resource.

The latter represents a severe obstacle to extending the result [15] to qubits. We overcome it by invoking the additional assumption, Condition P1. The phenomenon of state-independent contextuality with Pauli observables most pronouncedly illustrates that qubits are different.

B. Why are qubits different?

The difference between qudits and qubits derives from the fact that Heisenberg-Weyl operators behave differently—in ways that matter for contextuality and Wigner functions—depending on whether the Hilbert space dimension is even or odd. The HW operators are important for QCSI because,

by construction, all freely measurable observables are of this type.

Denote by d the dimension of the local Hilbert space, and by \mathbf{x} and \mathbf{z} two HW operators acting on that space,

$$\mathbf{x} := \sum_{i=0}^{d-1} |i + 1 \bmod d\rangle\langle i|, \quad \mathbf{z} := \sum_{i=0}^{d-1} \omega^i |i\rangle\langle i|, \quad (1)$$

where $\omega = e^{2\pi i/d}$. The definition of the HW operator \mathbf{x} in Eq. (1) requires addition mod d , and in result, addition mod d is a common operation when reasoning about Heisenberg-Weyl operators in dimension d . What distinguishes the qubit from the qudit case is that $2^{-1} \bmod d$ is well defined if d is odd, i.e., there exists an $x \in \mathbb{Z}_d$ such that $2x = 1 \bmod d$. But $2^{-1} \bmod d$ is undefined if d is even, specifically if $d = 2$.

Let us see how this mathematical difference is relevant for physics. First, we remark that the nonexistence of the inverse of 2 (mod d) affects the adaption of Wigner functions from the infinite-dimensional to the finite-dimensional case. The standard construction [12,13] works only for odd d , precisely for the above reason.

Here, we discuss the existence of state-independent proofs of contextuality based on HW operators (HW-SIC), such as the Peres-Mermin square [3,28] and Mermin star [3], in tensor product Hilbert spaces of dimension d^n . As we discussed in the introduction, if state-independent contextuality is present, then contextuality cannot be a resource possessed only by some magic states.

The Heisenberg-Weyl operators $\tau_{\mathbf{a}}^\gamma$ in the Hilbert space of n systems of dimension d are

$$\tau_{\mathbf{a}}^\gamma := \omega^{\gamma(\mathbf{a})} Z(\mathbf{a}_Z) X(\mathbf{a}_X),$$

where the function γ is an arbitrary phase convention, $\mathbf{a} = (\mathbf{a}_Z, \mathbf{a}_X)$, and $Z(\mathbf{a}_Z) = \bigotimes_{i=1}^n (\mathbf{z}^{(i)})^{a_{z,i}}$.

There is no HW-SIC if d is odd. First, note that the existence or nonexistence of HW-SIC is independent of the choice of γ . We may thus choose a convenient such function, which for our purposes is

$$\gamma_0(\mathbf{a}) := \mathbf{a}_Z \mathbf{a}_X / 2 \bmod d. \quad (2)$$

This requires the existence of $2^{-1} \bmod d$, and thus only works if d is odd. By direct calculation we find that

$$\tau_{\mathbf{a}}^{\gamma_0} \tau_{\mathbf{b}}^{\gamma_0} = \tau_{\mathbf{a}+\mathbf{b}}^{\gamma_0},$$

whenever $[\tau_{\mathbf{a}}^{\gamma_0}, \tau_{\mathbf{b}}^{\gamma_0}] = 0$. Thus, $\lambda(\tau_{\mathbf{a}}^{\gamma_0}) = 1$ is a consistent context-independent value assignment. Hence, there is no state-independent contextuality with HW operators for odd d , as claimed.

HW-SIC thus points to a difference between odd and even d . As will become clear in the subsequent sections, the presence or absence of state-independent contextuality with respect to Heisenberg-Weyl operators is linked to almost every subject in this paper.

C. Summary of results

Our main results are the following:

(1) *Computational universality.* For all n -qubit QCSI satisfying Condition P1, contextuality of the magic states

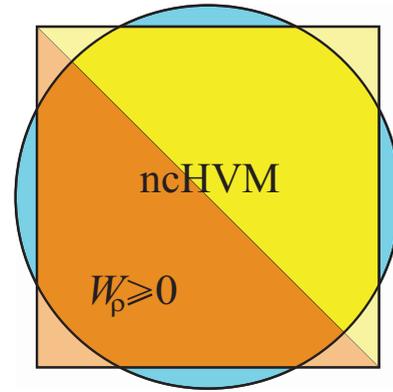


FIG. 1. Schematic view of the n -qubit state space. The disk represents the space of proper quantum states, the square shows the states describable by a noncontextual hidden variable model, and the lower triangle indicates the states with non-negative Wigner function.

is necessary for quantum computational universality (Theorem 4).

(2) *Quantum speedup.* For n -qubit QCSI schemes which satisfy Condition P1 and for which the value assignments of the ncHVM can be efficiently evaluated, contextuality is necessary for speedup (Theorem 5).

(3) *Existence.* There is at least one family of QCSI schemes and matching Wigner function which satisfies Conditions P1 and P2; see Sec. VI.

Points 1 and 2 also hold if “ncHVM” is replaced with “non-negative Wigner function”. However, this notion of classicality is more restricted; see Fig. 1. In contrast to the qudit case, the existence of an ncHVM does not imply non-negativity of the Wigner function. Therefore, in the present setting, non-negativity of the Wigner function is an unnecessarily restrictive notion of classicality.

There are additional points of technical interest:

(4) *Preservation of classicality.* If the input quantum state ρ_{in} of an n -qubit QCSI satisfying Condition P1 can be described by an ncHVM, then the state of the quantum register at all later times can be described by an ncHVM (Theorem 3).

(5) Non-negativity of Wigner functions is in general *not* preserved under free unitary gates for QCSI schemes satisfying P1; see Sec. IV C 2. This does not affect efficient classical simulability (Theorem 1).

(6) For any given n -qubit QCSI, the state space \mathcal{S} of the corresponding ncHVM consists of multiple copies of the phase space $V = \mathbb{Z}_2^n \times \mathbb{Z}_2^n$ on which the Wigner function is defined. If an n -qubit state is noncontextual, its representation by a probability distribution over \mathcal{S} is typically not unique.

Comparing points 4 and 5, we find that existence of an ncHVM for a given state ρ and a non-negative of Wigner function W_ρ are no longer the same, as they were in the qudit case [15,24]. Contextuality still implies Wigner negativity (Theorem 2), but not the other way around.

Remark. In our results on efficient simulation by sampling (Theorems 1 and 5), we assume the sampling sources as given, and only count the operational cost of processing the samples in the simulation. This assumption holds, for example, when each magic state injected to the computation has support only on a bounded number of qubits [16,17]. However, there is

strong indication that probability distributions exist which can be efficiently prepared by quantum means but are hard to sample from classically [29–34]. In view of those, Theorems 1 and 5 specify the computational cost of classical simulation relative to a sampling source, similar to the complexity of an algorithm relative to an oracle.

D. Outline

This paper is structured as follows. In Secs. III–V, we analyze the general structure of QCSI schemes defined by Conditions P1 and P2, and in Sec. VI we explicitly construct a QCSI scheme on qubits for which contextuality in the magic states is a necessary quantum mechanical resource. Regarding the former part, in Sec. III, we work out the implications of Conditions P1 and P2 for QCSI schemes. We give a prescription for how to construct QCSI schemes starting from the phase convention γ for the Heisenberg-Weyl operators. Section IV discusses the role of Wigner functions for QCSI. In particular, we present an efficient classical simulation of QCSI for magic states with non-negative Wigner function (Algorithm 1). Section V is on the role of contextuality. We show that state-independent contextuality is absent from all QCSI schemes satisfying Conditions P1 and P2, clarify the relation between Wigner function negativity and state-dependent contextuality, and establish the latter as a necessary resource for QCSI with magic states. Finally, we describe an efficient classical simulation algorithm for QCSI for magic states with a noncontextual HVM (Algorithm 2). It contains Algorithm 1 as a special case. We conclude in Sec. VII.

We also refer to a companion paper [35] of this article which focuses solely on the role of contextuality in QCSI. Wigner functions—and all the conceptual puzzles they give rise to in the multiqubit setting—are bypassed. The flip side of this approach is that Wigner functions can no longer be used to characterize the near-classical free sector of operations in QCSI. Instead, the free sector is specified by the set of available measurements, and Condition P1 is replaced by the requirement that the available measurements do not exhibit state-independent contextuality. Reference [35] provides a shorter approach for those readers whose main interest is in contextuality.

III. COMPUTATIONAL SETTING AND CONSISTENCY CONDITIONS

In this section we describe how qubit QCSI schemes are constructed starting from Conditions P1 and P2. We demonstrate that by requiring P1 and P2, the choice of Wigner function completely determines the free sector of the corresponding QCSI scheme. Namely, P1 defines the set \mathcal{O} of observables that can be measured for free. Then, the free unitaries are the maximal set of Clifford unitaries that map the set \mathcal{O} to itself under conjugation. Condition P2 is merely used to select the true n -qubit schemes. This construction is explained in Secs. III C–III G.

In Sec. III A, we briefly review the model of QCSI. In Sec. III B, we discuss the general concept of an operational restriction, how it overcomes the phenomenon of state-independent contextuality, and why that is necessary for

establishing contextuality of the magic states as a resource for QCSI.

A. The computational setting

Every QCSI scheme consists of four constituents, namely (i) a set Ω of states that can be prepared within the scheme (the free states), (ii) the set \mathcal{O} of observables which can be directly measured and which in the present discussion always consists solely of Pauli operators, (iii) a group G of unitary gates (the free gates), typically taken as the Clifford group or a subgroup thereof, and (iv) the set \mathcal{M} of magic states which render the scheme computationally universal. A general QCSI scheme is thus characterized by the quadruple $(\mathcal{O}, G, \Omega, \mathcal{M})$.

The first three of these four constituents are considered free. The justification for this terminology is that quantum computations built solely from the free operations cannot have a quantum speedup. This near-classicality of the free operations is made precise by an efficient classical simulation algorithm (see Sec. IV). It states that if the Wigner function of the initial quantum state ρ_{in} can be efficiently sampled from then so can the outcome distribution resulting from evolving ρ_{in} under the free unitary gates and measurements. This simulation result is the very justification for invoking a Wigner function in the description of QCSI.

B. Operational restrictions

When transitioning from local systems of odd prime Hilbert space dimension (qudits) to local systems of Hilbert space dimension 2 (qubits), one encounters a new phenomenon: state-independent contextuality among Pauli observables [3,36]. It is incompatible with viewing contextuality as a resource injected into the computation along with the magic states.

The reasons are twofold. First, within the framework of QCSI, Pauli measurements are supposed to be free, and if contextuality is already present in those operations, how can it be a resource? Perhaps even worse, for systems of two or more qubits, a contextuality witness can be constructed that classifies *all* quantum states of $n \geq 2$ qubits as contextual [15], including the completely mixed state. Again, how can contextuality be a resource if it is generic?

In this paper, the strategy for coping with state-independent contextuality is to place operational restrictions on the Pauli observables that can be measured in a QCSI scheme. The very concept of QCSI already invokes the notion of an operational restriction, since the operations in QCSI are nonuniversal by design. Here, additional constraints are placed by Condition P1. The rebit case [16] shall serve as a model scenario for the concept of operational restrictions, and we briefly review it for illustration.

The Peres-Mermin square [3,28,37] embeds into real quantum mechanics (see Fig. 2), and confining it to rebit quantum states does therefore not remove state independence all by itself. Rather, the following operational restriction is put in place. The directly measurable observables are restricted from the set of real Pauli operators to tensor products of Pauli operators Z_i only or X_i only. Accordingly, the free unitaries are restricted from all real Clifford gates to those which

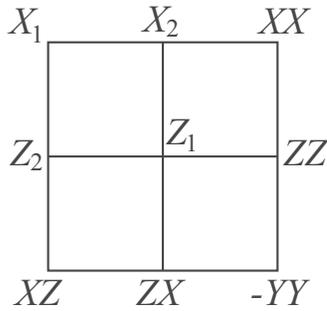


FIG. 2. Peres-Mermin square. For the restriction to CSS-ness-preserving operations, the six observables in the top two rows are in the set \mathcal{O} while the observables in the bottom row are in M but not in \mathcal{O} . In rebit QCSI, they can be measured individually but not jointly. The figure is adapted from Ref. [3].

preserve the set of Calderbank-Shor-Steane (CSS) stabilizer states [38,39].

Let us analyze the free measurements, for the case of $n = 2$ rebits, with the CSS restriction. The set \mathcal{O} of directly measurable observables is

$$\mathcal{O} = \{I, Z_1, Z_2, Z_1 Z_2, X_1, X_2, X_1 X_2\} \times \{\pm 1\}.$$

By directly measuring observables from the above set, measurement outcomes of further Pauli observables can be inferred. For example, $X_1 Z_2 \notin \mathcal{O}$. Yet, a value for $X_1 Z_2$ can be inferred by measuring the commuting observables X_1 and Z_2 separately, and multiplying the outcomes. By applying this construction to all possible pairs of commuting observables in \mathcal{O} , we find the set M of observables whose value can be inferred, namely

$$M = \mathcal{O} \cup \{X_1 Z_2, Z_1 X_2, Y_1 Y_2\} \times \{\pm 1\}.$$

M is thus the set of all real and Hermitian two-qubit Pauli operators. By measurement of observables in the smaller set \mathcal{O} it is thus possible to fully reconstruct all two-rebit density operators.

The next question of interest is which Pauli operators can be measured *jointly*. For example, while both the observables $X_1 Z_2$ and $Z_1 X_2$ are in M and even though they commute, in rebit QCSI they cannot have their values inferred simultaneously. Inferring the value of $X_1 Z_2$ necessitates the physical measurement of the observables X_1 and Z_2 , and inferring the value for $Z_1 X_2$ requires the physical measurement of Z_1 and X_2 . However, the four observables X_1 , Z_2 , Z_1 , and X_2 do not all commute. The measurement of Z_1 and X_2 to infer the outcome of $Z_1 X_2$ wipes out the value of $X_1 Z_2$, and vice versa.

The fact that the observables $X_1 Z_2$ and $Z_1 X_2$ cannot have their values inferred simultaneously is critical for state-independent contextuality. Namely, the consistency constraint among measurement outcomes for observables in the bottom row of the Mermin Peres square can no longer be experimentally checked and is thus effectively removed from the square. As a consequence, the remaining available measurements can be described by a noncontextual hidden variable model (HVM). For example, the value assignment $\lambda = 1$ for all observables in the Peres-Mermin square becomes consistent.

In this way, by imposing an operational restriction, state-independent contextuality disappears from QCSI.

This concludes the review of the rebit case. In the subsequent sections, we generalize the notions introduced above and apply them to a wider range of settings. As a final remark, earlier in this section we stated that the operational restrictions must obey certain consistency conditions. The above discussion points to two of them: To give rise to a tomographically complete scheme of QCSI on qubits, the set \mathcal{O} of directly measurable observables must be large enough for the derived set M to comprise all Pauli operators. At the same time, \mathcal{O} must be small enough to dispense with state-independent contextuality.

C. Consistency conditions on G and Ω

We now begin to describe the consistency conditions which must hold between the group G of free unitary gates in QCSI, the set \mathcal{O} of directly measurable observables, and the set Ω of free states. We require that these constituents of QCSI satisfy two constraints, namely

$$g^\dagger O g \in \mathcal{O}, \quad \forall O \in \mathcal{O}, \quad \forall g \in G \quad (3)$$

and

$$g|\psi\rangle \in \Omega, \quad \forall |\psi\rangle \in \Omega, \quad \forall g \in G. \quad (4)$$

Regarding Eq. (3), if O can be measured, so can $g^\dagger O g$, namely by first applying g , then measuring O , and then applying g^\dagger . Likewise, if $|\psi\rangle$ can be prepared, so can $g|\psi\rangle$.

We regard the set \mathcal{O} of directly measurable observables as primary among the constituents of the free sector of QCSI, and define the group G of free gates and the set Ω of free states in reference to it. Namely, G is the largest subgroup of the n -qubit Clifford group Cl_n that satisfies the property Eq. (3),

$$G := \{g \in Cl_n \mid g^\dagger O g \in \mathcal{O}, \forall O \in \mathcal{O}\}. \quad (5)$$

The free states are those that can be prepared by measurement of observables in \mathcal{O} . All other states are considered resources and must be provided externally if needed. That is, $|\psi\rangle \in \Omega$ if and only if there exists an ordered set $\mathcal{O}_{|\psi\rangle} \subset \mathcal{O}$ such that

$$|\psi\rangle\langle\psi| \sim \left(\prod_{O \in \mathcal{O}_{|\psi\rangle}} \left[\frac{I \pm O}{2} \right] \right) (I/2^n). \quad (6)$$

The projectors on the left-hand side of Eq. (6) do not necessarily commute. Their temporal order is specified by the ordering in $\mathcal{O}_{|\psi\rangle}$. The angular brackets denote superoperators. With Eq. (5), $\mathcal{P}_n \subset G$ always holds. Therefore, a totally depolarizing twirl may be implemented, producing $I/2^n$ from any n -qubit state.

The free sector of a QCSI scheme is thus fully specified via Eqs. (5) and (6) by the set \mathcal{O} of directly measurable observables. In Sec. III E we turn to the question of how \mathcal{O} itself is constructed.

D. Wigner functions

A Wigner function is a means of description of QCSIs. The reason for invoking Wigner functions is to characterize the near classicality of the sector of free operations in QCSI. This

proceeds by way of the efficient classical simulation algorithm described in Sec. IV B.

The Wigner functions considered here are defined on a phase space $V := \mathbb{Z}_2^n \times \mathbb{Z}_2^n$, starting from the Heisenberg-Weyl operators

$$T_{\mathbf{a}}^\gamma = i^{\gamma(\mathbf{a})} Z(\mathbf{a}_Z) X(\mathbf{a}_X). \quad (7)$$

Therein, $Z(\mathbf{a}_Z) := \bigotimes_{i=1}^n Z_i^{\mathbf{a}_Z \cdot \mathbf{z}_i}$, $X(\mathbf{a}_X) := \bigotimes_{i=1}^n X_i^{\mathbf{a}_X \cdot \mathbf{x}_i}$.

The possible phase conventions $\gamma : V \rightarrow \mathbb{Z}_4$ are constrained only by the requirement that all $T_{\mathbf{a}}^\gamma$, $\mathbf{a} \in V$, are Hermitian. As we show later, the QCSI schemes considered here and the Wigner functions describing them are both fully specified by γ .

We consider Wigner functions of the form $W_\rho^\gamma(\mathbf{u}) = 1/2^n \text{Tr}(A_{\mathbf{u}} \rho)$, for all $\mathbf{u} \in V = \mathbb{Z}_2^{2n}$, where $A_{\mathbf{u}} = T_{\mathbf{u}}^\gamma A_{\mathbf{0}} (T_{\mathbf{u}}^\gamma)^\dagger$,

$$A_{\mathbf{0}}^\gamma = \frac{1}{2^n} \sum_{\mathbf{a} \in V} T_{\mathbf{a}}^\gamma. \quad (8)$$

This definition satisfies the minimal conditions required of a Wigner function [11], namely that (i) W^γ is a quasiprobability distribution defined on a state space $V = \mathbb{Z}_2^{2n}$, (ii) W^γ transforms covariantly under the Pauli group, $W_{T_{\mathbf{a}} \rho T_{\mathbf{a}}^\dagger}^\gamma(\mathbf{u}) = W_\rho(\mathbf{u} + \mathbf{a})$, for all $\mathbf{u}, \mathbf{a} \in V$, and (iii) there is a suitable notion of marginals.

Remark. To simplify the notation, we subsequently omit the superscript γ in the Wigner functions, unless the precise choice of γ matters.

All previous works on the role of positive Wigner functions for QCSI—Refs. [16–19]—are based on a particular family of Wigner functions for finite-dimensional state spaces introduced by Gibbons *et al.* [11]. This is, indirectly, also the case for the present Wigner function, and we therefore briefly describe its genealogy. Gibbons *et al.* introduced a family of Wigner functions for finite-dimensional state spaces based on the concepts of mutually unbiased bases and lines in phase space. Among this family, for the special case of odd local dimension, Gross [12] identified a Wigner function which is the most sensible finite-dimensional analog of the infinite-dimensional case [10]. This Wigner function was written in the form of Eqs. (7) and (8) in Ref. [17], with a special phase convention γ . For local Hilbert space dimension 2, this special function γ does not exist, and in the present approach γ is left as a parameter to vary. The freedom of choosing the function γ replaces the freedom of choosing quantum nets in Ref. [11].

In addition to the above properties (i)–(iii), the Wigner functions defined in Eqs. (7) and (8) have two further relevant properties. First, for any pair ρ and σ of operators acting on the Hilbert space \mathbb{C}^{2^n} , it holds that

$$\text{Tr}(\rho\sigma) = 2^n \sum_{\mathbf{u} \in V} W_\rho(\mathbf{u}) W_\sigma(\mathbf{u}). \quad (9)$$

Second, for any admissible function γ , we have the following relation between a quantum state ρ and its Wigner functions W_ρ :

$$\rho = \sum_{\mathbf{u} \in V} W_\rho(\mathbf{u}) A_{\mathbf{u}}.$$

E. Condition P1 defines \mathcal{O}

The set \mathcal{O} of directly measurable observables is, in the present setting, always a set of Hermitian Pauli operators,

$$\mathcal{O} = \{\pm T_{\mathbf{a}}, \mathbf{a} \in V_{\mathcal{O}}\},$$

where $V_{\mathcal{O}}$ is a subset of $V = (\mathbb{Z}_2)^{2n}$.

In Sec. III C, we described how to construct the set Ω of free states given the set \mathcal{O} of directly measurable observables. But how is the set \mathcal{O} itself constructed? To answer this question, we return to the function γ in Eq. (7) from which everything follows in the present setting. The function $\gamma : V \rightarrow \mathbb{Z}_4$ specifies a function $\beta : V \times V \rightarrow \mathbb{Z}_4$ defined via

$$T_{\mathbf{a}+\mathbf{b}} = i^{\beta(\mathbf{a},\mathbf{b})} T_{\mathbf{a}} T_{\mathbf{b}}. \quad (10)$$

The function β is related to the function γ introduced in Eq. (7) via

$$\beta(\mathbf{a}, \mathbf{b}) = -\gamma(\mathbf{a}) - \gamma(\mathbf{b}) + \gamma(\mathbf{a} + \mathbf{b}) + 2\mathbf{a}_X \mathbf{b}_Z \pmod{4}. \quad (11)$$

Thus, β is fully determined by γ . The converse is not true. Every valid function β (i.e., one that derives from a function γ) determines γ only up to a translation in phase space.

The function β constrains the Pauli operators that can possibly be contained in the set \mathcal{O} . Namely, we have the following lemma:

Lemma 1. For any $\mathbf{a} \in V$, the measurement of an observable $\pm T_{\mathbf{a}}$ does not introduce negativity into the Wigner function if and only if

$$\beta(\mathbf{a}, \mathbf{b}) = 0, \quad \forall \mathbf{b} \in V \mid [\mathbf{a}, \mathbf{b}] = 0. \quad (12)$$

In Eq. (12), $[\cdot, \cdot]$ is the symplectic bilinear form defined by $[\mathbf{a}, \mathbf{b}] := \mathbf{a}_X \cdot \mathbf{b}_Z + \mathbf{a}_Z \cdot \mathbf{b}_X \pmod{2}$, for all $\mathbf{a}, \mathbf{b} \in V$.

Proof of Lemma 1. “Only if”: Assume that the condition Eq. (12) does not hold, i.e., there exists a Pauli operator $T_{\mathbf{b}}$ such that $[\mathbf{a}, \mathbf{b}] = 0$ and $\beta(\mathbf{a}, \mathbf{b}) \neq 0 = 2$ (Hermiticity).

Further assume that the system is in the mixed state $(I - T_{\mathbf{b}})/2^n$, which has non-negative W , and that $T_{\mathbf{a}}$ is measured. Without loss of generality (w.l.o.g.) assume that the outcome is -1 . The resulting state is $\rho = (I - T_{\mathbf{a}} - T_{\mathbf{b}} + T_{\mathbf{a}} T_{\mathbf{b}})/2^n = (I - T_{\mathbf{a}} - T_{\mathbf{b}} - T_{\mathbf{a}+\mathbf{b}})/2^n$. Thus, $W_\rho(\mathbf{0}) = -2/4^n < 0$. Thus, if $\beta(\mathbf{a}, \mathbf{b}) \neq 0$ for some $\mathbf{b} \in V$, the measurement of $T_{\mathbf{a}}$ can introduce negativity into Wigner functions, and hence $\pm T_{\mathbf{a}} \notin \mathcal{O}$. Negation of this statement proves the result.

“If”: We assume that the Wigner function W_ρ of the state ρ before the measurement is non-negative,

$$W_\rho(\mathbf{u}) \geq 0, \quad \forall \mathbf{u} \in V,$$

and that the measured observable $T_{\mathbf{a}}$ is such that $\beta(\mathbf{a}, \mathbf{b}) = 0$, for all $\mathbf{b} \in V$. The state ρ' after the measurement of the observable $T_{\mathbf{a}}$ with outcome $s \in \{0, 1\}$ is $\rho' \sim \frac{I + (-1)^s T_{\mathbf{a}}}{2} \rho \frac{I + (-1)^s T_{\mathbf{a}}^\dagger}{2}$, and the value of the corresponding Wigner function at the phase space point $\mathbf{u} \in V$ is

$$p_{\mathbf{a}}(s) W_{\rho'}(\mathbf{u}) = \frac{1}{2^n} \text{Tr} \left(\frac{I + (-1)^s T_{\mathbf{a}}^\dagger}{2} A_{\mathbf{u}} \frac{I + (-1)^s T_{\mathbf{a}}}{2} \rho \right). \quad (13)$$

Therein, $p_a(s)$ is the probability of obtaining the outcome s in the measurement of T_a . Now,

$$\begin{aligned} & \frac{I + (-1)^s T_a^\dagger}{2} A_u \frac{I + (-1)^s T_a}{2} \\ &= \frac{I + (-1)^s T_a^\dagger}{2} \left[\frac{1}{2^n} \sum_{\mathbf{b} \in V} (-1)^{[\mathbf{u}, \mathbf{b}]} T_{\mathbf{b}} \right] \frac{I + (-1)^s T_a}{2} \\ &= \frac{I + (-1)^s T_a}{2} \left[\frac{1}{2^n} \sum_{\mathbf{b} \in V | [\mathbf{b}, \mathbf{a}] = 0} (-1)^{[\mathbf{u}, \mathbf{b}]} T_{\mathbf{b}} \right] \\ &= \frac{1}{2^{n+1}} \sum_{\mathbf{b} \in V | [\mathbf{b}, \mathbf{a}] = 0} (-1)^{[\mathbf{u}, \mathbf{b}]} [T_{\mathbf{b}} + (-1)^s T_a T_{\mathbf{b}}] \\ &= \frac{1}{2^{n+1}} \sum_{\mathbf{b} \in V | [\mathbf{b}, \mathbf{a}] = 0} (-1)^{[\mathbf{u}, \mathbf{b}]} [T_{\mathbf{b}} + (-1)^s T_{\mathbf{a} + \mathbf{b}}] \\ &= \frac{1}{2^{n+1}} \sum_{\mathbf{b} \in V | [\mathbf{b}, \mathbf{a}] = 0} (-1)^{[\mathbf{u}, \mathbf{b}]} [1 + (-1)^{s + [\mathbf{a}, \mathbf{u}]}] T_{\mathbf{b}} \\ &= \frac{\delta_{s, [\mathbf{a}, \mathbf{u}]}}{2^n} \sum_{\mathbf{b} \in V | [\mathbf{b}, \mathbf{a}] = 0} (-1)^{[\mathbf{u}, \mathbf{b}]} T_{\mathbf{b}} \\ &= \frac{\delta_{s, [\mathbf{a}, \mathbf{u}]}}{2} (A_{\mathbf{u}} + A_{\mathbf{u} + \mathbf{a}}). \end{aligned}$$

Above, we have used the assumption that $\beta(\mathbf{a}, \mathbf{b}) = 0$ for all $\mathbf{b} \in V$ when transitioning from line 4 to line 5. Applying the result to Eq. (13), we find that

$$p_a(s) W_{\rho'}(\mathbf{u}) = \frac{\delta_{s, [\mathbf{a}, \mathbf{u}]}}{2} [W_{\rho}(\mathbf{u}) + W_{\rho}(\mathbf{u} + \mathbf{a})]. \quad (14)$$

By assumption, the right-hand side is always non-negative. For the outcome s to possibly occur, it is required that $p_a(s) > 0$. Hence, $W_{\rho'}(\mathbf{u}) \geq 0$, for all $\mathbf{u} \in V$. ■

We define the set \mathcal{O} of directly measurable observables to be the largest possible set of Pauli observables allowed by Lemma 1,

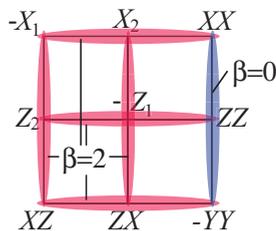
$$\mathcal{O} := \{\pm T_{\mathbf{a}}, \mathbf{a} \in V | \beta(\mathbf{a}, \mathbf{u}) = 0, \forall \mathbf{u} \in V_{\mathbf{a}}\}, \quad (15)$$

with $V_{\mathbf{a}} := \{\mathbf{u} \in V | [\mathbf{a}, \mathbf{u}] = 0\}$.

Example. To illustrate the usefulness of Lemma 1, consider the following choice for γ . For brevity, we restrict to two rebits. W^γ is specified by

$$\begin{aligned} A_0^\gamma &= \frac{1}{4} (I - Z_1 + Z_2 + Z_1 Z_2 - X_1 + X_2 + X_1 X_2 \\ &\quad + X_1 Z_2 + Z_1 X_2 - Y_1 Y_2). \end{aligned}$$

Arranging all observables in A_0 apart from the identity in the Peres-Mermin square,



it is evident that every observable $T_{\mathbf{a}}$ is part of at least one commuting triple with $\beta \neq 0$. Hence, apart from the identity,

no observable is in \mathcal{O} , i.e., $\mathcal{O} = \{I\}$. The corresponding QCSI scheme is thus the exact opposite of tomographically complete: Nothing can be measured at all. We find that not for every function γ the Wigner function W^γ can be paired with a matching QCSI scheme.

For further illustration of Lemma 1, we have the following implication:

Lemma 2. Consider a Wigner function as defined in Eqs. (8) and (7), for $n \geq 2$ qubits. Then, there always exists a Pauli observable whose measurement does not preserve positivity.

Thus, no QCSI scheme in which all Pauli observables are directly measurable can satisfy the Property P1. The original QCSI scheme [20] on qubits is one of those schemes, and it is therefore out of scope of the present analysis.

Proof of Lemma 2. Consider the Peres-Mermin square as displayed in Fig. 2. Irrespective of the sign conventions of the Pauli observables contained in it, there is always at least one context with $\beta = 2$. Thus, by Lemma 1, for $n \geq 2$ the measurement of at least three Pauli observables introduces negativity into the Wigner function. ■

Furthermore, regarding the free states, an immediate consequence of Property P1 is that all free states $|\Psi\rangle \in \Omega$ are non-negatively represented by W . All free states $|\Psi\rangle \in \Omega$ can be created from the completely mixed state $I/2^n$, by measurement of observables in \mathcal{O} , and $W_{I/2^n} \geq 0$ for any γ . Then, with Property P1, $W_{|\Psi\rangle\langle\Psi|} \geq 0$, for all $|\Psi\rangle \in \Omega$. By Eq. (4), free states remain non-negatively represented upon action of free unitary gates $g \in G$.

F. Inferability and tomographic completeness

We know from Lemma 2 that for $n \geq 2$ qubits, the set \mathcal{O} of directly measurable observables is always strictly smaller than the set of all Pauli observables. How is that not in conflict with tomographic completeness (Property P2)?

The reason—as was already mentioned in Sec. III B—is that some Pauli observables, while not directly measurable, can nonetheless have their value inferred by measurement. For example, consider three Pauli observables $T_a, T_b \in \mathcal{O}, T_c \notin \mathcal{O}$, such that $[a, b] = 0$ and $T_c = T_a T_b$. Then T_c can have its value inferred by measuring T_a and T_b , and multiplying the outcomes.

For suitable sets \mathcal{O} , all Pauli observables can have their value inferred, even if not directly measured. This suffices to satisfy Property P2.

Definition 1. $M = \{\pm T_{\mathbf{a}} | \mathbf{a} \in V_M\}$ is the set of Pauli observables whose value can be inferred from a single copy of the given quantum state, by measurement of other Pauli observables from the set \mathcal{O} and classical postprocessing.

The set M is typically larger than the set \mathcal{O} of observables which can be directly measured. This was illustrated by an example in Sec. III B, namely $\mathcal{O} = \{I, X_1, Z_2\}, M = \mathcal{O} \cup \{X_1 Z_2\}$. In terms of V_M , Condition P2 of tomographic completeness reads

$$V_M = V. \quad (16)$$

We now provide a general characterization of the set M generated by the set \mathcal{O} .

Lemma 3. For any γ , the set V_M has the properties that (i) $V_O \subseteq V_M$, and (ii) for any $\mathbf{a} \in V_O$, $\mathbf{b} \in V$ with $[\mathbf{a}, \mathbf{b}] = 0$, it holds that $\mathbf{a} + \mathbf{b} \in V_M$ if and only if $\mathbf{b} \in V_M$.

Proof of Lemma 3. Property (i) merely states that what can be directly measured can have its value inferred. Regarding (ii), the observable $T_{\mathbf{a}+\mathbf{b}}$ has its value inferred as follows. First, $T_{\mathbf{a}} \in \mathcal{O}$ is measured directly. Then, the procedure for inferring the value of $T_{\mathbf{b}}$ is applied. Since $T_{\mathbf{a}}$ commutes with $T_{\mathbf{b}}$, the former measurement does not interfere with the latter, and $\mu(T_{\mathbf{a}}T_{\mathbf{b}}) = \mu(T_{\mathbf{a}})\mu(T_{\mathbf{b}})$. Finally, with Eq. (12), $\mu(T_{\mathbf{a}+\mathbf{b}}) = \mu(T_{\mathbf{a}})\mu(T_{\mathbf{b}})$. Thus, if $\mathbf{b} \in V_M$ then $\mathbf{a} + \mathbf{b} \in V_M$. The reverse direction holds by symmetry in $\mathbf{b} \longleftrightarrow \mathbf{a} + \mathbf{b}$. ■

Example. Assume that $X_1, Z_2, Y_1Y_2 \in \mathcal{O}$. The outcome of the observable Z_1X_2 can then be inferred by measurement, i.e., $Z_1X_2 \in M$. The procedure for the measurement of the observable Z_1X_2 , given the above set \mathcal{O} of directly measurable observables, is the following. First, the observable Y_1Y_2 is measured, and second, the commuting observables X_1 and Z_2 are measured. The measurement outcome $\mu(Z_1X_2) \in \{\pm 1\}$ then is

$$\mu(Z_1X_2) = \mu(Y_1Y_2)\mu(X_1)\mu(Z_2).$$

The key point of this example is that not all pairs among the measured Pauli observables X_1 , Z_2 , and Y_1Y_2 commute, yet in the above expression for $\mu(Z_1X_2)$ we treated them as if they did. The reason that this is possible is the following:

Since Y_1Y_2 does not commute with X_1 and Z_2 , the measurements of X_1 and Z_2 after the measurement of Y_1Y_2 —if taken separately—do not reveal any information about the initial state. Individually, their outcomes are completely random, whatever the state prior to the Y_1Y_2 measurement is. However, X_1 and Z_2 mutually commute, and hence the separate measurement of X_1 and Z_2 implies a valid measurement outcome for the correlated observable X_1Z_2 , namely $\mu(X_1Z_2) = \mu(X_1)\mu(Z_2)$. Furthermore, since X_1Z_2 does commute with Y_1Y_2 , $\mu(X_1)\mu(Z_2)$ represents the outcome of a X_1Z_2 measurement on the initial state, and $\mu(Z_1X_2) = \mu(Y_1Y_2)\mu(X_1Z_2) = \mu(Y_1Y_2)\mu(X_1)\mu(Z_2)$, as claimed.

Let us now verify that $Z_1X_2 \in M$ follows from the properties established in Lemma 3. First, with property (i) of Lemma 3, $X_1 \in \mathcal{O}$ implies $X_1 \in M$. Then, using property (ii) with $X_1 \in M$, $Z_2 \in \mathcal{O}$, it follows that $X_1Z_2 \in M$. Finally, again with property (ii), since $Y_1Y_2 \in \mathcal{O}$ and $X_1Z_2 \in M$, it follows that $Z_1X_2 \in M$.

We note that the above procedure of inferring measurement outcomes by the physical measurement of noncommuting observables is reminiscent of the syndrome measurement in subsystem codes [40,41], with the Bacon-Shor code [42,43] and topological subsystem codes [44] as prominent examples.

Back to the general scenario, an observable $T_{\mathbf{a}}$ can have its value inferred, i.e., $T_{\mathbf{a}} \in M$, if there exists a resolution

$$\mathbf{a} = \mathbf{a}_1 + \{\mathbf{a}_2 + [\mathbf{a}_3 + \cdots (\mathbf{a}_{N-1} + \mathbf{a}_N) \cdots]\}, \quad (17)$$

where all $\mathbf{a}_i \in V_O$, and

$$\left[\mathbf{a}_i, \sum_{j=i+1}^N \mathbf{a}_j \right] = 0, \quad \forall i = 1, \dots, N-1. \quad (18)$$

The resolution, Eq. (17), of \mathbf{a} represents a measurement sequence for inferring the value of $T_{\mathbf{a}}$, starting with the measurement of \mathbf{a}_1 and ending with the measurement of \mathbf{a}_N . The inferred value is $\lambda(T_{\mathbf{a}}) = \prod_{i=1}^N \lambda(T_{\mathbf{a}_i})$.

Finally, we introduce a generalization of the set M of Pauli observables whose value can be inferred. Namely, we denote by C , $C \subset M$, a set of observables whose value can be inferred jointly in QCSI. For short, we call such a set C a “set of jointly measurable observables.”

Definition 2. A set C , $C \subset M$, of commuting Pauli observables is jointly measurable if the outcomes for all observables in C can be simultaneously inferred from measurement of observables in \mathcal{O} on a single copy of the given quantum state ρ .

The sets C of simultaneously measurable observables will become important in Sec. V C, where we discuss the relation between contextuality and negativity of Wigner functions.

Some examples for possible sets C are (i) $C = \{O\}$, for any $O \in M$, (ii) any commuting subset of \mathcal{O} , and (iii) $C = \{A, B, AB\}$, for $A \in M$, $B \in \mathcal{O}$ and $[A, B] = 0$.

We have the following characterization of the sets C of simultaneously measurable observables.

Lemma 4. Consider a set C of simultaneously measurable observables, and $T_{\mathbf{a}}, T_{\mathbf{b}} \in C$. Then, $T_{\mathbf{a}+\mathbf{b}} = T_{\mathbf{a}}T_{\mathbf{b}}$.

We postpone the proof of Lemma 4 until Sec. V B.

G. Constructing QCSI schemes from γ

Once the phase convention γ for the n -qubit Pauli operators is given [cf. Eq. (7)], the Wigner function Eq. (8) and the free sector of the corresponding QCSI scheme satisfying P1 are fully specified. They are obtained through the following steps:

(1) Construct the Wigner function W via its definition Eqs. (7) and (8).

(2) Compute the function β defined in Eq. (10) from the function γ . Construct the set \mathcal{O} of directly measurable observables via Eq. (15).

(3) Construct the group G of free unitary gates via Eq. (5) and set Ω of free states via Eq. (6).

In addition, for tomographic completeness (Condition P2) it needs to be checked that every Pauli observable $T_{\mathbf{a}}$ has a resolution Eq. (17).

To summarize, in this section we have stated minimal requirements for any QCSI scheme on qubits and its corresponding Wigner function. We have shown that once the function γ is provided, the free sector of the corresponding QCSI scheme is fully specified. To reflect this fact in our notation, we subsequently denote QCSI schemes by (γ, \mathcal{M}) instead of $(\mathcal{O}, G, \Omega, \mathcal{M})$. Implicit in this notation is that matching pairs of a QCSI scheme (γ, \mathcal{M}) and a Wigner function W^γ satisfy P1.

IV. EFFICIENT CLASSICAL SIMULATION OF QCSI FOR NON-NEGATIVE STATES

The purpose of this section is to clarify the role of Wigner functions for QCSI schemes satisfying Condition P1. Wigner functions endow the free operations with a notion of classicality, based on efficient simulability by sampling. In Secs. IV A and IV B we demonstrate the following result.

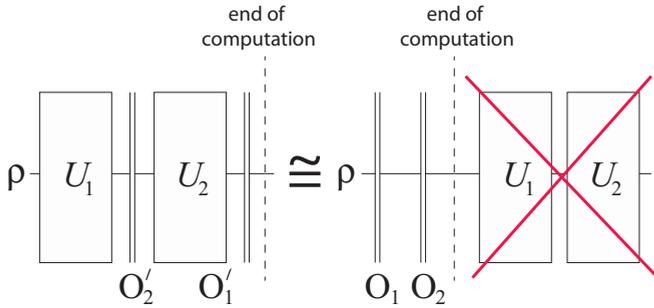


FIG. 3. The measurements of observables O'_i are propagated backwards in time to act on the initial state, by conjugation under the interspersed unitaries. Since only the measurement statistics is of interest, the trailing unitaries may be removed from the resulting circuit.

Theorem 1. For any QCSI scheme (γ, \mathcal{M}) , if (i) the Wigner function $W_{\rho_{\text{in}}}^\gamma$ of the initial state ρ_{in} can be efficiently sampled from, and (ii) the phase convention $\gamma(\mathbf{a})$ can be efficiently evaluated for all $\mathbf{a} \in V_{\mathcal{O}}$, then the distribution of measurement outcomes can be efficiently sampled from.

Wigner functions do the same work as in the qudit case [17]. However, as we show in Sec. IV C, there also are differences. For example, the free unitary gates may introduce negativity into the Wigner function at hand, without compromising efficient classical simulability. The negativity of Wigner functions is thus no witness of quantumness.

A. Reformulation of the simulation problem

For the purpose of classical simulation, we make an alternation to the present QCSI scheme, which, however, does not affect its computational power. Namely, we absorb the unitary gates in the measurements, such that only state preparations and measurements remain as free operations. Here we take the viewpoint that all there is to simulate about a quantum computation is the joint outcome distribution of measurements performed in course of the computation. If the unitaries can be removed without altering the outcome distribution, then there is certainly no loss in removing them. But there is a gain: As is made explicit in Sec. IV C, the simulation algorithm of Sec. IV B can handle free unitaries $g \in G$ that introduce negativity into the Wigner function of the processed quantum state.

The general procedure is outlined in Fig. 3. Consider a QCSI circuit which is an alternation of unitary gates $g_i(\mathbf{s}_{<i})$ and projective measurements represented by projectors $P'_i(\mathbf{s}_{<i}, s_i)$,

$$\mathcal{C} = \prod_{i=1}^{t_{\max}} P'_i(\mathbf{s}_{<i}, s_i) g_i(\mathbf{s}_{<i}), \quad (19)$$

where the factors in the product are ordered from right to left. Therein, \mathbf{s} is the binary vector of all measurement outcomes, and $\mathbf{s}_{<i}$ is \mathbf{s} restricted to the measurement outcomes obtained prior to the gate g_i . We thus allow unitary gates and measurements to depend on previously obtained measurement outcomes. Such conditioning is essential for the working of QCSI.

Now denote by $G_i(\mathbf{s}_{<i})$ the unitaries accumulated up to step i ,

$$G_i(\mathbf{s}_{<i}) = \prod_{j=1}^i g_j(\mathbf{s}_{<j}). \quad (20)$$

Therein, the ordering of operations is the same as in Eq. (19). The circuit \mathcal{C} of Eq. (19) may then be rewritten as

$$\begin{aligned} \mathcal{C} &= \prod_{i=1}^{t_{\max}} P'_i(\mathbf{s}_{<i}, s_i) g_i(\mathbf{s}_{<i}) \\ &= G_{t_{\max}}(\mathbf{s}) \prod_{i=1}^{t_{\max}} G_i(\mathbf{s}_{<i})^\dagger P'_i(\mathbf{s}_{<i}, s_i) G_i(\mathbf{s}_{<i}) \\ &\cong \prod_{i=1}^{t_{\max}} G_i(\mathbf{s}_{<i})^\dagger P'_i(\mathbf{s}_{<i}, s_i) G_i(\mathbf{s}_{<i}). \end{aligned}$$

Thus, if the measured observables in the original sequence of operations were $O'_i(\mathbf{s}_{<i})$, the corresponding observables in the equivalent sequence are

$$O_i(\mathbf{s}_{<i}) = G_i(\mathbf{s}_{<i})^\dagger O'_i(\mathbf{s}_{<i}) G_i(\mathbf{s}_{<i}). \quad (21)$$

By Eq. (3), if $O'_i(\mathbf{s}_{<i}) \in \mathcal{O}$, then $O_i(\mathbf{s}_{<i}) \in \mathcal{O}$. Therefore, a QCSI scheme with set \mathcal{O} of measurable observables and group G of unitary gates is equivalent to a QCSI scheme with set \mathcal{O} of measurable observables and no unitaries at all.

B. Simulation algorithm

The classical simulation algorithm for the setting of Theorem 1 is given in Table I.

Any sample $\mathbf{u} \in V$ from a non-negative Wigner function has a definite value assignment for all observables in the Pauli group. Namely, for any $\mathbf{a} \in V$, the measurement outcome for the Pauli observable $T_{\mathbf{a}}$ is

$$\lambda_{\mathbf{u}}(\mathbf{a}) = (-1)^{[\mathbf{a}, \mathbf{u}]}. \quad (22)$$

The value assignment Eq. (22) is a direct consequence of the update rule Eq. (14). Namely, in the left-hand side of Eq. (14) assume that the probability $p_{\mathbf{a}}(s)$ for obtaining the outcome $s \in \{0, 1\}$ in a measurement of a Pauli observable $T_{\mathbf{a}}$ on a state ρ is nonzero. Then, the right-hand side of Eq. (14) implies that $s = [\mathbf{a}, \mathbf{u}]$, or, equivalently, $\lambda_{\mathbf{u}} = (-1)^{[\mathbf{a}, \mathbf{u}]}$.

TABLE I. Algorithm 1 for the classical simulation of n -qubit QCSI with non-negative Wigner function of the initial state.

Algorithm 1

- (1) Draw a sample $\mathbf{v} \in V$ from $W_{\rho_{\text{in}}}$, and set $\mathbf{v}_1 := \mathbf{v}$.
- (2) For all the measurements of observables $T_{\mathbf{a}_i} \in \mathcal{O}$ comprising the circuit, starting with the first,
 - (a) output the result $s_i = [\mathbf{a}_i, \mathbf{v}_i]$ for the measurement of the observable $T_{\mathbf{a}_i}$,
 - (b) flip a fair coin, and update the sample

$$\mathbf{v}_i \longrightarrow \mathbf{v}_{i+1} = \begin{cases} \mathbf{v}_i, & \text{if "heads"} \\ \mathbf{v}_i + \mathbf{a}_i, & \text{if "tails"} \end{cases}$$
 until the measurement sequence is complete.
- (3) Repeat until sufficient statistics is gathered.

For illustration of the state update rule in the above classical simulation algorithm, consider two measurement sequences for the state of one qubit, namely (i) repeated measurements of the Pauli observable $Z = \pm T_{\mathbf{a}(Z)}$ and (ii) alternating measurements of the Pauli observables Z and $X = \pm T_{\mathbf{a}(X)}$. Assume that the sample from the Wigner function of the initial state is $\mathbf{u} \in V$. Regarding (i), according to the classical simulation algorithm, the ontic states after one or a larger number of measurements are \mathbf{u} or $\mathbf{u} + \mathbf{a}(Z)$. Either way, the reported measurement outcome is $(-1)^{[\mathbf{u}, \mathbf{a}(Z)]}$, since $[\mathbf{a}(Z), \mathbf{a}(Z)] = 0$. For any sample \mathbf{u} from the input distribution, the sequence of measurement outcomes is thus constant, as required. Regarding (ii), since $[\mathbf{a}(Z), \mathbf{a}(X)] = 1$, from the second measurement onwards the outcomes produced by the classical simulation are completely random and uncorrelated, as required.

Proof of Theorem 1. The theorem follows from the efficiency and the correctness of the above classical simulation algorithm.

Efficiency. Both the preprocessing of removing the unitaries from the circuit and the classical simulation algorithm itself need to be considered.

(1) *Preprocessing.* We need to track the evolution of Pauli observables $T_{\mathbf{a}}$ under conjugation by gates $g \in G$, as described in Eq. (21). This can be done efficiently within the stabilizer formalism [25]. However, the stabilizer formalism uses its own phase convention $\tilde{\gamma}$ for the Pauli operators, $\tilde{T}_{\mathbf{a}} := i^{\tilde{\gamma}(\mathbf{a})} Z(\mathbf{a}_Z) X(\mathbf{a}_X)$, such that $\tilde{\gamma}$ can be efficiently evaluated. Suppose $g^\dagger \tilde{T}_{\mathbf{a}} g = i^{\tilde{\phi}_g(\mathbf{a})} \tilde{T}_{g(\mathbf{a})}$. Then, $g^\dagger T_{\mathbf{a}} g = i^{\phi_g(\mathbf{a})} T_{g(\mathbf{a})}$, with

$$\phi_g(\mathbf{a}) = \tilde{\phi}_g(\mathbf{a}) + [\tilde{\gamma}(g\mathbf{a}) - \tilde{\gamma}(\mathbf{a})] - [\gamma(g\mathbf{a}) - \gamma(\mathbf{a})].$$

By assumption, γ can be efficiently evaluated, and hence can ϕ_g , for any $g \in G$.

(2) *Classical simulation algorithm.* The efficiency of the above classical simulation algorithm is evident.

Correctness. Assume that the classical simulation algorithm samples correctly from the Wigner function of the state ρ_t after the t th measurement in the sequence. We now show that under this assumption the above classical simulation algorithm (i) produces the correct probability distribution for the $(t+1)$ -th measurement and (ii) correctly samples from the Wigner function of the conditional state $\rho_{t+1}(s_{t+1})$ after the $(t+1)$ -th measurement.

(i) According to the value assignment Eq. (22) of the classical simulation algorithm, the probability $p_{\mathbf{a}}(s)$ for obtaining the outcome $s \in \{0, 1\}$ in the measurement of the observable $T_{\mathbf{a}}$ on the state ρ_t is

$$p_{\mathbf{a}}(s) = \sum_{\mathbf{u} \in V} \delta_{[\mathbf{a}, \mathbf{u}], s} W_{\rho_t}(\mathbf{u}).$$

As is easily verified by direct calculation,

$$W_{\frac{I + (-1)^s T_{\mathbf{a}}}{2}}(\mathbf{u}) = \frac{1}{2^n} \delta_{[\mathbf{a}, \mathbf{u}], s}. \quad (23)$$

Combining the last two equations, and using the property Eq. (9), we find that

$$\begin{aligned} p_{\mathbf{a}}(s) &= 2^n \sum_{\mathbf{u} \in V} W_{\frac{I + (-1)^s T_{\mathbf{a}}}{2}}(\mathbf{u}) W_{\rho_t}(\mathbf{u}) \\ &= \text{Tr} \left(\frac{I + (-1)^s T_{\mathbf{a}}}{2} \rho_t \right), \end{aligned}$$

which is the quantum-mechanical expression.

(ii) Consider the Wigner function W_{ρ_t} for the state ρ_t after step t in the expansion $W_{\rho_t} = \sum_{\mathbf{u} \in V} W_{\rho_t}(\mathbf{u}) \delta_{\mathbf{u}}$. At time $t+1$, the observable $T_{\mathbf{a}}$ is measured, with outcome $s \in \{0, 1\}$. Then, from the value assignment Eq. (22), only the phase space points $\mathbf{u} \in V$ with $s = [\mathbf{u}, \mathbf{a}]$ contribute to conditional density matrix $\rho_{t+1}(s)$. Furthermore, per step 2(b) of the classical simulation algorithm, the update for δ distributions over phase space is $\delta_{\mathbf{u}} \mapsto (\delta_{\mathbf{u}} + \delta_{\mathbf{u}+\mathbf{a}})/2$. Therefore, the Wigner function for the (normalized) conditional state $\rho_{t+1}(s)$, according to the classical simulation algorithm, is

$$p_{\mathbf{a}}(s) W_{\rho_{t+1}(s)} = \sum_{\mathbf{u} \in V} \delta_{[\mathbf{a}, \mathbf{u}], s} W_{\rho_t}(\mathbf{u}) \frac{\delta_{\mathbf{u}} + \delta_{\mathbf{u}+\mathbf{a}}}{2}.$$

Hence, $p_{\mathbf{a}}(s) W_{\rho_{t+1}(s)}(\mathbf{v}) = \delta_{[\mathbf{a}, \mathbf{v}], s} \frac{W_{\rho_t}(\mathbf{v}) + W_{\rho_t}(\mathbf{v}+\mathbf{a})}{2}$, which is the quantum-mechanical expression Eq. (14).

By assumption of Theorem 1, the Wigner function of the initial state $\rho_{\text{in}} = \rho_0$ is correctly sampled from. Thus, with the above statements (i) and (ii), it follows by induction that all sequences of measurement outcomes occur with the correct probabilities. \blacksquare

C. Discussion

A notable property of the above simulation method is that, for any Wigner function employed therein, covariance and preservation of positivity under the group G of free gates are not required. This is a consequence of the reformulation of QCSI in Sec. IV A, where the free unitary gates were eliminated. It is in sharp contrast to the previously considered cases of qudits [15, 17] and rebits [16], where covariance and preservation of positivity under G were critical for the classical simulation by sampling. These points, and the roles remaining for covariance and positivity preservation in the present simulation method, are discussed below.

1. Covariance

As an example, consider a quantum circuit for a single qubit consisting of a Hadamard gate followed by a measurement of the Pauli observable Z . Given is a source that samples from the non-negative Wigner function of the initial state ρ_{in} , and the task is to sample from the output distribution of the measurement.

A classical simulation method based on Wigner function covariance would, in the first step, convert the source that samples from $W_{\rho_{\text{in}}}$ into a source that samples from $W_{H\rho_{\text{in}}H^\dagger}$, using covariance. In the second step, it would, for each sample \mathbf{u} drawn from $W_{H\rho_{\text{in}}H^\dagger}$, output the value $(-1)^{[\mathbf{a}(Z), \mathbf{u}]}$, with $\mathbf{a}(Z)$ such that $T_{\mathbf{a}(Z)} = Z$; cf. Eq. (22). But there is a problem.

Lemma 5. For any number n of qubits, no Wigner function of the type defined in Eqs. (7) and (8) is covariant under a Hadamard gate on a single qubit.

The first step of the above procedure cannot be performed.

Proof of Lemma 5. We only discuss $n = 1$; the generalization to other n is straightforward. Consider the phase point operator $A_0 = (I + i^{\gamma_x} X + i^{\gamma_y} Y + i^{\gamma_z} Z)/2$. For W to be covariant under H , we require that $H^\dagger A_0 H = A_{\mathbf{u}}$, for some \mathbf{u} . Now consider the sum of signs $\eta = \gamma_x + \gamma_y + \gamma_z \pmod{4}$, and how it transforms under H . Since $H^\dagger X H = Z$, $H^\dagger Y H = -Y$,

and $H^\dagger ZH = X$, it follows that $\eta \longrightarrow \eta' = \eta + 2 \pmod{4}$. However, under the transformation $A_0 \longrightarrow A_{\mathbf{u}} = T_{\mathbf{u}}^\dagger A_0 T_{\mathbf{u}}$, the signs of an even number of $\{X, Y, Z\}$ are flipped, and hence η remains unchanged mod 4. Thus $H^\dagger A_0 H \neq A_{\mathbf{u}}$ for any \mathbf{u} , for any γ . ■

2. Preservation of positivity

As an example, consider a quantum circuit for two qubits consisting of a Hadamard gate H_1 on the first qubit followed by a measurement of the Pauli observable Z_1 . Assume the initial state ρ_{in} is the completely mixed state, for which each Wigner function of the type Eq. (8) is positive and can be efficiently sampled from. The task is to sample from the output distribution of the measurement.

Again, a classical simulation method based on the preservation of Wigner function positivity under free unitaries again runs into a problem.

Lemma 6. For $n \geq 2$, for no Wigner function of type Eq. (8) positivity is preserved for all states under a Hadamard gate on a single qubit.

Proof of Lemma 6. Consider a real stabilizer state ρ of two qubits, $\rho = (I + T_{\mathbf{a}} + T_{\mathbf{b}} + T_{\mathbf{a}}T_{\mathbf{b}})/4$, and w.l.o.g the Hadamard gate H_1 on the first qubit. As a consequence of Lemma 1, $W_\rho \geq 0$ if and only if $\beta(\mathbf{a}, \mathbf{b}) = 0$. Likewise, for the transformed state, $W_{H_1 \rho H_1^\dagger} \geq 0$ if and only if $\beta(H_1 \mathbf{a}, H_1 \mathbf{b}) = 0$. Now consider the Peres-Mermin square. There are six contexts, i.e., sets of commuting Pauli observables such that within each set the observables multiply to the identity times ± 1 . Whatever the phase convention γ for the nine Pauli operators in the square, there is always an odd number of contexts for which $\beta \pmod{4} = 2$. Namely, for the standard phase convention, there is one such context. If the sign of any of the Pauli observables is flipped, then $\beta \longrightarrow \beta + 2 \pmod{4}$ in the corresponding horizontal and vertical context. Hence the number of contexts with $\beta \pmod{4} = 2$ remains odd.

The action of H_1 subdivides the set of the six contexts into three orbits of size 2. Since the number of nonzero values of β is odd, there must be at least one orbit in which one β has the value 0 and the other has the value 2. Within this orbit, choose \mathbf{a}, \mathbf{b} such that $\beta(\mathbf{a}, \mathbf{b}) = 0$. Hence, $\beta(H_1 \mathbf{a}, H_1 \mathbf{b}) = 2$. Thus, positivity is not preserved under H_1 , for any phase convention γ . ■

On the other hand, the simulation method of Sec. IV A has no problem with the above example circuit. Namely, there are Wigner functions of the type Eq. (8) for which the Hadamard gate on a single (the first) qubit is in the group of free gates, $H_1 \in G$, and Z_1 is in the set of directly measurable observables, $Z_1 \in \mathcal{O}$. An example for such a Wigner function is given in Sec. VI.

We observe that the negativity which can be introduced into a Wigner function by the free unitary gates G is of a very special kind. Namely, it can be lifted by redefinition of the Wigner function according to $A_{\mathbf{v}} \mapsto A'_{\mathbf{v}} = g A_{\mathbf{v}} g^\dagger$, $\forall \mathbf{v} \in V$, for some $g \in G$.

To summarize, while in the present framework the free measurements are required to preserve positivity of the Wigner function, no such constraint needs to be imposed on the free unitaries. The amount of negativity introduced into the Wigner functions by the free unitaries can be large, as measured by

sum negativity [45]. However, it is always of a special kind. In this sense, our observation complements the recent finding [46] that a small amount of sum negativity—of any kind—does not compromise the efficiency of a suitable classical simulation algorithm.

V. CONTEXTUALITY

In this section, we establish contextuality of magic states as a computational resource, for qubit QCSI schemes satisfying Condition P1. We also clarify how contextuality and Wigner function negativity are related and state our most general classical simulation algorithm for qubit QCSI.

A. Noncontextual hidden variable models

Recall that \mathcal{O} is the set of Pauli observables which can be directly measured in QCSI, M is the set of observables which can have their value inferred by measurement of observables in \mathcal{O} , and any $C \subset M$ is a set of Pauli observables which can have their value inferred jointly, from a single copy of the given quantum state.

Definition 3. Consider a quantum state ρ and a set \mathcal{O} of observables grouping into contexts C of simultaneously measurable observables. A noncontextual hidden variable model $(\mathcal{S}, q_\rho, \Lambda)$ consists of a probability distribution q_ρ over a set \mathcal{S} of internal states and a set $\Lambda = \{\lambda_{\mathbf{v}}\}_{\mathbf{v} \in \mathcal{S}}$ of value assignment functions $\lambda_{\mathbf{v}} : \mathcal{O} \rightarrow \mathbb{R}$ that meet the following criteria:

(i) Each $\lambda_{\mathbf{v}} \in \Lambda$ is consistent with quantum mechanics: For any set C of jointly measurable observables there exists a quantum state $|\psi\rangle$ such that

$$A|\psi\rangle = \lambda_{\mathbf{v}}(A)|\psi\rangle, \quad \forall A \in C. \quad (24)$$

(ii) The distribution q_ρ satisfies

$$\text{tr}(A\rho) = \sum_{\mathbf{v} \in \mathcal{S}} \lambda_{\mathbf{v}}(A)q_\rho(\mathbf{v}), \quad \forall A \in M. \quad (25)$$

We say that a quantum state ρ is contextual if no noncontextual HVM according to Definition 3 exists that correctly reproduces the probability distributions $p_{C,\rho}(\mathbf{s}_C)$ of measurement outcomes for all sets C of jointly measurable observables.

The states $|\psi\rangle$ in Eq. (24) of Definition 3 are auxiliary. Their purpose is to ensure that the value assignments $\lambda_{\mathbf{v}}$ correspond to compatible eigenvalues. As a direct consequence of Eq. (24), the noncontextual value assignments $\lambda_{\mathbf{v}} \in \Lambda$ must all satisfy a set of compatibility constraints.

Lemma 7. For any triple $A, B, AB \in M$ of simultaneously measurable observables and any internal state $\mathbf{v} \in \mathcal{S}$ of an NCHVM $(\mathcal{S}, q_\rho, \Lambda)$ it holds that

$$\lambda_{\mathbf{v}}(AB) = \lambda_{\mathbf{v}}(A)\lambda_{\mathbf{v}}(B). \quad (26)$$

Returning to Definition 3, it may *a priori* seem that the condition (ii) is not sufficiently stringent and that one should rather require all outcome distributions for sets C of jointly measurable observables to be correctly reproduced by the

hidden variable model. That is,

$$p_{C,\rho}(\mathbf{s}_C) = \sum_{\nu \in \mathcal{S}} p(\mathbf{s}_C|\nu)q_\rho(\nu). \quad (27)$$

Therein, $p_{C,\rho}$ is the probability distribution for the measurement outcomes \mathbf{s}_C of the set C of simultaneously measurable observables given the quantum state ρ , and $p(\mathbf{s}_C|\nu)$ is the conditional probability for the measurement outcomes \mathbf{s}_C given the HVM internal state ν . However, Eq. (27) is implied by Eq. (25).

Lemma 8. An ncHVM according to Definition 3 that correctly reproduces all expectation values of observables $A \in M$ via Eq. (25) also correctly reproduces the outcome probability distributions Eq. (27) for all sets $C \subset M$ of jointly measurable observables.

Proof of Lemma 8. Assume the observables in C are algebraically independent, i.e., there are no nontrivial product relations among them, and denote by $\text{span}(C)$ the set of all products of observables in C .

Quantum mechanically, $p_{C,\rho}(\mathbf{s}_C) = \text{Tr}(E_C(\mathbf{s}_C)\rho)$, where the effect $E_C(\mathbf{s}_C)$ is

$$\begin{aligned} E_C(\mathbf{s}_C) &:= \prod_{A \in C} \frac{I + (-1)^{s(A)}A}{2} \\ &= \frac{1}{2^{|C|}} \sum_{B \in \text{span}(C)} (-1)^{s(B)} B. \end{aligned}$$

Therein, in the last line we have used that the measured eigenvalues $(-1)^{s(B)}$ satisfy the same consistency relation Eq. (26) as the noncontextual value assignments. Then,

$$\begin{aligned} p_{C,\rho}(\mathbf{s}_C) &= \frac{1}{2^{|C|}} \sum_{B \in \text{span}(C)} (-1)^{s(B)} \text{Tr}(B\rho) \\ &= \frac{1}{2^{|C|}} \sum_{B \in \text{span}(C)} (-1)^{s(B)} \sum_{\nu \in \mathcal{S}} \lambda_\nu(B) q_\rho(\nu) \\ &= \frac{1}{2^{|C|}} \sum_{\nu \in \mathcal{S}} q_\rho(\nu) \sum_{B \in \text{span}(C)} (-1)^{s(B)} \lambda_\nu(B) \\ &= \sum_{\nu \in \mathcal{S}} q_\rho(\nu) \delta_{(-1)^{s_C}, \lambda_\nu|_C} \\ &= \sum_{\nu \in \mathcal{S}} q_\rho(\nu) p(\mathbf{s}_C|\nu). \end{aligned}$$

In the last line above, we have used the fact that the value assignments λ are deterministic and that the conditional probabilities $p(\mathbf{s}_C|\nu)$ are thus δ functions. ■

B. The absence of state-independent contextuality

Consider the value assignment

$$\lambda(T_{\mathbf{a}}) = 1, \quad \forall \mathbf{a} \in V_M. \quad (28)$$

First, by Eq. (12), for any three commuting and directly measurable observables $T_{\mathbf{a}}, T_{\mathbf{b}}, T_{\mathbf{a}+\mathbf{b}} \in \mathcal{O}$ we have $T_{\mathbf{a}+\mathbf{b}} = +T_{\mathbf{a}}T_{\mathbf{b}}$. Thus, the above value assignment is compatible with all available direct measurements.

Second, the value of any observable $T_{\mathbf{a}+\mathbf{b}} \in M \setminus \mathcal{O}$ is inferred by measuring a suitable observable $T_{\mathbf{a}} \in \mathcal{O}$ for which

$[T_{\mathbf{a}+\mathbf{b}}, T_{\mathbf{a}}] = 0$, and then running a procedure to infer the value of $T_{\mathbf{b}}$. With Eq. (12), for all such observables $T_{\mathbf{a}}$ it holds that $T_{\mathbf{a}+\mathbf{b}} = +T_{\mathbf{a}}T_{\mathbf{b}}$, and the assignment Eq. (28) is thus consistent.

While the Peres-Mermin square and its cousins are present, the operational restriction enforced by condition Eq. (12) prevents obstructions to the assignment Eq. (28) from being established as experimental facts. Hence at least one consistent assignment exists, and there is no state-independent contextuality in this setting.

We are now in the position to prove Lemma 4 of Sec. III E.

Proof of Lemma 4. If there is a set C with $T_{\mathbf{a}}, T_{\mathbf{b}} \in C$ such then $T_{\mathbf{a}+\mathbf{b}} = -T_{\mathbf{a}}T_{\mathbf{b}}$ then the value assignment $\lambda(T_{\mathbf{a}}) = 1, \forall \mathbf{a} \in V$, is inconsistent. Contradiction. ■

Remark. In the companion paper [35], Condition P1 is replaced by the requirement that the set M of observables with inferable values is free of state-independent contextuality, given the set \mathcal{O} of directly measurable observables. With the above, we find that all such QCSI schemes are included in the present classification. The converse also holds: For every qubit QCSI scheme in which the set M of observables is free of state-independent contextuality, there is a Wigner function W^γ such that Condition P1 holds.

This can be seen as follows. If M is free of state-independent contextuality given the set \mathcal{O} of directly measurable observables, then there exists at least one consistent value assignment $\lambda : V_M \rightarrow \{\pm 1\}$. We may now rephrase the inferable observables, $T_{\mathbf{a}} \mapsto T'_{\mathbf{a}} = \lambda(\mathbf{a})^{-1}T_{\mathbf{a}}$, for all $\mathbf{a} \in V_M$, such that the new observables $\{T'_{\mathbf{a}}, \mathbf{a} \in V_M\}$ have a consistent value assignment $\lambda' \equiv 1$. This means that for all $T'_{\mathbf{a}} \in \mathcal{O}$, $T'_{\mathbf{b}}, T'_{\mathbf{a}+\mathbf{b}} \in M$ with $[\mathbf{a}, \mathbf{b}] = 0$ it holds that $T'_{\mathbf{a}+\mathbf{b}} = +T'_{\mathbf{a}}T'_{\mathbf{b}}$. With Eq. (10) it thus follows that $\beta'(\mathbf{a}, \mathbf{b}) = 0$ for all $\mathbf{a} \in V_{\mathcal{O}}, \mathbf{b} \in V_M$ with $[\mathbf{a}, \mathbf{b}] = 0$. Thus, with Lemma 1, the measurement of any observable $\pm T'_{\mathbf{a}}$ does not introduce negativity into the Wigner function defined by $A_0 = 1/2^n \sum_{\mathbf{a} \in V} T'_{\mathbf{a}}$.

C. Contextuality implies Wigner negativity

In accordance with existing results [15,22], also for the present setting of QCSI on qubits it holds that a non-negative Wigner function always implies the viability of a noncontextual hidden variable model.

Theorem 2. Consider a quantum state ρ with Wigner function W_ρ^γ given by Eq. (8). If $W_\rho^\gamma \geq 0$ then the measurement of all Pauli observables $T_{\mathbf{a}} \in \mathcal{O}$ can be described by a noncontextual hidden variable model.

Proof of Theorem 2. The Wigner function itself constitutes a noncontextual HVM, with set of internal states $\mathcal{S} = V$, probability distribution $q_\rho(\mathbf{u}) = W_\rho(\mathbf{u})$ over the internal states, and the conditional probabilities $p(\mathbf{s}_C|\mathbf{u})$ given by the Wigner functions of the effects.

Using Eq. (9), the probability $p_C(\mathbf{s}_C)$ for measuring the context C and obtaining the set \mathbf{s}_C of measurement outcomes is

$$p_C(\mathbf{s}_C) = \text{Tr}(\rho E_C(\mathbf{s}_C)) = \sum_{\mathbf{u} \in V} W_\rho(\mathbf{u}) [2^n W_{E_C(\mathbf{s}_C)}(\mathbf{u})]. \quad (29)$$

If $W_\rho \geq 0$, then W_ρ may be regarded as a probability distribution over the space V of internal states of a hidden variable model. If furthermore $0 \leq 2^n W_{E_C(\mathbf{s}_C)}(\mathbf{u}) \leq 1$ for all $\mathbf{u} \in V$, then we may regard $2^n W_{E_C(\mathbf{s}_C)} =: p(\mathbf{s}_C|\mathbf{u})$ as the

conditional probability for obtaining the outcome \mathbf{s}_C in the measurement of the observables in C , given the HVM internal state $\mathbf{u} \in V$. Then, Eq. (29) is Bayes's rule for computing the probability $p_C(\mathbf{s}_C)$.

It remains to check that the conditional probabilities $p(\mathbf{s}_C|\mathbf{u})$ assigned by the Wigner functions of the effects $E_C(\mathbf{s}_C)$ are compatible with Definition 3 of an ncHVM.

Item (i) of Definition 3. Consider contexts with a single observable $T_{\mathbf{a}}$. A definite value assignment for all internal states $\mathbf{u} \in V$ has already been established in Eq. (22). It corresponds to

$$p_{\mathbf{a}}(s|\nu) = 2^n W_{E_{\mathbf{a}}(s)}(\mathbf{u}) = \delta_{s, [\mathbf{a}, \mathbf{u}]}.$$
 (30)

The conditional probabilities $p_{\mathbf{a}}(s|\nu)$ thus have values in the required range $0 \dots 1$.

Consistency of the value assignments. Equation (22) yields $\lambda_{\mathbf{u}}(T_{\mathbf{a}+\mathbf{b}}) = \lambda_{\mathbf{u}}(T_{\mathbf{a}})\lambda_{\mathbf{u}}(T_{\mathbf{b}})$, for all $\mathbf{a}, \mathbf{b} \in V$. By Lemma 4, if $\{T_{\mathbf{a}}, T_{\mathbf{b}}\}$ form a jointly measurable set, then $T_{\mathbf{a}+\mathbf{b}} = T_{\mathbf{a}}T_{\mathbf{b}}$. Hence, $\lambda_{\mathbf{u}}(T_{\mathbf{a}}T_{\mathbf{b}}) = \lambda_{\mathbf{u}}(T_{\mathbf{a}})\lambda_{\mathbf{u}}(T_{\mathbf{b}})$ for such pairs $\mathbf{a}, \mathbf{b} \in V$, as required by Lemma 7. Therefore, for all contexts C , the observables $T_{\mathbf{a}} \in C$ have a joint eigenstate $|\psi_0\rangle$ with $T_{\mathbf{a}}|\psi_0(C)\rangle = +|\psi_0(C)\rangle$, and thus, the state $|\psi_{\mathbf{u}}(C)\rangle := T_{\mathbf{u}}|\psi_0(C)\rangle$ has eigenvalues $(-1)^{[\mathbf{a}, \mathbf{u}]}$ for the observables $T_{\mathbf{a}} \in C$. The condition Eq. (24) is thus satisfied for the value assignment Eq. (22).

Item (ii) of Definition 3. Since all observables $T_{\mathbf{a}} \in M$ only have eigenvalues ± 1 ,

$$\begin{aligned} \text{Tr}(T_{\mathbf{a}}\rho) &= \text{Tr}[(E_{\mathbf{a}}(0) - E_{\mathbf{a}}(1)]\rho) \\ &= 2^n \sum_{\mathbf{u} \in V} W_{\rho}(\mathbf{u})(W_{E_{\mathbf{a}}(0)} - W_{E_{\mathbf{a}}(1)}) \\ &= \sum_{\mathbf{u} \in V} W_{\rho}(\mathbf{u})(\delta_{0, [\mathbf{a}, \mathbf{u}]} - \delta_{1, [\mathbf{a}, \mathbf{u}]}) \\ &= \sum_{\mathbf{u} \in V} W_{\rho}(\mathbf{u})(-1)^{[\mathbf{a}, \mathbf{u}]} \\ &= \sum_{\mathbf{u} \in V} q_{\rho}(\mathbf{u})\lambda_{\mathbf{u}}(\mathbf{a}). \end{aligned}$$

Above, the second line follows by Eq. (9), the third by Eq. (30), and the fifth by Eq. (22). ■

The converse of Theorem 2 does not hold: There are quantum states with a noncontextual HVM description for which all considered Wigner functions are negative. This is illustrated in Fig. 4 for the example of a single qubit, where all physically allowed states have an HVM description [1]. The one-qubit states are all of the form $\rho = (I + \mathbf{r}\vec{\sigma})/2$, and the physical such states are constrained by $|\mathbf{r}| \leq 1$. The set of states describable in terms of a noncontextual HVM is a cube, $|r_x|, |r_y|, |r_z| \leq 1$, containing all physical states. The eight extremal states i of this cube have definite value assignments $\lambda_i(X), \lambda_i(Y), \lambda_i(Z) = \pm 1$ for the observables X, Y, Z .

Up to equivalence under translation, there are two one-qubit Wigner functions of type Eq. (8), namely the Wigner function W defined by the phase point operator at the origin $A_0 = (I + X + Z + Y)/2$, and the Wigner function \bar{W} defined by $\bar{A}_0 = HA_0H^\dagger = (I + X + Z - Y)/2$. The phase space for these Wigner functions is $\mathbb{Z}_2 \times \mathbb{Z}_2$, and W, \bar{W} thus have four extremal states each. If these extremal states are combined, the extremal states of the noncontextual HVM are recovered.

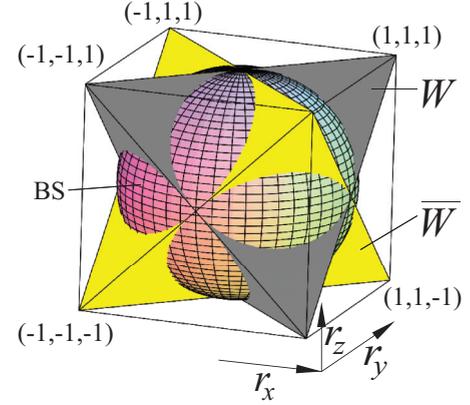


FIG. 4. State space for the one-qubit states $\rho = (I + \mathbf{r}\vec{\sigma})/2$. The physical states lie within or on the Bloch sphere (BS). The two tetrahedra contain the states positively represented by the Wigner functions W and \bar{W} , respectively. The state space describable by a noncontextual HVM is a cube with corners $(\pm 1, \pm 1, \pm 1)$; also see Ref. [47]. It contains the Bloch ball.

Each Wigner function by itself has only half of the extremal states of the HVM, and the set of positively represented states is thus smaller. Furthermore, there are physical states which are negatively represented by both W and \bar{W} ; see Fig. 4. Contextuality and negativity of the Wigner functions Eq. (8) are thus not the same.

Remark. While there are physical one-qubit quantum states which are negatively represented by both W and \bar{W} , every state $\rho = (I + \mathbf{r}\vec{\sigma})/2$, with $|r_x|, |r_y|, |r_z| \leq 1$, can be described by an ensemble

$$\mathcal{E}(\rho) = \{(p_1, \rho_1), (p_2, \rho_2)\},$$

such that $W_{\rho_1} \geq 0$ and $\bar{W}_{\rho_2} \geq 0$. A generalization of this fact to n -qubit systems will be of relevance in Sec. V D.

D. Contextuality as a resource

Theorem 3. For any QCSI scheme (γ, \mathcal{M}) , if the input magic state ρ_{in} can be described by a noncontextual HVM, then the quantum state $\rho_t(\mathbf{s}_{<t})$ at time t , conditioned on the prior measurement record $\mathbf{s}_{<t}$, can be described by a noncontextual HVM, for any t and any \mathbf{s} .

Before we give the proof of Theorem 3, we need to set up some more notation. We observe that the state space of a general noncontextual HVM is larger than the state space of an HVM deriving from a non-negative Wigner function; see the discussion of a single qubit in Sec. V C, Fig. 4. The enlarged state space $\mathcal{S} = \{\nu\}$ is finite, yet maximal in the sense that for every value assignment $\lambda(\cdot)$ satisfying the consistency conditions Eq. (26), there is a corresponding internal state $\nu \in \mathcal{S}$ such that $\lambda_{\nu}(\cdot) \equiv \lambda(\cdot)$.

We choose to have this state space \mathcal{S} acted upon by the group V , dividing \mathcal{S} into orbits. Namely, given an element $\nu \in \mathcal{S}$ specified by the value assignment $\lambda_{\nu} : V \rightarrow \{\pm 1\}$, there is another internal state $\nu + \mathbf{u}$, defined through the value assignment

$$\lambda_{\nu+\mathbf{u}}(\mathbf{a}) = \lambda_{\nu}(\mathbf{a})(-1)^{[\mathbf{u}, \mathbf{a}]}, \quad \forall \mathbf{a} \in V, \quad (31)$$

for all $\mathbf{u} \in V$. In Eq. (31), we have set $\lambda_v(\mathbf{a}) := \lambda_v(T_{\mathbf{a}})$ for notational simplicity. It is easily seen that $v \in \mathcal{S} \Leftrightarrow v + \mathbf{u} \in \mathcal{S}$, for all $\mathbf{u} \in V$. The condition to check is the consistency of the value assignment in item (ii) of Definition 3. Equation (26) is preserved under the change $\lambda_v(\mathbf{a}) \mapsto \lambda_v(\mathbf{a})(-1)^{[\mathbf{u}, \mathbf{a}]}$, for any $\mathbf{u} \in V$. The group action of V on \mathcal{S} defined through Eq. (31) labels the elements of \mathcal{S} in a fashion convenient for the subsequent discussion.

Proof of Theorem 3. The proof of Theorem 3 is by induction. We assume that there exists an HVM with probability distribution $q_{t, \mathbf{s}_{<t}}$ which describes the quantum state $\rho_t(\mathbf{s}_{<t})$, conditioned on the previous measurement record $\mathbf{s}_{<t}$. We then show that there is an HVM with probability distribution $q_{t+1, \mathbf{s}_{<t+1}}$ which describes the quantum state $\rho_{t+1}(\mathbf{s}_{<t+1})$.

To establish this result, we need the relation between $q_{t+1, \mathbf{s}_{<t+1}}$ and its precursor $q_{t, \mathbf{s}_{<t}}$. Denoting the observable measured in the t th time step of the computation by $T_{\mathbf{a}_t} \in \mathcal{O}$ and the corresponding measurement outcome by $s_t \in \mathbb{Z}_2$, the required relation is

$$q_{t+1, \mathbf{s}_{<t+1}}(v) = \frac{\delta_{(-1)^{s_t}, \lambda_v(\mathbf{a}_t)} q_{t, \mathbf{s}_{<t}}(v) + q_{t, \mathbf{s}_{<t}}(v + \mathbf{a}_t)}{\bar{p}_t(s_t | \mathbf{s}_{<t}) 2}, \quad (32a)$$

$$\bar{p}_t(s_t | \mathbf{s}_{<t}) = \sum_{v \in \mathcal{S}} \delta_{(-1)^{s_t}, \lambda_v(\mathbf{a}_t)} q_{t, \mathbf{s}_{<t}}(v). \quad (32b)$$

Therein, $\bar{p}_t(s_t | \mathbf{s}_{<t})$ is the HVM prediction for the probability of obtaining the outcome s_t in the measurement of $T_{\mathbf{a}_t}$, given a prior measurement record $\mathbf{s}_{<t}$. Equation (32) will be justified *a posteriori*. Namely, with these assignments, the induction argument works out.

With Eq. (25) in Definition 3, the induction assumption is

$$\langle T_{\mathbf{a}} \rangle_{\rho_t} = \langle T_{\mathbf{a}} \rangle_{q_t}, \quad \forall \mathbf{a} \in V.$$

Therein, we have suppressed the dependence on the measurement record, to simplify the notation. We need to show that

$$\langle T_{\mathbf{a}} \rangle_{\rho_{t+1}} = \langle T_{\mathbf{a}} \rangle_{q_{t+1}}, \quad \forall \mathbf{a} \in V,$$

and that $\bar{p}_t(s_t | \mathbf{s}_{<t}) = p_t(s_t | \mathbf{s}_{<t})$, with $p_t(s_t | \mathbf{s}_{<t})$ the quantum mechanical value for the probability of the outcome s_t given the prior measurement record $\mathbf{s}_{<t}$.

First, regarding the probability of finding s_t ,

$$\begin{aligned} \bar{p}_t(s_t | \mathbf{s}_{<t}) &= \sum_{v \in \mathcal{S}} \frac{1 + (-1)^{s_t} \lambda_v(\mathbf{a}_t)}{2} q_{t, \mathbf{s}_{<t}}(v) \\ &= \frac{1}{2} \sum_{v \in \mathcal{S}} q_{t, \mathbf{s}_{<t}}(v) + \frac{(-1)^{s_t}}{2} \sum_{v \in \mathcal{S}} q_{t, \mathbf{s}_{<t}}(v) \lambda_v(\mathbf{a}_t) \\ &= \frac{\langle I \rangle_{\rho_t(\mathbf{s}_{<t})} + (-1)^{s_t} \langle T_{\mathbf{a}_t} \rangle_{\rho_t(\mathbf{s}_{<t})}}{2} \\ &= \text{Tr} \left(\rho_t(\mathbf{s}_{<t}) \frac{I + (-1)^{s_t} T_{\mathbf{a}_t}}{2} \right) \\ &= p_t(s_t | \mathbf{s}_{<t}). \end{aligned}$$

We thus reproduce the quantum mechanical expression within the HVM. Above, in transitioning from the second to the third line, we have invoked the induction assumption.

Second, regarding the expectation values of the $T_{\mathbf{a}}$ on $\rho_{t+1}(\mathbf{s}_{<t+1})$, the HVM prediction is

$$\begin{aligned} \langle T_{\mathbf{a}} \rangle_{q_{t+1}} &= \sum_{v \in \mathcal{S}} q_{t+1, \mathbf{s}_{<t+1}}(v) \lambda_v(\mathbf{a}) \\ &= \sum_{v \in \mathcal{S}} \frac{1 + (-1)^{s_t} \lambda_v(\mathbf{a}_t)}{4 p_t(s_t | \mathbf{s}_{<t})} q_{t, \mathbf{s}_{<t}}(v) \lambda_v(\mathbf{a}) \\ &\quad + \sum_{v \in \mathcal{S}} \frac{1 + (-1)^{s_t} \lambda_v(\mathbf{a}_t)}{4 p_t(s_t | \mathbf{s}_{<t})} q_{t, \mathbf{s}_{<t}}(v + \mathbf{a}_t) \lambda_v(\mathbf{a}). \end{aligned}$$

Reordering the sum via the substitution $v + \mathbf{a}_t \rightarrow v$, and using Eq. (31), the second term in the last line equals

$$\sum_{v \in \mathcal{S}} \frac{1 + (-1)^{s_t} \lambda_v(\mathbf{a}_t)}{4 p_t(s_t | \mathbf{s}_{<t})} q_{t, \mathbf{s}_{<t}}(v) \lambda_v(\mathbf{a}) (-1)^{[\mathbf{a}, \mathbf{a}_t]}.$$

We now distinguish between the case where $T_{\mathbf{a}}, T_{\mathbf{a}_t}$ commute and where they do not.

Case (i). $[\mathbf{a}, \mathbf{a}_t] = 1$. Then, $\langle T_{\mathbf{a}} \rangle_{q_{t+1}} = 0$, which is the correct quantum mechanical expression.

Case (ii). $[\mathbf{a}, \mathbf{a}_t] = 0$. Then, the expression for $\langle T_{\mathbf{a}} \rangle_{q_{t+1}}$ simplifies to

$$\begin{aligned} \langle T_{\mathbf{a}} \rangle_{q_{t+1}} &= \sum_{v \in \mathcal{S}} \frac{1 + (-1)^{s_t} \lambda_v(\mathbf{a}_t)}{2 p_t(s_t | \mathbf{s}_{<t})} q_{t, \mathbf{s}_{<t}}(v) \lambda_v(\mathbf{a}) \\ &= \frac{1}{2 p_t(s_t | \mathbf{s}_{<t})} \sum_{v \in \mathcal{S}} q_{t, \mathbf{s}_{<t}}(v) \lambda_v(\mathbf{a}) \\ &\quad + \frac{(-1)^{s_t}}{2 p_t(s_t | \mathbf{s}_{<t})} \sum_{v \in \mathcal{S}} q_{t, \mathbf{s}_{<t}}(v) \lambda_v(\mathbf{a} + \mathbf{a}_t). \end{aligned}$$

Here we have used the relation $\lambda_v(\mathbf{a} + \mathbf{a}_t) = \lambda_v(\mathbf{a}_t) \lambda_v(\mathbf{a})$, which arises as follows. Since $T_{\mathbf{a}_t} \in \mathcal{O}$, $T_{\mathbf{a}} \in \mathcal{M}$, and $[T_{\mathbf{a}_t}, T_{\mathbf{a}}] = 0$ by the case assumption, $\{T_{\mathbf{a}_t}, T_{\mathbf{a}}\}$ is a jointly measurable set of observables; cf. example (iii) after Def. 2. (The procedure is to measure $T_{\mathbf{a}_t} \in \mathcal{O}$ first, and then run the measurement sequence for $T_{\mathbf{a}} \in \mathcal{M}$.) Thus, by property (ii) of Definition 3 for noncontextual HVMs, $\lambda_v(T_{\mathbf{a}_t}, T_{\mathbf{a}}) = \lambda_v(T_{\mathbf{a}_t}) \lambda_v(T_{\mathbf{a}})$. Finally, with Lemma 1, $T_{\mathbf{a}_t}, T_{\mathbf{a}} = T_{\mathbf{a} + \mathbf{a}_t}$, which yields the stated relation.

Next we use the induction assumption and obtain

$$\begin{aligned} \langle T_{\mathbf{a}} \rangle_{q_{t+1}} &= \frac{1}{2 p_t(s_t | \mathbf{s}_{<t})} (\langle T_{\mathbf{a}} \rangle_{\rho_t} (-1)^{s_t} + \langle T_{\mathbf{a} + \mathbf{a}_t} \rangle_{\rho_t}) \\ &= \frac{1}{p_t(s_t | \mathbf{s}_{<t})} \text{Tr} \left(\rho_t \frac{I + (-1)^{s_t} T_{\mathbf{a}_t}}{2} T_{\mathbf{a}} \right) \\ &= \frac{\text{Tr} \left(\left[\frac{I + (-1)^{s_t} T_{\mathbf{a}_t}}{2} \rho_t \frac{I + (-1)^{s_t} T_{\mathbf{a}_t}}{2} \right] T_{\mathbf{a}} \right)}{p_t(s_t | \mathbf{s}_{<t})} \\ &= \langle T_{\mathbf{a}} \rangle_{\rho_{t+1}}. \end{aligned}$$

We thus reproduce the quantum mechanical expression within the HVM. This completes the induction step.

The induction starts at time $t = 1$, where $\rho_1 = \rho_{\text{in}}$ has an HVM description, by assumption of Theorem 3. Thus, by induction, for every time $t \geq 1$ and every history $\mathbf{s}_{<t}$ of measurement outcomes, the conditional state $\rho_t(\mathbf{s}_{<t})$ has a description in terms of a noncontextual HVM. ■

Corollary 1. For any QCSI scheme (γ, \mathcal{M}) , if the input magic state ρ_{in} can be described by a noncontextual HVM,

then for the measurement of any sequence of observables $\{T_{\mathbf{a}_t}, t = 1..t_{\max}\} \subset \mathcal{O}$, the probability distribution $p(\mathbf{s}) = p(s_1, s_2, \dots, s_{t_{\max}})$ of outcomes is fixed by the HVM for ρ_{in} . The $T_{\mathbf{a}_t}$ may be mutually noncommuting and dependent on previous measurement outcomes.

Proof of Corollary 1. By Bayes' rule, the joint probability of the outcomes \mathbf{s} can be written as

$$p(\mathbf{s}) = \prod_{t=1}^{t_{\max}} p_t(s_t | \mathbf{s}_{<t}).$$

By Theorem 3, the conditional probabilities $p_t(s_t | \mathbf{s}_{<t}) = \bar{p}_t(s_t | \mathbf{s}_{<t})$ are all correctly obtained from the probability distributions $q_{t, \mathbf{s}_{<t}}$; cf. Eq. (32b). The distributions $q_{t, \mathbf{s}_{<t}}$, for $t = 2, \dots, t_{\max}$, in turn follow from the distribution $q_{1, \mathbf{s}_{<1} = \emptyset}$ (describing ρ_{in} at $t = 1$), by Eq. (32a). Thus, $p(\mathbf{s})$ is fully specified by the probability distribution $q_{1, \mathbf{s}_{<1} = \emptyset}$ over the state space \mathcal{S} of the HVM. ■

We now discuss the implications of Theorem 3 with regards to universality of quantum computation. We want to capture in our analysis the case where a QCSI scheme running on n qubits is universal only on a subspace supporting k encoded qubits. (This does of course include the unencoded case, where every logical qubit is represented by one physical qubit.) We use the following notion of computational universality.

Definition 4. We say that a QCSI scheme is encoded universal if the following operations can be performed.

U1 Encoded inputs. Prepare a set of encoded orthonormal input states $\mathcal{E}(|x\rangle)$, $x \in \{0, 1\}^k$ up to an arbitrarily small error ϵ , where $\mathcal{E} : \mathbb{C}^{2^k} \rightarrow \mathbb{C}^{2^n}$ is an isometry of k logical qubits into n physical qubits.

U2 Encoded gates. For any $V \in \text{SU}(\mathbb{C}^{2^k})$ and any encoded input state $\mathcal{E}(|\phi\rangle)$ prepare the encoded output state $\mathcal{E}(V|\phi\rangle)$, up to an arbitrarily small error ϵ .

U3 Encoded outputs. Measure the value of any logical observable $\mathcal{E}(X_i)$, i.e., $\{\mathcal{E}(X_i), i = 1, \dots, k\} \subset \mathcal{O}$.

Requirement U3 means that it is possible to physically measure any logical qubit in the standard basis.

We then have the following result:

Theorem 4. A QCSI scheme (γ, \mathcal{M}) on $k \geq 3$ (possibly encoded) qubits satisfying U1–U3 is universal only if its magic states are contextual.

The full proof of Theorem 4 is given in Appendix. Here we prove Theorem 4 under the simplifying assumption that every encoded qubit can be measured in two complementary bases rather than one basis. That is, U3 is replaced by

$$\text{U3}' \{ \mathcal{E}(X_i), \mathcal{E}(Y_i), i = 1, \dots, k \} \subset \mathcal{O}.$$

While more stringent than U3, the condition U3' is not unreasonable. It grants the measurement device the power to measure two complementary observables for each encoded qubit, and thus to be genuinely quantum. However, the main reason for invoking U3' is that it removes a substantial amount of technical complication from the proof, while preserving its general structure.

Proof of Theorem 4 under U3'. We consider a QCSI where the available initial (magic) states all have an ncHVM description.

Now assume that the QCSI scheme is universal for quantum computation. Then, it must be possible to create an encoded

Greenberger-Horne-Zeilinger (GHZ) state $\mathcal{E}(|\text{GHZ}\rangle)$, with $|\text{GHZ}\rangle = (|000\rangle + |111\rangle)/\sqrt{2}$, on a subset of the qubits from an initial state $\mathcal{E}(|0\rangle^{\otimes n})$. Now consider the expectation value

$$\mathcal{W} = \langle \mathcal{E}(X_1 X_2 X_3) - \mathcal{E}(X_1 Y_2 Y_3) - \mathcal{E}(Y_1 X_2 Y_3) - \mathcal{E}(Y_1 Y_2 X_3) \rangle.$$

\mathcal{W} is a contextuality witness derived from Mermin's star [3]. Since the observables $\mathcal{E}(X_i)$, $i = 1, \dots, 3$, are directly measurable by assumption U3', their product $\mathcal{E}(X_1 X_2 X_3)$ is inferable, and for any internal state ν of the ncHVM it holds that

$$\lambda_\nu(\mathcal{E}(X_1 X_2 X_3)) = \prod_{i=1}^3 \lambda_\nu(\mathcal{E}(X_i)).$$

The same holds for the other three measurement contexts $(\mathcal{E}(X_1), \mathcal{E}(Y_2), \mathcal{E}(Y_3))$, $(\mathcal{E}(Y_1), \mathcal{E}(X_2), \mathcal{E}(Y_3))$, and $(\mathcal{E}(Y_1), \mathcal{E}(Y_2), \mathcal{E}(X_3))$. Since for all ncHVM states ν , $\lambda_\nu(\mathcal{E}(X_i)), \lambda_\nu(\mathcal{E}(Y_i)) = \pm 1$, for all states ρ describable by an ncHVM it holds that $\mathcal{W}_\rho \leq 2$, which is the Mermin inequality [7]. $\mathcal{W}_{|\text{GHZ}\rangle} = 4$, and the encoded state $\mathcal{E}(|\text{GHZ}\rangle)$ is thus contextual. With Theorem 3, it cannot be prepared by the given QCSI with a nonzero probability of success: contradiction.

The indirect assumption is thus wrong. Hence, if the initial (magic) states are noncontextual, the resulting QCSI scheme is not universal. ■

Remark. The same conclusion holds when an error ϵ is allowed in the quantum computation, due to the finite gap between of 2 between $\mathcal{W}_{|\text{GHZ}\rangle} = 4$ and $\mathcal{W}_{\rho_{\text{HVM}}} \leq 2$.

E. Generalized simulation algorithm

Theorem 5. For any QCSI scheme (γ, \mathcal{M}) , if (i) the input magic state ρ_{in} can be described by a noncontextual HVM with state space \mathcal{S} and value assignments $\lambda_\nu : V \rightarrow \{\pm 1\}$, for all $\nu \in \mathcal{S}$, (ii) this HVM can be efficiently sampled from, and (iii) the value assignments $\lambda_\nu(\mathbf{a})$ and the phase convention $\gamma(\mathbf{a})$ can be efficiently evaluated for all $\mathbf{a} \in V_{\mathcal{O}}$, then any resulting QCSI can be efficiently classically simulated.

Theorem 5 is proved constructively, i.e., by providing a classical simulation algorithm. This algorithm is given in Table II. It is an almost exact copy of the simulation algorithm

TABLE II. Algorithm 2 for the classical simulation of n -qubit QCSI with an ncHVM for the initial state. Addition on the HVM state labels ν in Eq. (33) is defined through Eq. (31).

Algorithm 2

- (1) Draw a sample $\nu \in \mathcal{S}$ from the probability distribution $q_{1, \mathbf{s}_{<1} = \emptyset}$ describing ρ_{in} in the HVM, and set $\nu_1 := \nu$.
- (2) For all the measurements of observables $T_{\mathbf{a}_t} \in \mathcal{O}$ comprising the circuit, starting with the first,
 - (a) output the measurement outcome $\lambda_{\nu_t}(\mathbf{a}_t) \in \{\pm 1\}$ for the observable $T_{\mathbf{a}_t}$,
 - (b) flip a fair coin, and update the sample

$$\nu_t \rightarrow \nu_{t+1} = \begin{cases} \nu_t, & \text{if "heads"} \\ \nu_t + \mathbf{a}_t, & \text{if "tails"} \end{cases}, \quad (33)$$

until the measurement sequence is complete.

- (3) Repeat until sufficient statistics is gathered.
-

encountered in Sec. IV B, and we comment on the resemblance in Sec. V F.

Before we proceed to the proof of Theorem 5, we briefly discuss what sampling from conditional probability distributions means for the above algorithm. For any sample v drawn in step 1, while looping through step 2, a measurement record \mathbf{s} is built up. In every iteration t of step 2, the updated sample v_t may be regarded as being drawn from a probability distribution $\tilde{q}_{t,\mathbf{s}_{<t}}$, conditioned on the previous measurement record $\mathbf{s}_{<t}$. So the above simulation algorithm definitely samples. The question is whether it samples from the *correct* distributions, i.e., whether $\tilde{q}_{t,\mathbf{s}_{<t}} = q_{t,\mathbf{s}_{<t}}$, for all $t = 1, \dots, t_{\max}$ and for all \mathbf{s} .

Proof of Theorem 5. The proof proceeds by demonstrating the correctness and efficiency of the above classical simulation algorithm.

Correctness. We first show that for each time t and measurement record $\mathbf{s}_{<t}$, the above classical simulation algorithm (i) produces the correct quantum mechanical conditional probability $p_t(s_t|\mathbf{s}_{<t})$ of obtaining the outcome s_t in the measurement of the observable $T_{\mathbf{a}_t} \in \mathcal{O}$, and (ii) samples from the correct conditional probability distribution $q_{t,\mathbf{s}_{<t}}$ of the HVM, which is given by Eq. (32).

The proof is by induction. We assume that at time t , the classical simulation algorithm samples from the correct distribution $q_{t,\mathbf{s}_{<t}}$.

Regarding (i): Denote the conditional probabilities produced by the simulation algorithm as $\tilde{p}_t(s_t|\mathbf{s}_{<t})$. A state $v \in \mathcal{S}$ contributes its probability weight $q_{t,\mathbf{s}_{<t}}(v)$ to $\tilde{p}_t(0|\mathbf{s}_{<t})$ or $\tilde{p}_t(1|\mathbf{s}_{<t})$ if $\lambda_v(T_{\mathbf{a}_t}) = +1$ or $\lambda_v(T_{\mathbf{a}_t}) = -1$, respectively. Therefore,

$$\tilde{p}_t(s_t|\mathbf{s}_{<t}) = \sum_{v \in \mathcal{S}} \delta_{\lambda_v(T_{\mathbf{a}_t}), (-1)^{s_t}} q_{t,\mathbf{s}_{<t}}(v) = \bar{p}_t(s_t|\mathbf{s}_{<t}).$$

The second equality follows by comparison with Eq. (32b). Furthermore, $\bar{p}_t(s_t|\mathbf{s}_{<t}) = p_t(s_t|\mathbf{s}_{<t})$ was already demonstrated in the proof of Theorem 3. Thus, $\tilde{p}_t(s_t|\mathbf{s}_{<t}) = p_t(s_t|\mathbf{s}_{<t})$, as required.

Re (ii): Through the value assignment in step 2(a), an internal state $v_t \in \mathcal{S}$ contributes to

$$\begin{aligned} \tilde{q}_{t+1,(\mathbf{s}_{<t}, s_t=0)}, & \quad \text{if } \lambda_{v_t}(\mathbf{a}_t) = +1, \\ \tilde{q}_{t+1,(\mathbf{s}_{<t}, s_t=1)}, & \quad \text{if } \lambda_{v_t}(\mathbf{a}_t) = -1. \end{aligned}$$

The updated rule for step 2(a) is thus

$$q_{t,\mathbf{s}_{<t}}(\tau) \xrightarrow{s_t} q'_{t+1,\mathbf{s}_{<t+1}}(\tau) = q_{t,\mathbf{s}_{<t}}(\tau) \frac{\delta_{\lambda_{\tau}(\mathbf{a}_t), (-1)^{s_t}}}{p_t(s_t|\mathbf{s}_{<t})},$$

for all $\tau \in \mathcal{S}$, and $p_t(s_t|\mathbf{s}_{<t})$ appears for normalization.

In step 2(b), with Eq. (33), the updated rule is

$$q'_{t+1,\mathbf{s}_{<t+1}} \longrightarrow \tilde{q}_{t+1,\mathbf{s}_{<t+1}} = q'_{t+1,\mathbf{s}_{<t+1}} * \frac{\delta_{\mathbf{0}} + \delta_{\mathbf{a}_t}}{2},$$

where the asterisk (*) stands for convolution. Using Eq. (31), the resulting expression for $\tilde{q}_{t+1,\mathbf{s}_{<t+1}}(v)$ matches the expression in Eq. (32a), i.e., $\tilde{q}_{t+1,\mathbf{s}_{<t+1}}(v) = q_{t+1,\mathbf{s}_{<t+1}}(v)$, as required. This completes the induction step.

The induction assumption is satisfied for $t = 1$, by the first assumption of Theorem 5. Thus, by induction, the above algorithm samples from the correct conditional outcome probabilities $p(s_t|\mathbf{s}_{<t})$ for measurement outcomes s_t and from

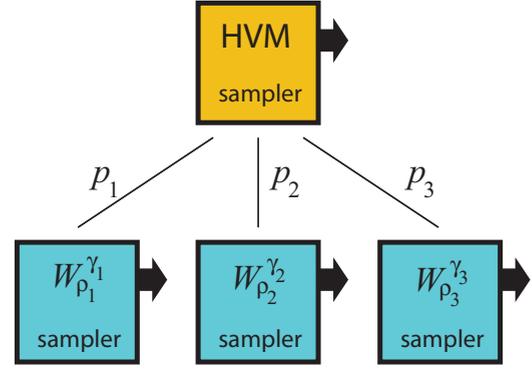


FIG. 5. Relation between Algorithms 1 and 2. Sampling from the probability distribution underlying a noncontextual HVM may be viewed as a two-stage process. Stage 1: Sampling from a probability distribution $\{p_i\}$ over Wigner functions, Stage 2: Sampling from the phase space with respect to the Wigner function chosen in the first stage.

the correct HVM distributions $q_{t,\mathbf{s}_{<t}}$ describing $\rho_t(\mathbf{s}_{<t})$, for all times t and all outcome histories \mathbf{s} .

Efficiency. The classical preprocessing of removing the unitaries $g \in G$ from the circuit is efficient if the function $\gamma : V \rightarrow \mathbb{Z}_4$ can be efficiently computed, which holds by assumption. See the proof of Theorem 1.

Regarding the simulation algorithm itself, the critical step is 2(a), the evaluation of the function λ_v on some $\mathbf{a} \in V$. Again, the efficiency of this function evaluation is guaranteed by the assumption of the theorem. ■

F. Relation between Algorithms 1 and 2

Algorithms 1 and 2 are very similar. They only differ in the sampling source they have access to. In this section, we explain that Algorithm 2 can be understood as a master algorithm calling Algorithm 1 as a subroutine; see Fig. 5. This illustrates that an ncHVM for a set of magic states can be viewed as a probabilistic mixture of non-negative Wigner functions.

By Theorem 2, the sampling source for Algorithm 2, based on noncontextual HVMs, is at least as powerful as the sampling source for Algorithm 1, based on non-negative Wigner functions. By the 1-qubit example of Sec. V C, the former source is indeed more powerful. The root of the connection between the two algorithms is that if the initial quantum state ρ_{in} can be described by a noncontextual HVM, then it can be represented by an ensemble

$$\mathcal{E}_{\rho_{\text{in}}} = \{(p_i, \rho_i)\},$$

such that there are Wigner functions W^{γ_i} for which (i) $W^{\gamma_i}_{\rho_i} \geq 0, \forall i$, and (ii) the measurement of observables from the set \mathcal{O} preserves positivity of the $W^{\gamma_i}, \forall i$.

Therefore, Algorithm 2 can be simulated by a master algorithm that merely draws samples $v \in \mathcal{S}$ from the noncontextual HVM, and then employs Algorithm 1 as a subroutine for dealing with the samples. This interpretation of Algorithm 2 is developed below.

The state space \mathcal{S} of the HVM can be partitioned into orbits $[v]$ of V ,

$$[v] = \{v + \mathbf{u}, \mathbf{u} \in V\} \in \mathcal{S}/V.$$

Then there exists a special orbit $[0] \in \mathcal{S}/V$ defined by the property that there is a $\tau_{[0]} \in [0]$ for which the value assignment is constant, $\lambda_{\tau_{[0]}}(\cdot) \equiv 1$. With Eq. (31) it then follows that

$$\lambda_{\tau_{[0]}+\mathbf{u}}(\mathbf{a}) = (-1)^{[\mathbf{a}, \mathbf{u}]}, \forall \mathbf{a} \in V.$$

Comparing with Eq. (22), we find that the above value assignment $\lambda_{\tau_{[0]}+\mathbf{u}}(\mathbf{a})$ agrees with the value assignment made by a positive Wigner function Eq. (8) considered as an HVM, if we identify, for all $\mathbf{u} \in V$,

$$(\tau_{[0]} + \mathbf{u}) \in [0] \longleftrightarrow \mathbf{u} \in V.$$

This motivates the definition of a quantum state $\rho_{[0]}$ associated with the orbit $[0]$, via

$$W_{\rho_{[0]}}(\mathbf{u}) := \frac{q(\tau_{[0]} + \mathbf{u})}{p_{[0]}}, \quad \forall \mathbf{u} \in V, \quad (34)$$

where $p_{[0]} = \sum_{\mathbf{u} \in V} q(\tau_{[0]} + \mathbf{u})$ to ensure proper normalization. The state $\rho_{[0]}(\mathbf{u})$ is not required to be a valid quantum state, i.e., to be positive semidefinite. The only requirement is a non-negative Wigner function, which is satisfied by definition. The fact that classical sampling algorithms can handle states which have a positive Wigner function but are not proper quantum states is familiar from the qudit case [17].

In analogy with Eq. (34), we will now define states $\rho_{[v]}$ for all orbits $[v] \in \mathcal{S}/V$. The state $\rho_{[0]}$ and its cousins will then be used in the interpretation of Algorithm 2.

For any $[v] \in \mathcal{S}/V$, pick a $\tau_{[v]} \in [v]$ and define

$$T_{\mathbf{a}}^{\gamma_{[v]}} := \lambda_{\tau_{[v]}}^{-1}(\mathbf{a})T_{\mathbf{a}}, \quad \forall \mathbf{a} \in V, \quad (35)$$

where on the right-hand side $T_{\mathbf{a}} = T_{\mathbf{a}}^{\gamma}$, as defined in Eq. (7). Denoting $\lambda_{\tau_{[v]}}(\mathbf{a}) = (-1)^{s_{[v]}(\mathbf{a})}$, for all $\mathbf{a} \in V$, we thus have the relation

$$\gamma_{[v]} \equiv \gamma + 2s_{[v]} \pmod{4}. \quad (36)$$

From the above definition of $T_{\mathbf{a}}^{\gamma_{[v]}}$, $\lambda_{\tau_{[v]}}(T_{\mathbf{a}}^{\gamma_{[v]}}) = 1$, for all $\mathbf{a} \in V$. We can thus reproduce for any orbit $[v]$ the previous argument for $[0]$. First, with Eq. (31),

$$\lambda_{\tau_{[v]}+\mathbf{u}}(T_{\mathbf{a}}^{\gamma_{[v]}}) = (-1)^{[\mathbf{a}, \mathbf{u}]}, \quad \forall \mathbf{a}, \mathbf{u} \in V. \quad (37)$$

Again by comparison with Eq. (22), the value assignments by the HVM and by the Wigner function $W^{\gamma_{[v]}}$ match if we identify, for all $\mathbf{u} \in V$,

$$(\tau_{[v]} + \mathbf{u}) \in [v] \longleftrightarrow \mathbf{u} \in V. \quad (38)$$

A state $\rho_{[v]}$ associated with any orbit $[v] \in \mathcal{S}/V$ can now be defined, via

$$W_{\rho_{[v]}}^{\gamma_{[v]}}(\mathbf{u}) := \frac{q(\tau_{[v]} + \mathbf{u})}{p_{[v]}}, \quad \forall \mathbf{u} \in V. \quad (39)$$

Therein, $p_{[v]} = \sum_{\mathbf{u} \in V} q(\tau_{[v]} + \mathbf{u})$. As before with $\rho_{[0]}(\mathbf{u})$, the state $\rho_{[v]}(\mathbf{u})$ is not required to be positive semidefinite.

Remarks. (i) For each $[v] \in \mathcal{S}/V$, the choice of the representative $\tau_{[v]}$ in Eq. (39) is arbitrary. Different choices lead to different $\gamma_{[v]}$, which are, however, related in a simple

way. Namely, the corresponding Wigner functions differ only by translation. By contrast, the Wigner functions $W^{\gamma_{[v'()]}}$ and $W^{\gamma_{[v()]}}$, for any $[v'] \neq [v]$, are not equivalent under translation.

(ii) We note that multiple Wigner functions have previously been discussed in relation to QCSI [18,19]. Therein, a quantum state is considered classical if *all* its Wigner functions are positive. Our viewpoint is the opposite. For a state to be considered classical, not even a single one of its Wigner functions has to be positive.

The states $\rho_{[v]}$, defined in Eq. (39) have the following relation to the state ρ of the quantum register.

Lemma 9. For any QCSI scheme (γ, \mathcal{M}) , if the state ρ of the quantum register has a noncontextual HVM description, then the states $\rho_{[v]}$ provide an ensemble representation $\mathcal{E}_{\rho} = \{(\rho_{[v]}, p_{[v]}), [v] \in \mathcal{S}/V\}$ of ρ , i.e.,

$$\rho = \sum_{[v] \in \mathcal{S}/V} p_{[v]} \rho_{[v]}. \quad (40)$$

When relating the classical simulation Algorithms 1 and 2, we apply Lemma 9 in particular to the input state ρ_{in} of the computation, i.e., magic state.

Proof of Lemma 9. Since the $T_{\mathbf{a}}$, $\mathbf{a} \in V$, form a basis of Hermitian operators on n qubits, it suffices to show that $\langle T_{\mathbf{a}} \rangle_{\rho_{\text{in}}} = \langle T_{\mathbf{a}} \rangle_{\sum_{[v]} p_{[v]} \rho_{[v]}}$, for all $\mathbf{a} \in V$.

$$\begin{aligned} \langle T_{\mathbf{a}} \rangle_{\sum_{[v]} p_{[v]} \rho_{[v]}} &= \sum_{[v]} p_{[v]} \langle T_{\mathbf{a}} \rangle_{\rho_{[v]}} \\ &= \sum_{[v]} p_{[v]} \lambda_{\tau_{[v]}}(\mathbf{a}) \langle T_{\mathbf{a}}^{\gamma_{[v]}} \rangle_{\rho_{[v]}} \\ &= \sum_{[v]} p_{[v]} \lambda_{\tau_{[v]}}(\mathbf{a}) 2^n \sum_{\mathbf{u} \in V} W_{\rho_{[v]}}^{\gamma_{[v]}}(\mathbf{u}) W_{T_{\mathbf{a}}^{\gamma_{[v]}}}^{\gamma_{[v]}}(\mathbf{u}) \\ &= \sum_{[v]} \lambda_{\tau_{[v]}}(\mathbf{a}) \sum_{\mathbf{u} \in V} q(\tau_{[v]} + \mathbf{u}) (-1)^{[\mathbf{a}, \mathbf{u}]} \\ &= \sum_{v \in \mathcal{S}} q(v) \lambda_{\tau_{[v]}}(\mathbf{a}) \lambda_v(T_{\mathbf{a}}^{\gamma_{[v]}}) \\ &= \sum_{v \in \mathcal{S}} q(v) \lambda_v(T_{\mathbf{a}}) \\ &= \langle T_{\mathbf{a}} \rangle_{\rho_{\text{in}}}, \end{aligned}$$

as required. We used Eq. (35) in line 2 above, Eq. (39) in line 4, Eq. (37) in line 5, and Eq. (35) in line 6. ■

With Lemma 9, we can now reinterpret the sampling from the HVM as the following two-stage process. In the first stage, equivalence classes $[v] \in \mathcal{S}/V$ are sampled from, according to the probabilities $\{p_{[v]}\}$. In the second stage, given a particular class $[v]$, the phase space V is sampled from, according to the conditional probability distribution $q_{[v]}/p_{[v]}$. The conditional probability distributions $q_{[v]}/p_{[v]}$ over V are regarded as Wigner functions $W_{\rho_{[v]}}^{\gamma_{[v]}}$ of states $\rho_{[v]}$ associated with the orbits $[v]$; cf. Eq. (39). See Fig. 5 for illustration.

Algorithm 2 can thus be simulated by a master algorithm calling Algorithm 1 as a subroutine, as follows. Step 1: A sample $v \in \mathcal{S}$ is drawn. This sample is converted into the into the pair $([v] \in \mathcal{S}, \mathbf{u} \in V)$, such that $v = \tau_{[v]} + \mathbf{u}$. Step 2: Algorithm 1 is called, with the sample $\mathbf{u}_1 := \mathbf{u}$ being handed over.

The orbit $[\nu]$ has no influence on how Algorithm 1 runs, but needs to be taken into account when the simulated measurement outcomes are returned. Namely, Algorithm 1 returns the values for $T_{\mathbf{a}}^{\gamma_{[\nu]}}$, not for the $T_{\mathbf{a}}$, with the standard phase convention γ . A conversion of those values is thus necessary, which proceeds by Eq. (35).

There is one more item to check: To employ sampling from Wigner functions $W^{\gamma_{[\nu]}}$ as a subroutine, the measurement of observables which leave positive Wigner functions W positive must also leave all Wigner functions $W^{\gamma_{[\nu]}}$ positive. Denote by $\mathcal{O}_{[\nu]}$ the nonextendable set of directly measurable observables with respect to the phase convention $\gamma_{[\nu]}$ (i.e., the set of Pauli observables whose measurement preserves non-negativity of the Wigner function $W^{\gamma_{[\nu]}}$). Then, we have the following result.

Lemma 10. For all $[\nu] \in \mathcal{S}/V$, it holds that $\mathcal{O}_{[\nu]} = \mathcal{O}$.

Proof of Lemma 10. By Eq. (10), for any phase convention γ it holds that

$$\beta(\mathbf{a}, \mathbf{b}) = \gamma(\mathbf{a}) + \gamma(\mathbf{b}) - \gamma(\mathbf{a} + \mathbf{b}) + 2\mathbf{a}_X \mathbf{b}_Z \pmod{4}.$$

Then, by Eq. (36), the function β based on a specific γ and the functions $\beta_{[\nu]}$ based on the corresponding $\gamma_{[\nu]}$ are related via

$$\beta_{[\nu]}(\mathbf{a}, \mathbf{b}) = \beta(\mathbf{a}, \mathbf{b}) + 2[s_{[\nu]}(\mathbf{a}) + s_{[\nu]}(\mathbf{b}) - s_{[\nu]}(\mathbf{a} + \mathbf{b})],$$

where the addition is again mod 4. Now assume that $T_{\mathbf{a}} \in \mathcal{O}$ and that $[\mathbf{a}, \mathbf{b}] = 0$. Then, $\{T_{\mathbf{a}}, T_{\mathbf{b}}\}$ is a jointly measurable set of observables, and thus, by property (ii) of Definition 3, $s_{[\nu]}(\mathbf{a}) + s_{[\nu]}(\mathbf{b}) - s_{[\nu]}(\mathbf{a} + \mathbf{b}) \pmod{2} = 0$. Hence,

$$\beta_{[\nu]}(\mathbf{a}, \mathbf{b}) = \beta(\mathbf{a}, \mathbf{b}), \quad \forall [\nu] \in \mathcal{S}/V,$$

for all above pairs $\mathbf{a} \in V_{\mathcal{O}}$, $\mathbf{b} \in V$. Thus, by Lemma 1, the measurement of $T_{\mathbf{a}}$ preserves positivity of the Wigner function W^{γ} if and only if it preserves positivity of the Wigner function $W^{\gamma_{[\nu]}}$, for any $[\nu] \in \mathcal{S}/V$. ■

This concludes the discussion of the relation between the classical simulation Algorithms 1 and 2. We have seen that if a given magic state can be described by an nCHVM, then this nCHVM can be viewed as the probabilistic combination of many non-negative Wigner functions, each compatible with the same set \mathcal{O} of measurable observables.

VI. A QUBIT SCHEME OF QCSI WITH MATCHING WIGNER FUNCTION

Four of the five theorems in the preceding sections begin with ‘‘For any QCSI scheme $(\gamma, \mathcal{M}) \dots$ ’’. We must thus ask: Are there any such schemes at all? This is the case for any number of qubits, as we now show by example.

A. Definition of the Wigner function

In this section, we focus on the properties of a single Wigner function. We make the choice

$$\gamma_0(\mathbf{a}) = \mathbf{a}_Z \cdot \mathbf{a}_X \pmod{4}, \quad (41)$$

which has the important and rare consequence that the corresponding Wigner function factorizes $W_{\rho \otimes \sigma} = W_{\rho} W_{\sigma}$ for all states ρ, σ . In fact, the factorization property already holds on the level of the Heisenberg-Weyl operators Eq. (7),

$$T_{\mathbf{a}} \otimes T_{\mathbf{b}} = T_{\mathbf{a}+\mathbf{b}}. \quad (42)$$

B. The set \mathcal{O} of directly measurable observables

Lemma 11. For $\gamma_0(\mathbf{a}) = \mathbf{a}_Z \cdot \mathbf{a}_X \pmod{4}$, the set \mathcal{O} of directly measurable observables is

$$\mathcal{O} = \{\pm X_i, \pm Y_i, \pm Z_i | i = 1, \dots, n\}.$$

This means first of all that the corresponding Wigner function W^{γ_0} has a corresponding QCSI scheme, and, perhaps surprisingly, spatial locality plays a role in it. Below, we first prove Lemma 11, and then flesh out the QCSI scheme corresponding to W^{γ_0} .

Proof of Lemma 11. We first show that the set $\mathcal{O} = \{X_i, Y_i, Z_i | i = 1, \dots, n\}$ satisfies the defining conditions Eqs. (12) and (16). Consider two commuting Pauli observables $T_{\mathbf{b}}, T_{\mathbf{c}}$ such that $T_{\mathbf{b}}$ is local to qubit k , and $T_{\mathbf{c}}$ is written as $T_{\mathbf{c}} = T_{\mathbf{c}'+\mathbf{c}''} = T_{\mathbf{c}'} \otimes T_{\mathbf{c}''}$, where $T_{\mathbf{c}'}$ acts nontrivially only on qubit k , and $T_{\mathbf{c}''}$ acts nontrivially only on the complement of qubit k . Then,

$$\begin{aligned} T_{\mathbf{b}} T_{\mathbf{c}} &= T_{\mathbf{b}} T_{\mathbf{c}'} \otimes T_{\mathbf{c}''} \\ &= i^{-\beta(\mathbf{b}, \mathbf{c}')} T_{\mathbf{b}+\mathbf{c}'} \otimes T_{\mathbf{c}''} \\ &= i^{-\beta(\mathbf{b}, \mathbf{c}')} T_{(\mathbf{b}+\mathbf{c}')+\mathbf{c}''} \\ &= i^{-\beta(\mathbf{b}, \mathbf{c}')} T_{\mathbf{b}+\mathbf{c}}. \end{aligned}$$

Therein, in lines 1 and 3 we used the property Eq. (42).

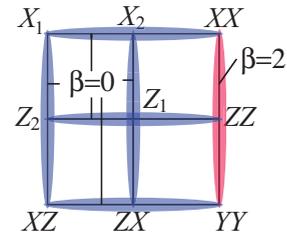
Since $T_{\mathbf{b}}$ and $T_{\mathbf{c}}$ are commuting, $\beta(\mathbf{b}, \mathbf{c}') \in \{0, 2\}$. Since $T_{\mathbf{b}}$ is local, by going through all the cases of local Pauli operators we find that $\beta(\mathbf{b}, \mathbf{c}') \in \{0, \pm 1\}$. Thus, $\beta(\mathbf{b}, \mathbf{c}') = 0$ is the only choice that satisfies both constraints. Therefore,

$$T_{\mathbf{b}+\mathbf{c}} = T_{\mathbf{b}} T_{\mathbf{c}}, \quad (43)$$

whenever $[T_{\mathbf{b}}, T_{\mathbf{c}}] = 0$, and $T_{\mathbf{b}}$ is local. Condition (12) is thus satisfied.

Next, since every multilocal Pauli operator can be written as a tensor product of local Pauli operators, and the local Pauli operators in such an expansion trivially commute, it follows that $V_{\mathcal{M}} = V$, as required by condition (16). We have thus shown that $\{X_i, Y_i, Z_i | i = 1, \dots, n\}$ is a possible set \mathcal{O} .

It remains to prove that the above \mathcal{O} is maximal, i.e., that \mathcal{O} cannot contain any additional observable without violating the condition Eq. (12). For $n = 1$ this is clear, and we only need to discuss the case of $n \geq 2$. To this end, consider the two-local Pauli operators, beginning with $Y \otimes Y$.



$Y \otimes Y$ is part of a context with $\beta = 2 \neq 0 \pmod{4}$. Therefore, with Lemma 1, $Y \otimes Y \notin \mathcal{O}$.

Now, conjugate the observables in the above diagram under a local Clifford unitary, and readjust the signs such that only observables $T_{\mathbf{a}}$ appear. In this way, any two-local Pauli observable can appear in the bottom left corner of the diagram. We observe that the four observables in the top left corner of

the diagram will remain local under such a transformation. As we have shown, all local Pauli operators $T_{\mathbf{b}}$ satisfy Eq. (43) for all commuting $T_{\mathbf{c}}$. Hence, (i) the four β s involving local observables remain $\beta = 0$ and (ii) the six β appearing in the square must sum to $2 \pmod 4$, as per Mermin's argument. Combining these two facts, we find that the two β involving the observable in the bottom right corner of the diagram cannot simultaneously be zero. Hence this observable cannot be in \mathcal{O} . Thus, no two-local Pauli observable is in \mathcal{O} .

Next, consider a Pauli observable $T_{\mathbf{b}}$ with a support of size greater than 2. Be J a set of two sites in the support of $T_{\mathbf{b}}$, $J = \{j, k\} \subset \text{supp}(T_{\mathbf{b}})$, and denote by $T_{\mathbf{b}'}$ the restriction of $T_{\mathbf{b}}$ to J , and by $T_{\mathbf{b}''}$ the restriction of $T_{\mathbf{b}}$ to the complement of J . Then, with Eq. (42), $T_{\mathbf{b}} = T_{\mathbf{b}'+\mathbf{b}''} = T_{\mathbf{b}'} \otimes T_{\mathbf{b}''}$. Now consider a second Pauli operator $T_{\mathbf{c}}$ that commutes with $T_{\mathbf{b}}$ and has support on J only. Then, using the property Eq. (42),

$$\begin{aligned} T_{\mathbf{b}}T_{\mathbf{c}} &= T_{\mathbf{b}''} \otimes T_{\mathbf{b}'}T_{\mathbf{c}} \\ &= i^{\beta(\mathbf{b}',\mathbf{c})} T_{\mathbf{b}'} \otimes T_{\mathbf{b}'+\mathbf{c}} \\ &= i^{\beta(\mathbf{b}',\mathbf{c})} T_{\mathbf{b}'+(\mathbf{b}'+\mathbf{c})} \\ &= i^{\beta(\mathbf{b}',\mathbf{c})} T_{\mathbf{b}+\mathbf{c}}. \end{aligned}$$

By the previous argument for two-local operators, for any $T_{\mathbf{b}}$ with support on two or more qubits, a commuting two-local Pauli operator $T_{\mathbf{c}}$ can be found such that $\beta(\mathbf{b}',\mathbf{c}) = 2$. Then, with Lemma 1, $\pm T_{\mathbf{b}} \notin \mathcal{O}$. ■

From Eq. (6) it follows that the set Ω of free states are tensor products of one-qubit stabilizer states. The group of free unitary gates therefore is the local Clifford gates.

C. Magic states and universality

From the perspective of computational universality of QCSI, all we do not know yet is what the magic states are. Since all state preparations and measurements are local in the present situation, any entanglement needed in the computation must be brought in by the magic state. That is, there is only one big entangled magic state. Factors of tensor product states cannot be coupled by the free operations.

In fact, one possibility is to use as magic state a slightly modified cluster state. We consider a set of qubits located on the vertices of a square lattice graph. We denote the set of its sites by \mathcal{V} and its adjacency matrix by Γ . We single out a subset $\mathcal{R} \subset \mathcal{V}$ of sites which are sufficiently sparse. Denote by A the observable $\frac{X+Y}{\sqrt{2}}$. With those definitions, the resource state $|\Psi\rangle$ is the unique joint eigenstate with eigenvalue 1 of the stabilizer operators

$$K_a^X := X_a \bigotimes_{b \in \mathcal{V}} Z_b^{\Gamma_{ab}}, \quad \text{if } a \in \mathcal{V} \setminus \mathcal{R}, \quad (44a)$$

$$K_a^A := A_a \bigotimes_{b \in \mathcal{V}} Z_b^{\Gamma_{ab}}, \quad \text{if } a \in \mathcal{R}. \quad (44b)$$

That this leads to universal quantum computation is easily shown by standard arguments pertaining to measurement-based quantum computation (MBQC). See Fig. 6 for illustration.

While being a valid scheme of QCSI, this is also MBQC. The distinction between MBQC and QCSI is thus blurred. By various equivalence transformations, we can make the

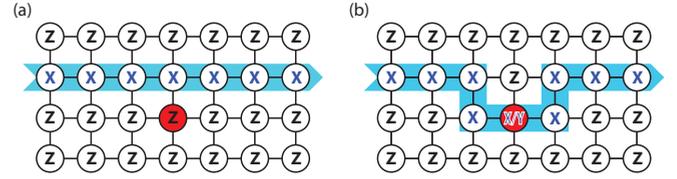


FIG. 6. QCSI with modified cluster state of Eqs. (44a) and (44b) as magic state, which is subjected to measurements of local Pauli operators X_i, Y_j, Z_k , for all $i, j \neq i, k \neq i, j \in \mathcal{V}$. The role of the Z measurements is to cut out of the plane a web corresponding to some layout of a quantum circuit, and the X measurements drive the MBQC simulation of this circuit [48]. The qubit in \mathcal{R} is displayed in red. By rerouting a wire piece, one may choose between implementing and not implementing a non-Clifford gate. (a) Identity operation on the logical state space. (b) Logical gate $e^{i\pi/4 Z}$.

above computational scheme look more like the known QCSI schemes, or more like standard MBQC.

Equivalent scheme 1. In all QCSI schemes worked out to date [15–18], the magic states are local to single or at most two particles. Although this is by no means necessary, we are used to those states being injected into the computation one by one. If desired, we may convert the above computational scheme into such a form, by conjugating it—the resource state, the measurable observables in \mathcal{O} , and the Wigner function W^0 —under the unitary transformation

$$U_{\text{Ising}} = \prod_{i,j \in \mathcal{V}} (\Lambda(Z)_{i,j})^{\Gamma_{ij}}. \quad (45)$$

In this way, we arrive at the following equivalent computational scheme. The resource state $|\Psi\rangle$ is converted into a tensor product state of individual qubits being in the state $|+\rangle_i$, defined by $X|+\rangle = |+\rangle$, for $i \in \mathcal{V} \setminus \mathcal{R}$, and $|A\rangle_j$, defined by $A|A\rangle = |A\rangle$, for $j \in \mathcal{R}$. The new magic states are thus the local states $|A\rangle_j$.

The new set \mathcal{O}_1 of directly measure observables is $\mathcal{O}_1 = \{K_a^X, K_a^Y, Z_a, a = 1, \dots, n\}$, where $K_a^Y = Y_a \bigotimes_{b \in \mathcal{V}} Z_b^{\Gamma_{ab}}$.

Equivalent scheme 2. We note in Eq. (44b) that stabilizer operators K^A of the magic state $|\Psi\rangle$ are not exactly stabilizer operators of cluster states. Therefore, we may apply the equivalence transformation

$$U_{\text{loc}} = \bigotimes_{j \in \mathcal{R}} e^{-i\pi/4 Z_j},$$

and obtain as the new magic state the standard cluster state. The new measurable observables are

$$\mathcal{O}_2 = \{X_i, Y_i, A_j, A'_j, Z_k | a \in \mathcal{V} \setminus \mathcal{R}, j \in \mathcal{R}, k = 1 \dots n\},$$

where $A' = (X - Y)/\sqrt{2}$. We note that the measurable observables which are not Z s are of the form

$$\mathcal{O}_i = \cos \phi_i X_i \pm \sin \phi_i Y_i,$$

as standard in MBQC [48]. A minor deviation from the standard remains. Namely, for each site i , only a single setting out of two is available for the measurement angle, either $\phi_i = 0$ or $\phi_i = \pi/4$. In standard MBQC, any angle $\phi_i \in [0, \pi/2]$ may be chosen. However, the present restriction does not affect computational universality.

VII. CONCLUSION

We have shown that for all schemes of quantum computation with magic states on qubits that satisfy the condition (P1), contextuality and Wigner function negativity are necessary resources for quantum computational universality and quantum speedup. This extends the earlier results [15,17] for qudits to qubits. Key to our construction is the additional condition (P1) imposed on QCSI schemes, which removes the phenomenon of state-independent contextuality from the free operations.

Despite the interchangeability of “contextuality” and “Wigner function negativity” in the above statement about computational resources, the results on contextuality are stronger. They imply their counterparts on the negativity of Wigner functions but are not implied by them. This arises because the equivalence between the existence of a noncontextual hidden variable model and non-negativity of the Wigner function—which characterized the qudit case [15,24]—no longer holds for qubits. Magic states can be described by an ncHVM but nonetheless have a negative Wigner function.

A related matter is the preservation of positivity under free QCSI operations. For any given QCSI scheme satisfying Condition P1, we find that the existence of an ncHVM for the state of the quantum register is preserved under all free operations. The non-negativity of Wigner functions is only preserved under free measurements, not under all free unitaries. The amount of negativity introduced by the free unitaries can be very large as measured by the sum negativity [45], without compromising efficient classical simulability.

We conclude with two questions:

(1) We have shown that QCSI schemes satisfying Conditions P1 and P2 exist for any number n of qubits, but we have not classified them. From initial numerical studies we know that (i) for $n = 2$ there are plenty of QCSI schemes—as specified by the function γ in Eq. (7)—that satisfy both Conditions P1 and P2. In particular, there are at least two distinct classes under Clifford equivalence. This tells us that there are solutions fundamentally different from Eq. (41). (ii) For $n \geq 3$, the solutions of Conditions P1 and P2 for γ are very sparse (we did not find any in 10^4 random trials). Yet, by Lemma 11 we know that solutions exist for any n . Can they be classified?

(2) Determining the cost of classically simulating universal quantum computing is a fascinating area of research [46,49–51]. Can the simulation methods developed here be extended to qubit QCSI schemes that do not satisfy the Condition P1?

ACKNOWLEDGMENTS

R.R. thanks Terry Rudolph and the Horodecki family for discussion. This work has been funded by NSERC (N.D., C.O., and R.R.), SIQS (J.B.V.), AQuS (J.B.V.), Cifar (N.D. and R.R.), and IARPA (R.R.). R.R. acknowledges support from a fellowship from the Cifar (Canadian Institute for Advanced Research) Quantum Information Program.

APPENDIX: PROOF OF THEOREM 4

Before we give the proof, we need to set up some additional notation. Namely, we extend the value assignments λ_ν . As per

Eq. (33) in the simulation, Algorithm 2 of Sec. VE, the update of the internal state ν in the measurement of an observable $T_{\mathbf{a}_t}$ invokes a random bit c_t , $\nu_t \rightarrow \nu_{t+1} = \nu_t + c_t \mathbf{a}_t$. Assume that the measurement sequence representing a given quantum algorithm has a maximum length t_{\max} . Since the values $\{c_t, t = 1, \dots, t_{\max}\}$ are random and uncorrelated with everything, it does not matter whether they are determined at runtime or before the computation, and we may include those values in the ncHVM value assignments λ_ν . That is, we extend the internal states $\nu \in \mathcal{S}$ to

$$\nu^* = (\nu, c_1, c_2, \dots, c_{t_{\max}}).$$

Correspondingly, the set of internal states becomes $\mathcal{S}^* := \mathcal{S} \times \mathbb{Z}_2^{\times t_{\max}}$.

A consequence of this definition is that the function λ_ν can now assign a value to a potentially larger set of observables than the inferable ones. Namely, consider a unitary U which can be implemented using free gates, free measurements, and ancilla states that can be described by an ncHVM, and an inferable observable $O \in M$. Then, the observable $U^\dagger O U$ can also be measured, namely by first implementing U and then measuring O . However, it is not *a priori* clear that $U^\dagger O U \in M$. Yet, by the general simulation Algorithm 2, given the initial internal state ν^* , a value $\lambda_{\nu^*}(U^\dagger O U)$ can be assigned to the observable $U^\dagger O U$.

An important point to note is that the above value assignments involving quantum circuits are in general context dependent. That is, they depend on how the unitary U is implemented as a circuit and which ancilla states are being consumed in the process. Further, for two given observables $O \in M$, O' , related via $O' = U^\dagger O U$, the unitary U is not unique, and $\lambda_{\nu^*}(U^\dagger O U)$ may again depend on the choice made. It is this potential contextuality of value assignments involving circuits that complicates the proof of Theorem 4.

Let us illustrate this property in an example. Suppose there is a QCSI setting where all observables X_i can be directly measured but Y_3 cannot. Further, assume that the unitary S_3 , with the property $Y_3 = S_3^\dagger X_3 S_3$, can be implemented using only free operations and ancillas that have an ncHVM description. The circuit for measuring Y_3 in this example is to first implement S_3 and then measure X_3 . Now the task is to measure Y_3 after first measuring either X_1 or X_2 . Are the corresponding values $\lambda_{\nu^*}(Y_3)$ assigned by the internal state ν^* via the simulation Algorithm 2 the same in those two situations?

This is in general not the case. Suppose that ν^* is such that $c_1 = 1$. Then, if X_1 was measured first, the HVM state after the first measurement is $\nu + \mathbf{a}_{X_1}$. Analogously, if X_2 was measured first, the HVM state after the first measurement is $\nu + \mathbf{a}_{X_2}$. With the second measurement in the sequence begins the implementation of the measurement of Y_3 . By construction, the measurement strategy does not depend on the outcome of the first measurement. Therefore, the second measured observable is the same $T_{\mathbf{a}_2}$ in both cases. However, the value assigned to this observable by the internal state ν^* through Eq. (31) differs if $[\mathbf{a}_2, \mathbf{a}_{X_1}] \neq [\mathbf{a}_2, \mathbf{a}_{X_2}]$. In this situation, the HVM prediction for the outcome obtained in the second measurement depends on which observable was measured first. Since the choice of observable in the third measurement may depend on the outcome of the second measurement, the two measurement sequences may now run down completely

different tracks, and there is no guarantee whatsoever that the value assigned to Y_3 will be the same. Hence, the value assignments involving circuits are potentially contextual.

Proof of Theorem 4. The outline of the proof is as follows. Suppose the statement is wrong, i.e., it is possible that the initial magic state has a description of an ncHVM and simultaneously gives rise to universal quantum computation. Then, by assumptions U1 and U2, universality implies that an encoded GHZ state can be prepared. Further, by U2 and U3 it is possible to randomly measure in one of the four contexts $\mathcal{E}(X_1, X_2, X_3)$, $\mathcal{E}(X_1, Y_2, Y_3)$, $\mathcal{E}(Y_1, X_2, Y_3)$, $\mathcal{E}(Y_1, Y_2, X_3)$. In each such measurement, the three outcomes are multiplied and it is verified whether the product matches the corresponding stabilizer eigenvalue of the GHZ state.

We will show below that if the initial (magic) state has an ncHVM description, then the simulation Algorithm 2 will fail to reproduce the quantum prediction for the stabilizer eigenvalues of the GHZ state with a probability $\geq 1/8$. However, Algorithm 2 has been proven to be correct (see Theorem 5), and therefore computational universality and availability of an ncHVM description for the initial state cannot simultaneously apply.

We now describe the measurement procedures for the four measurement contexts. Case (i): $\mathcal{E}(X_1, X_2, X_3)$. The three observables $\mathcal{E}(X_1)$, $\mathcal{E}(X_2)$, $\mathcal{E}(X_3)$ are jointly measured, which is guaranteed to be possible by U3.

Case (ii): the three remaining contexts $\mathcal{E}(X_i, Y_j, Y_k)$, $i \neq j \neq k \in \{1, 2, 3\}$. For each of the three measurement contexts, we apply the following procedure:

(1) Measure the observable $\mathcal{E}(X_i)$.

(2) Implement the measurement sequence corresponding to the execution of the encoded gate $\mathcal{E}(S_1 \otimes S_2 \otimes S_3)$, giving rise to a free unitary $g \in G$ which is propagated out of the circuit; see Fig. 3.

(3) Measure the observables $g^\dagger \mathcal{E}(X_i) g$, for $i = 1, \dots, 3$. The outcomes for $\mathcal{E}(Y_1)$, $\mathcal{E}(Y_2)$, and $\mathcal{E}(Y_3)$ are thereby obtained. Among them, the value for $\mathcal{E}(Y_i)$ is discarded.

Step 1 of this procedure is possible by U3, and step 2 is possible by U2. Regarding step 3, the unitary g is in general dependent on them measurement outcomes obtained in implementing the circuit for $\mathcal{E}(S_1 \otimes S_2 \otimes S_3)$, but in all cases g is free ($g \in G$). Thus, the observables $g^\dagger \mathcal{E}(X_i) g$ are in \mathcal{O} and can be directly measured.

From the perspective of the hidden variable model, for any internal state $\nu^* \in \mathcal{S}^*$, the values $\lambda_{\nu^*}(\mathcal{E}(X_i))$, $\lambda_{\nu^*}(\mathcal{E}(Y_i))$, $i = 1, \dots, 3$ are assigned to the corresponding observables via the

simulation Algorithm 2 applied to the above measurement procedure. We now show that these value assignments are noncontextual if $c_1 = 0$.

(i) $\mathcal{E}(X_i)$. Since the $\mathcal{E}(X_i)$ are measured first in every context they appear in, $\lambda_{\nu^*}(\mathcal{E}(X_i)) = \lambda_{\nu}(\mathcal{E}(X_i))$. They are thus context independent.

(ii) $\mathcal{E}(Y_i)$. We show this by the example of $\mathcal{E}(Y_1)$, the other two cases are analogous. $\mathcal{E}(Y_1)$ appears in two measurement contexts, namely $\mathcal{E}(Y_1, X_2, Y_3)$ and $\mathcal{E}(Y_1, Y_2, X_3)$. The corresponding measurement procedure differ only in one respect. In the former context $\mathcal{E}(X_2)$ is measured in the first step, and in the latter context $\mathcal{E}(X_3)$ is measured in the first step. Now, by the assumption of $c_1 = 0$, the internal state ν remains unchanged under this first measurement in both cases, and in particular remains the same for both cases. After the first step, the procedures are thus identical, and the same value $\lambda_{\nu^*}(\mathcal{E}(Y_1))$ is assigned by the HVM in either context.

Thus, if $c_1 = 0$, all values $\lambda_{\nu^*}(\mathcal{E}(X_i))$, $\lambda_{\nu^*}(\mathcal{E}(Y_i))$, $i = 1, \dots, 3$ assigned by the HVM are thus noncontextual, as claimed. By Mermin's argument, such noncontextual value assignments fail to reproduce the respective stabilizer eigenvalue of the encoded GHZ state in at least one of the contexts. Thus, if the measurement contexts are randomly chosen and if $c_1 = 0$, then the HVM fails to match the quantum prediction with a probability of $\geq 1/4$.

Further, for any internal state $\nu \in \mathcal{S}$, the conditional probability for $c_1 = 0$ in ν^* given ν equals 1/2 (recall that c_1 is the result of an unbiased coin flip). Therefore, the HVM fails to match the quantum prediction with a probability $\geq 1/8$.

Finally, universality only requires the preparation of an encoded GHZ state $\mathcal{E}(|\text{GHZ}\rangle)$ up to an arbitrarily small error $\epsilon > 0$, and likewise only the approximate realization of the gates $\mathcal{E}(S_i)$, up to the same error ϵ . When taking such errors into account, the HVM prediction remains unchanged. If the initial state can be described by a noncontextual HVM, the measurement record will fail to reproduce the quantum mechanical prediction of the measurement record with a probability $\geq 1/8$.

The quantum prediction does now differ from the ideal scenario. With a slightly erroneous encoded GHZ state $\mathcal{E}(|\text{GHZ}\rangle)$ and slightly erroneous gates $\mathcal{E}(S_i)$, the measurement record will differ from the measurement record for the ideal circuit with a probability $p(\epsilon)$. However, $p(\epsilon)$ can be made arbitrarily small by decreasing ϵ . Thus, even in the presence of a small error ϵ , the quantum prediction can be distinguished from the HVM prediction. ■

[1] J. S. Bell, On the problem of hidden variables in quantum mechanics, *Rev. Mod. Phys.* **38**, 447 (1966).
 [2] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, *Journal of Mathematics and Mechanics* **17**, 59 (1967).
 [3] N. D. Mermin, Hidden variables and the two theorems of John Bell, *Rev. Mod. Phys.* **65**, 803 (1993).
 [4] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, *New J. Phys.* **13**, 113036 (2011).

[5] A. Cabello, S. Severini, and A. Winter, Graph-Theoretic Approach to Quantum Correlations, *Phys. Rev. Lett.* **112**, 040401 (2014).
 [6] A. Acín, T. Fritz, A. Leverrier, and A. Belén Sainz, A combinatorial approach to nonlocality and contextuality, *Comm. Math. Phys.* **334**, 533 (2015).
 [7] N. D. Mermin, Extreme Quantum Entanglement in a Superposition of Macroscopically Distinct States, *Phys. Rev. Lett.* **65**, 1838 (1990).

- [8] V. Scarani, A. Acin, E. Schenck, and M. Aspelmeyer, Nonlocality of cluster states of qubits, *Phys. Rev. A* **71**, 042325 (2005).
- [9] O. Gühne, G. Toth, P. Hyllus, and H. J. Briegel, Bell Inequalities for Graph States, *Phys. Rev. Lett.* **95**, 120405 (2005).
- [10] E. Wigner, On the quantum correction for thermodynamic equilibrium, *Phys. Rev.* **40**, 749 (1932).
- [11] K. S. Gibbons, M. J. Hoffman, and W. K. Wootters, Discrete phase space based on finite fields, *Phys. Rev. A* **70**, 062101 (2004).
- [12] D. Gross, Computational power of quantum many-body states and some results on discrete phase spaces, Ph.D. thesis, Imperial College, London, 2008 (unpublished).
- [13] A. Mari and J. Eisert, Positive Wigner Functions Render Classical Simulation of Quantum Computation Efficient, *Phys. Rev. Lett.* **109**, 230503 (2012).
- [14] S. Rahimi-Keshari, T. C. Ralph, and C. M. Caves, Sufficient Conditions for Efficient Classical Simulation of Quantum Optics, *Phys. Rev. X* **6**, 021039 (2016).
- [15] M. Howard *et al.*, Contextuality supplies the ‘magic’ for quantum computation, *Nature (London)* **510**, 351 (2014).
- [16] N. Delfosse, P. Allard-Guerin, J. Bian, and R. Raussendorf, Wigner Function Negativity and Contextuality in Quantum Computation on Rebits, *Phys. Rev. X* **5**, 021003 (2015).
- [17] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, Negative quasiprobability as a resource for quantum computation, *New J. Phys.* **14**, 113011 (2012).
- [18] E. F. Galvão, Discrete Wigner functions and quantum computational speedup, *Phys. Rev. A* **71**, 042302 (2005).
- [19] C. Cormick, E. F. Galvão, D. Gottesman, J. P. Paz, and A. O. Pittenger, Classicality in discrete Wigner functions, *Phys. Rev. A* **73**, 012301 (2006).
- [20] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Phys. Rev. A* **71**, 022316 (2005).
- [21] R. L. Hudson, When is the Wigner quasi-probability density non-negative? *Rep. Math. Phys.* **6**, 249 (1974).
- [22] R. W. Spekkens, Negativity and Contextuality are Equivalent Notions of Nonclassicality, *Phys. Rev. Lett.* **101**, 020401 (2008).
- [23] S. Bravyi, Universal quantum computation with the $\nu = 5/2$ fractional quantum Hall state, *Phys. Rev. A* **73**, 042313 (2006).
- [24] N. Delfosse, C. Okay, J. Bermejo-Vega, D. E. Browne, and R. Raussendorf, Equivalence between contextuality and negativity of the Wigner function for qudits, [arXiv:1610.07093](https://arxiv.org/abs/1610.07093).
- [25] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 1997.
- [26] H. Zhu, Permutation Symmetry Determines the Discrete Wigner Function, *Phys. Rev. Lett.* **116**, 040501 (2016).
- [27] H. Zhu, Multiqubit Clifford groups are unitary 3-designs, [arXiv:1510.02619](https://arxiv.org/abs/1510.02619).
- [28] A. Peres, Incompatible results of quantum measurements, *Phys. Lett. A* **151**, 107 (1990).
- [29] Y. Han, L. Hemaspaandra, and T. Thierauf, Threshold computation and cryptographic security, *SIAM J. Comput.* **26**, 59 (1997).
- [30] S. Aaronson, Quantum computing, post-selection and probabilistic polynomial time, *Proc. R. Soc. London, Ser. A* **461**, 3473 (2005).
- [31] S. Aaronson and A. Arkhipov, The computational complexity of linear optics, in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing, STOC '11* (ACM, New York, NY, 2011), pp. 333–342.
- [32] M. J. Bremner, R. Jozsa, and D. J. Shepherd, Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy, *Proc. R. Soc. London, Ser. A* **467**, 459 (2011).
- [33] T. Morimae, K. Fujii, and J. F. Fitzsimons, On the Hardness of Classically Simulating the One Clean Qubit Model, *Phys. Rev. Lett.* **112**, 130502 (2014).
- [34] M. J. Hoban, J. J. Wallman, H. Anwar, N. Usher, R. Raussendorf, and D. E. Browne, Measurement-Based Classical Computation, *Phys. Rev. Lett.* **112**, 140505 (2014).
- [35] J. Bermejo-Vega, N. Delfosse, C. Okay, D. E. Browne, and R. Raussendorf, Contextuality as a resource for qubit quantum computation, [arXiv:1610.08529](https://arxiv.org/abs/1610.08529).
- [36] M. Howard, E. Brennan, and J. Vala, Quantum contextuality with stabilizer states, *Entropy* **15**, 2340 (2013).
- [37] N. D. Mermin, Simple Unified form for No-Hidden-Variables Theorems, *Phys. Rev. Lett.* **65**, 3373 (1990).
- [38] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, *Phys. Rev. A* **54**, 1098 (1996).
- [39] A. Steane, Multiple particle interference and quantum error correction, *Proc. R. Soc. London, Ser. A* **452**, 2551 (1996).
- [40] D. W. Kribs, R. Laflamme, D. Poulin, and M. Lesosky, Operator quantum error correction, *Quantum Information & Computation* **6**, 382 (2006).
- [41] D. Poulin, Stabilizer Formalism for Operator Quantum Error Correction, *Phys. Rev. Lett.* **95**, 230504 (2005).
- [42] D. Bacon, Operator quantum error-correcting subsystems for self-correcting quantum memories, *Phys. Rev. A* **73**, 012340 (2006).
- [43] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
- [44] H. Bombin, Topological subsystem codes, *Phys. Rev. A* **81**, 032301 (2010).
- [45] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer computation, *New J. Phys.* **16**, 013009 (2014).
- [46] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating Outcome Probabilities of Quantum Circuits Using Quasiprobabilities, *Phys. Rev. Lett.* **115**, 070501 (2015).
- [47] J. J. Wallman and S. D. Bartlett, Nonnegative subtheories and quasiprobability representations of qubits, *Phys. Rev. A* **85**, 062121 (2012).
- [48] R. Raussendorf and H. J. Briegel, A One-Way Quantum Computer, *Phys. Rev. Lett.* **86**, 5188 (2001).
- [49] M. van den Nest, A monomial matrix formalism to describe quantum many-body states, *New J. Phys.* **13**, 123004 (2011).
- [50] S. Bravyi, G. Smith, and J. Smolin, Trading Classical and Quantum Computational Resources, *Phys. Rev. X* **6**, 021043 (2016).
- [51] M. Howard and E. T. Campbell, Application of a Resource Theory for Magic States to Fault-Tolerant Quantum Computing, *Phys. Rev. Lett.* **118**, 090501 (2017).